Undergraduate Student Agreement With Reformed Introductory Physics Classes

Matthew Wilcox
University of Central Florida

UNDERGRADUATE STUDENT AGREEMENT WITH REFORMED INTRODUCTORY PHYSICS CLASSES

by
MATTHEW WILCOX
B.S. University of Maryland Baltimore County, 2014
M.S. University of Central Florida, 2016

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Physics
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Summer Term
2018

Major Professor: Jacquelyn J. Chini

ABSTRACT

In this study, I investigate student "buy-in", defined as students' proper understanding of and agreement with the class format, for introductory studio physics classes that incorporate lectures, labs, and group problem-solving activities into one interactive environment. I also investigate the ways in which instructors try to gain student buy-in to their class. Research has shown that student resistance to reformed instruction is a barrier to an instructor's use of research-based instructional strategies that are common to the studio class. Expectancy value theory suggests that by gaining student buy-in to the reformed class format, student resistance will decrease thus allowing a more effective class. I created a survey to measure student agreement with their class and another survey to determine the strategies that instructors use to gain student buy-in. I describe the responses to the surveys and use hierarchical models to determine if student agreement predicts their performance in the class and if the instructor strategies have an effect on student agreement. To triangulate these findings, I also interviewed instructors and students. From the surveys, I found that students disagree with the time spent lecturing and the importance and time spent reading outside of class. This is important because student agreement with the time spent in class predicts favorable attitudes about physics and their agreement with the time spent outside of class predicts a higher expected final grade. From the interviews, I discovered that both instructors and students believe that using evidence to justify the class format would be an effective strategy to gain agreement. However, few instructors used evidence due to a lack of prepared materials. Future work should develop materials to support instructors in presenting evidence about studio's effectiveness and investigate the impact on student buy-in and other outcomes.

Dedicated to my family, whose only wish is that I am happy no matter what. To my wife, who dropped everything to come with me as I pursued this degree.

ACKNOWLEDGMENTS

Most importantly, I would like to acknowledge my adviser, Dr. Jacquelyn Chini. She has provided me with so many opportunities to grow as a researcher and as a teacher. She allowed me to take this research idea I developed after my first year in graduate school and turn it into a full dissertation. She gave me the freedom to conduct the research as I felt necessary, while providing excellent feedback to ensure that I was on the right path. Outside of my academic endeavors, she has shown interest and concern for me and my fellow graduate students in our everyday lives, which I truly appreciate.

I would also like to acknowledge those in our physics education research group: Westley James, Brian Zamarripa Roman, Constance Doty, Erin Scanlon, and the now graduated Jarrad Pond. Your friendship and insight have helped me along throughout my entire research.

Lastly, I must acknowledge the three people by my side throughout my first year of graduate school: Tommy Boykin, Becca Cebulka, and Walt Malone. We have become good friends and without them I surely would not have made it past the first year.

Thank you, everyone!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

CLASP: Collaborative Learning through Active Sense-making in Physics

CLASS: Colorado Learning Attitudes about Science Survey

CSEM: Conceptual Survey of Electricity and Magnetism

FCI: Force Concept Inventory

GTA: Graduate Teaching Assistant

HLM: Hierarchical Linear Model

HMEC: Hypothetical Most Effect Class

IMPEC: Integrated Math, Physics, Engineering, and Chemistry Curriculum

ISLE: Investigative Science Learning Environment

LA: Learning Assistant

MCAT: Medical College Admissions Test

MPEX: Maryland Physics Expectations Survey

PER: Physics Education Research

PEVA: Pedagogical Expectation Violation Assessment

PUMBA: Perceptions of Undergraduates Measure of Buy-in Achievement

RBIS: Research-Based Instructional Strategy

RIOT: Real-time Instructor Observing Tool

SCALE-UP: Student Centered Active Learning Environment with Upside-down Pedagogies

SIMBA: Survey of Instructor Methods for Buy-in Achievement

STEM: Science Technology Engineering and Mathematics

StRIP: Student Response to Instructional Practices

TEAL: Technology Enabled Active Learning

UCF: University of Central Florida

CHAPTER ONE: INTRODUCTION

In this dissertation, I study student agreement with the class format of reformed student-centered introductory physics classes. I want to know how well students agree with the class format and specific activities that occur for the class so that instructors may learn which aspects of their class they should target to increase student agreement. In increasing student agreement with the format of the class, students should be less resistant to the research-based instructional strategies used in the class, thus making those strategies more successful and effective for student learning. To test this, I explore how student agreement to a class affects their performance in the class and their attitudes about physics. Lastly, to provide practical advice to instructors who want to achieve greater student agreement with their class, and thus ensure less resistance to the class, I investigate the strategies instructors use to gain student agreement and the effectiveness of these strategies.

1.1 Motivations

My studies in Physics Education Research (PER) at the University of Central Florida (UCF) began by observing the teaching practices of Graduate Teaching Assistants (GTAs) in the lab component of our algebra-based introductory mechanics class (Wilcox, Yang, & Chini, 2016). The lab component for this class is unique; it differs from the more traditional lab style that is taught in the calculus-based courses of the introductory sequence in that the algebra-based introductory mechanics lab component is a hybrid of discussion and lab. For roughly the first hour and fifteen minutes, groups of students complete worksheets modeled from the University of Maryland's *Open Source Tutorials* (Elby, Scherr, Goertzen, & Conlin, 2008) with word problems based on Heller and Heller's context-rich written problems (2001). Following this discussion style portion of the class, there is a fifteen-minute group quiz. The remaining hour and twenty minutes is devoted to lab activities following Etkina and Van Heuvelen's Investigative Science Learning Environment, or ISLE (2001).

We call this discussion and lab hybrid format "mini-studio" for its resemblance to the studio classroom format, which is a class that combines lecture, discussion, and lab, in one setting. The mini-studio was developed with funding from the National Science Foundation Transforming Undergraduate STEM Education program, so it incorporates research-based instructional strategies that require GTAs to interact in different ways than in a traditional lab or recitation section. For example, in a traditional recitation section, one typically expects a GTA to be at the board showing students how to solve problems. However, in the *Open Source Tutorials*, GTAs are instructed to "not pick up a pen or pencil to show [the students] how to do something" (Redish, 2009). In regard to the lab portion, the ISLE method places a strong emphasis, as evidenced by the grading rubric focusing on scientific ability on the process of "doing science" in a lab instead of an outcome-driven, confirmatory lab (Etkina & Van Heuvelen, 2007). GTAs in mini-studios should allow their students to develop their own research question and the process with which they will answer that question. This practice is in contrast to a traditional lab where students may be told what the experimental question is and what they need to do to answer it. Thus, GTAs in mini-studios may need to show restraint and let their students go on divergent paths. As long as students are properly "doing science", they are still performing well in an ISLE lab.

Since the teaching methods needed for a successful mini-studio are a departure from the traditional expectations of a GTA, we were interested to see what exactly GTAs were doing in the class. Did their teaching practices meet the expectations of the lab coordinator? What might influence how they teach?

We measured the GTAs' teaching practices using the Real-time Instructor Observing Tool, or the RIOT (West, Paul, Webb, & Potter, 2013). Using the RIOT, we found a GTA's RIOT profile, which is the percentage of class time a GTA spends on actions such as explaining, listening to a question, engaging in open dialogue, and passively observing the classroom. In an interview with each GTA after the semester in which we observed them, we asked the GTAs to write down or predict four RIOT profiles: 1) How do you think the lab coordinator wants his GTAs to spend their time on these actions in the mini-studio? 2) What do you think is the most helpful way to spend your time on these actions in the mini-studio? 3) How do you think the students want their GTA to spend their time on these actions in the mini-studio? 4) How do you think you actually spent your time on these actions in the mini-studio?

By comparing the GTAs' responses to these items with other items or with their actual RIOT profile, we determined that GTAs agreed with how they felt they were expected to teach. What they felt was the most helpful RIOT profile was similar to what they thought the lab coordinator wanted from them. However, this did not translate to their actual teaching practice. We found that the RIOT profile for their actual teaching practice was most like what they felt the students wanted from them, which was not what the TAs felt was the most helpful way to teach and not how the class was meant to be taught. So, there appeared to be two major influences on GTAs: 1) The design of the course as expressed by the lab coordinator; and 2) the GTA's perceived desires of the students. Unfortunately, these two influences were not alike and ultimately the TAs' actions reflected that of the perceived student desires.

Based on these findings, I saw two possible solutions to support our GTAs in teaching as intended by the course developer. We could tell GTAs to ignore what the students want and reinforce the pedagogy of the course. We could also try to adjust the students' desires to better match the course pedagogy. As a teacher, it would not be best to ignore the wishes of your students; students should feel comfortable and satisfied with their class. Students who feel anxiety in their class due to the use of active learning strategies report lower grades (England, Brigati, & Schussler, 2017) and academic stress has been found as a predictor of students leaving the STEM fields (Barthelemy, Hedberg, Greenberg, & McKay, 2015). A teacher must endeavor to get students to agree with the design of the course so that the teacher can use the research-based instructional strategies and students can feel satisfied with their education, making the class successful for both parties. Doing so is no small task. It requires being clear about the format of the class from the very beginning of the semester. One should plan for the disagreements students may have with the reformed class design and address them at the beginning as well as throughout the semester. Exploring what instructors do to get students to agree with a reformed class design and what it means for the students when they agree with design forms the basis of my dissertation research. I feel that getting students to agree with the class format eliminates a major barrier for both instructors and students in reformed classrooms. If students in the mini-studios had agreed with the mini-studio format, meaning they wanted to be taught in that manner, then it would not have mattered if the GTAs were more influenced by the students or the course designer. Both influences on the GTA would have wanted the GTA to teach in the same manner, using the research-based instructional strategies. When students agree with the class format, they may be more cooperative and in a better state of mind to learn. It may also allow instructors to use research-based instructional strategies without resistance which would again be beneficial for the students. Thus, I hypothesize that students who buy in to the class format will perform better in the class.

This need not only apply to the mini-studio. Any class in which the pedagogy is a departure from the traditional may be one in which students disagree with the course design. Courses that implement the studio classroom design instead of the traditional lecture hall design may experience this issue. These studio classes are a relatively new way of teaching and heavily rely on research-based instructional strategies to be successful. In studio classes, all aspects of the class occur in the same environment with the same set of instructors, making it ideal for a research study since there are no other external aspects of the class (like a separate lab with different instructors) that could affect the study. Additionally, prior research that focused on student expectations (Chini, Gaffney, & Al-Rawi, 2013; Gaffney, Gaffney, & Beichner, 2010) and student response to active learning (DeMonbrun et al., 2017; Nguyen et al., 2017), studies similar to mine, used an active learning or studio class for their investigations. For these reasons, my research on student agreement will focus primarily on studio classrooms.

1.2 Definitions of Key Concepts and Terms

Before going further into my study, I should define several key concepts and terms that I use throughout the dissertation and that are fundamental to my research. First, I describe my definition of "student buy-in" to the class format. Next, I briefly define student affect and how it relates to student buy-in. Finally, I describe the studio classroom and list several of the research-based instructional strategies that a studio instructor may use.

1.2.1 Student Buy-in

"Student buy-in" is the degree to which a student: a) understands the format of the course; and b) agrees with the format of the course. For this research, the format of a course is defined by the time spent on activities for the class and the importance placed on these activities. While this research examines both aspects of student buy-in, student agreement with the format of the course is the primary focus.

1.2.2 Student Affect

Student affect, or their affective domain, has been defined as “the attitudes, interests, and values that students exhibit and acquire in school” (Popham, 2009). In general, student affect may be described as a student’s emotions or feelings about a topic. Student buy-in and agreement are descriptors of a how a student values aspects of the class and thus they are contained within the domain of student affect.

Figure 1. An image of a studio classroom.

1.2.3 Studio Classes

Studio physics classes are classrooms in which the traditional components of a physics class, i.e. lecture, labs, problem solving, and discussion, may happen in the same space. The class time may be spent on any of these activities in the classroom, flowing from one activity to the next as the instructor chooses. An important feature of studio style classes is an emphasis on group work and a de-emphasis on lecture. Visually, studio classes are distinct from traditional lecture halls. To facilitate group work, studio classrooms have group tables, such as those shown in Figure 1 which seat up to three groups of three students. Often, these tables have desktop or laptop computers for the students to use. The walls of a studio classroom are often entirely covered with whiteboards, which facilitate group work and allow for these groups to present their work to the class. Lastly, multiple screens or televisions are placed around the classroom so students facing any direction can see anything that an instructor might project such as a PowerPoint or a worksheet under a DocCam. Figure 1 shows the studio classroom at UCF, exhibiting large circular tables that seat nine students, three computers at each table, and white-boards surrounding the entire class. Also note the lab equipment interface devices for each group (the objects the flags are standing on), which allow for experiments to be completed within the studio room. Not shown in the figure are large screens that drop down at either end of the room.

1.2.4 Research-Based Instructional Strategies (RBISs)

RBISs are practices or pedagogies instructors may use that have been empirically shown to increase student learning. Numerous physics RBISs exist, including the *Open Source Tutorials* (Elby et al., 2008) and ISLE labs (Etkina & Heuvelen, 2001) used in UCF's mini-studios. Additional RBISs are Peer Instruction (Mazur & Somers, 1999), Context-Rich Problems (P. Heller & Hollabaugh, 1992), and Interactive Lecture Demonstrations (Sokoloff & Thornton, 1997). For more examples of RBISs, see Table I of Henderson, Dancy, and Niewiadomska-Bugaj's study on the use of RBISs in introductory physics (2012).

1.3 Research Questions

Four major questions guide this research. These questions explore the level of student buy-in for a class, and the ways in which instructors attempt to get a high level of student buy-in. They also explore the relationship between the methods instructors use to get buy-in and the level of student buy-in for their class as well as the relationship between student agreement and student performance.

The four major questions and their constituents are:

Q1: How well do students buy in to reformed instruction?

(a) How well do students understand how time will be spent in class?

(b) How well do students understand how frequently specific activities will occur in class?

(c) How well do students agree with the class format?

(d) What specific class activities do students agree or disagree with?

(e) Do student buy-in (understanding and agreement) measures change between the beginning and end of the semester?

Q2: What effect does student agreement have on student performance?

(a) Does students' agreement with the class format predict their final grade?

(b) Does students' agreement with the class format predict their performance on concept tests?

(c) Does students' agreement with the class format predict their performance on attitudinal surveys?

Q3: What do instructors do in an attempt to gain a high level of student buy-in?

(a) What class activities do instructors commonly talk about with their students to set their expectations and get their agreement with those expectations?

(b) What methods or topics do instructors commonly use to get student agreement?

(c) Through which formats do instructors commonly have these discussions?

(d) Do these things change from the beginning to the end of the semester?

Q4: What effect do instructors' discussions of the class format have on student buy-in?

(a) Is the use of certain methods or formats to get student agreement associated with higher levels of student agreement?

1.4 Overview of Methodologies

I developed an online survey given to studio physics students to quantify student agreement with the importance of and time spent on in-class and out-of-class activities. I made another online survey for the instructors of studio physics classes to determine what they discussed with the students about the class format, how they tried to get students to agree with the format, and the ways in which they had this discussion. Students and instructors took these surveys at the beginning and end of the semester. At the end of the semester, I collected student scores to concept tests, attitudinal surveys, and students' expected final grades. I used hierarchical linear regressions to explore the effect of student agreement on student outcomes and the effect of instructor methods on student agreement. To triangulate (using multiple methods to support and validate my findings) the results of the surveys, I conducted interviews with instructors and students.

1.5 Organization of Dissertation

In Chapter One, the introduction, I describe the foundation of this research. I explain the initial research I did that created my interest in student buy-in, define key terms that will be used frequently in this dissertation, and outline the research questions that guide my work. In Chapter Two, I describe the theoretical and conceptual framework that applies to my research. The goal in describing these frameworks is to allow the reader to understand my position and viewpoints on my research. Doing so provides disclosure on how I choose to interpret my data and insight into the reasoning for some of the decisions made throughout the research. In Chapter Three, the literature review, I describe the foundational research I use to support my work. I also describe the work done by other groups who are also interested in student response to instructional practices and the ways in which instructors can get student agreement. In this chapter, I also discuss the areas in which my work will extend beyond the already published material. In Chapter Four, I introduce the two surveys I developed to measure student buy-in and the ways in which instructors attempt to gain a high level of student buy-in. I describe the surveys, explain how to calculate buy-in from the results, and discuss how I found evidence of validity for the surveys. In Chapter Five, the methods section, I discuss the study population, interviews, and the analyses that are run on my data. In Chapter Six and Seven, I present my findings from the surveys and then the interviews. Finally, in Chapter Eight I summarize my research, discuss its implications, describe limitations of the research, and present possible future research directions.

CHAPTER TWO: THEORETICAL AND CONCEPTUAL FRAMEWORK

So the reader may better understand the structure of my study and my interpretations of the results, in this chapter I describe the lens with which I view my research. First, I discuss the mental model of student resistance that I have developed as a result of my initial research described in the "Motivation" section (1.1). Then, I introduce the theoretical framework that describes broadly the context of my research, what its goals are, and the common methodologies used for this type of research. Finally, I explain the conceptual framework that I use to describe the connections between the variables in my research.

2.1 My Mental Model

From my initial research into the instructional practices of GTAs in a reformed discussion/lab section, I found that their practices were affected by their perceptions of their students' desires for the course; GTAs taught how they thought their students wanted them to teach. This was, unfortunately, not in the manner that the course designer meant GTAs to teach in the mini-studio. To ensure that both instructor and students were satisfied with the class, and that the class was taught in the most effective manner, I decided that the most appropriate course of action would be to adjust the students' desires so that they want to be taught how the class is meant to be taught. Literature has shown that if instructors introduce class activities by discussing the purpose of the activity and the benefits of engaging with it (i.e. strategies for gaining student agreement), students may exhibit less resistance to these interactive activities (Shekhar et al., 2015). Although instructors may intend to implement research-based instructional strategies (RBISs), student resistance may cause them to fall back to the more traditional teaching methods that they think the students may desire (Henderson et al., 2012).

One of the overarching goals of my research is to learn how to reduce student resistance. I believe that instructor efforts to gain student agreement could significantly reduce student resistance. By default, without the instructor trying to gain student agreement, students would probably base their agreement on their experiences of the benefits or drawbacks of reformed instruction as they take the class. However, instructors should not wait for students to discover on their own if they agree with the class format because while students are experiencing the class and deciding on their agreement, they may be resisting the reformed instruction. Thus, it is up to the instructor to adjust student expectations and agreement early in the semester so resistance never occurs.

2.2 Instructional Communication Framework

The type of research I am doing is best described by the instructional communication framework. Instructional communication research is the intersection of research in pedagogy, educational psychology, and communication (see Figure 2). Research in pedagogy studies teaching and teaching practices (Mottet, Richmond, & McCroskey, 2006). For example, in a meta-analysis of science pedagogy research over 25 years, Schroeder et al. (2007) classified studies based on the instructional strategy, or pedagogy, under investigation. These pedagogies or teaching practices include question strategies, focusing strategies, and direct instruction. Studies in pedagogy also examine teaching strategies to motivate students. For example, a series of studies looked at the effect of content relevance to student motivation and found that frequently making the content relevant was positively correlated with student motivation (Frymier & Houser, 1998; Frymier & Shulman, 1995; Frymier, Shulman, & Houser, 1996).

Research in educational psychology "investigates the underlying psychological and intellectual processes that explain and predict student learning." (Mottet et al., 2006, p. 7) Bloom identified three domains of educational psychology: cognitive, behavioral, and affective learning. Cognitive learning focuses on how students understand and use learning. It is for the cognitive learning domain that Bloom created his famous taxonomy (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956). Behavioral learning focuses on developing physical skills. Affective learning focuses on "students' attitudes, beliefs, values, and underlying emotions or feelings as they relate to the knowledge and skills they are acquiring." (Mottet et al., 2006).

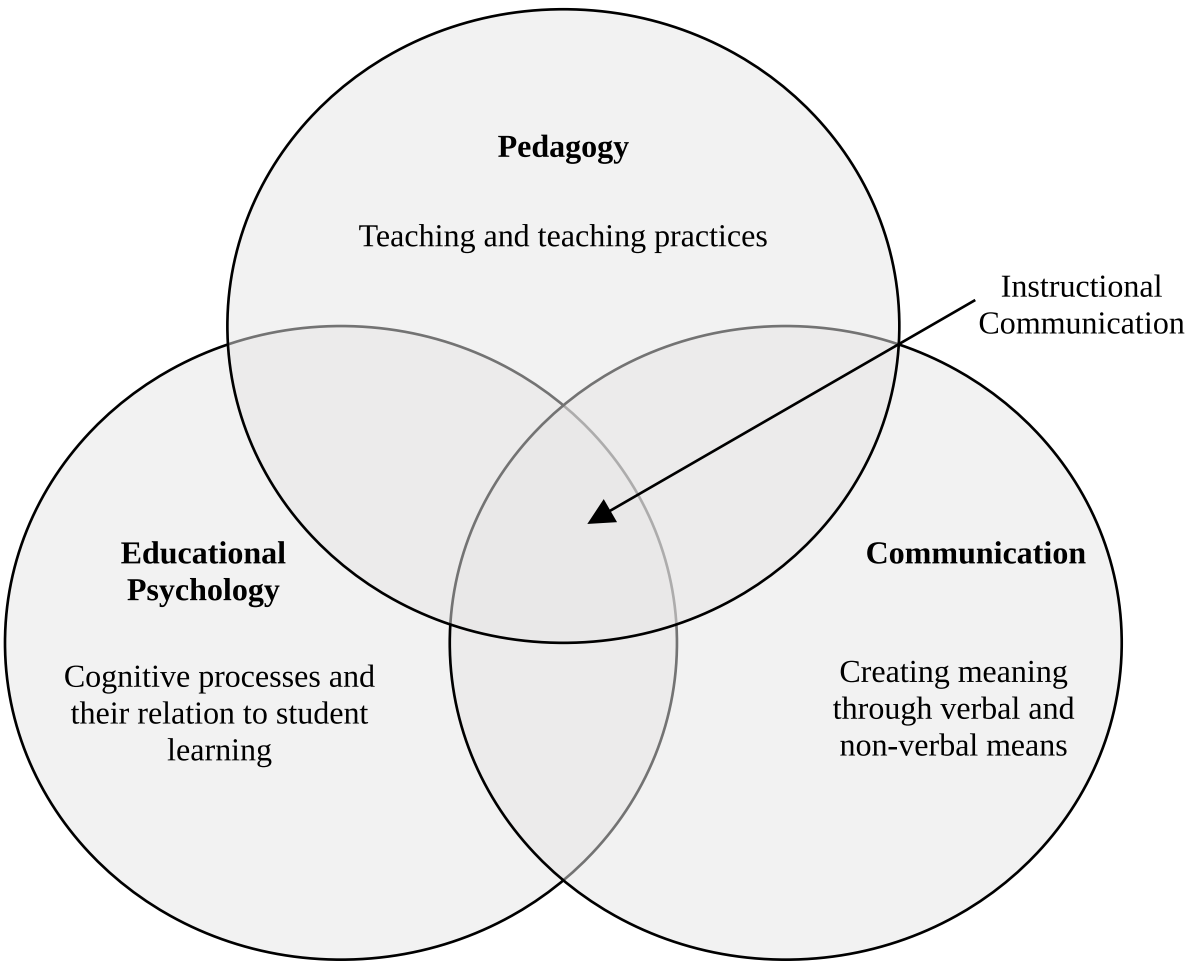


Figure 2. A diagram of the instructional communication framework and the research fields that comprise it.

McCroskey defined communication as the process of creating meaning for others through verbal and non-verbal means (1968). Research in communication studies how exactly this is done. In terms of the teacher-student dynamic, communication may be one-way, an interaction, or a transaction. In one-way communication, information flows from the source to the receiver, or teacher to student. In the communication as an interaction model, the information from the source may be altered based on the feedback from the receiver (Mottet et al., 2006, pp. 12-13). For example, a teacher may rephrase her explanation (the information) if she notices puzzled looks among her students (the feedback). Communication as a transaction occurs when two parties are simultaneously sharing responsibilities as the source and receiver of information (Mottet et al., 2006, pp. 12-14). In the classroom context, one might see communication as a transaction when both teacher and student are figuring out how to experimentally test a hypothesis. They must respect each other's ideas and build off of them until they come to a shared understanding.

One of the major goals of instructional communication is to study the communication behaviors that lead to effective teaching (L. McCroskey, Richmond, & McCroskey, 2002). From 2000 to 2016, the most common topic for research in instructional communication was affective learning and its relation to teaching effectiveness (Conley & Ah Yun, 2017). Krathwohl defined the affective domain of learning as "The realm which emphasized a feeling of tone, an emotion, or a degree of acceptance or rejection" (1964, p. 7). He and his colleagues developed a taxonomy of affective response which, beginning at the lowest level, are: receiving, responding, valuing, organizing, and value complex (see Table 1). At the lowest level of the affective domain is receiving or attending. At this level, a student is aware or willing to attend to phenomena. The level above receiving is responding and it differs slightly from receiving. At the responding level, students are motivated enough to actively attend to phenomena instead of simply being aware of it (Krathwohl, Bloom, & Masia, 1964).

Table 1. Levels of the affective domain and sub-levels within the "valuing" level.

Affective Domain Level		Description
5	Value Complex	The value is a characteristic of the student.
4	Organization	Student creates a system to organize new and previously held values.
3	Valuing	Student finds worth in the phenomena.
2	Responding	Student actively attends to the phenomena.
1	Receiving or Attending	Student is aware or willing to attend to the phenomena.

Value Level		Description
3	Commitment to Value	Students are certain in their belief.
2	Preference for a Value	Student pursues the belief.
1	Acceptance of Value	Student tentatively holds a belief.

The third level in the affective domain taxonomy is valuing. This is the realm of the affective domain that I am primarily interested in. At the valuing level, students find worth in the phenomena. Value is further categorized into three sub-levels: acceptance of value, preference for a value, and commitment. At the lowest level of value, acceptance of value, the student holds a belief, but it is tentative. The succeeding levels of value vary in the certainty a student has of their belief or value. A student at the next level, preference for a value, is committed and pursues that value. At the highest level of value, commitment, students are certain in their belief. Krathwohl et al. state that people at this level will seek to further the thing they value and increase their involvement with it (1964).

Students with higher levels of agreement with their class format may be found in this level, the valuing level, of the affective domain. They may agree with the class format, but it could be tenuous; perhaps agreeing with a few actions only. Students with stronger agreement may exhibit a preference for the class format in general; agreeing with most or all actions. Students with a commitment to the values of the class format likely agree with all actions. Further, these students might encourage others to agree with the class format as well. My measure of student agreement with the class format gives me a sense of where students may lie in the value level of the affective domain, if they are at that level. I can get a sense of whether their belief in the class format is tenuous, preferred, or if they exhibit a commitment to it based on the degree to which they agree and if they maintain that agreement over time.

The fourth level of the affective domain is organization. At the organization level, a student creates a system to organize new and previously held values that relate to each other and determines which are dominant. The final level of the affective domain is referred to as value complex. At this level, the value becomes a characteristic of a person, defining their behaviors. A value or system of values is completely integrated into the person's worldview (Krathwohl et al., 1964).

This description of the affective domain taxonomy supports my view on student resistance and the need to reduce student resistance to successfully implement RBISs. Students who reach the upper end of the value level regarding the class format will prefer and even seek out the class format. Since students at these levels will value the class format to the extent that they desire it, they should not be resistant to it. Thus, one of the goals of this research is to determine how instructors attempt to bring students to this level of the affective domain. By following this taxonomy, instructors must first make students aware of the class format, get them motivated to consider the class format, and then make strides to get students to agree with that. How exactly this is done is the overarching goal of my dissertation.

Like instructional communication, my research lies at the confluence of research in pedagogy, educational psychology, and communication. My interest is in the pedagogy and communication needed to influence the affective domain of student learning. Specifically, I want to study how instructors might modify student agreement with the class format so that they may become more effective teachers.

The methodology I use for my research aligns with the most common methods used in instructional communication research over the last 15 years (see Table 2). Surveys were the most frequent data collection method observed from the meta-analysis completed by Conley and Ah Yun (2017). Likewise, the majority of my data will come from two surveys. I will also support my findings with interviews which make up a small proportion of the observed data collection methods in the meta-analysis. College students and instructors were the two most common participants in instructional communication research over the last 15 years and are the two groups in my study. My research will deviate from the norm when it comes to the conceptual framework used. Most commonly, recent instructional communication research used grounded theory or expectancy violation theory. However, I use the less common rhetorical theory framework to help me interpret the relationship between the variables in my data.

Table 2. A comparison between the research elements commonly used in instructional communication research and my research.

Research Element	Most Commonly Used in Instructional Communication Research (Conley & Ah Yun, 2017)	My Research
Data Collection Method	Survey	Surveys
		Interviews
Participants	College students and instructors	College students and instructors
Conceptual Framework	Grounded theory	Rhetorical theory

2.3 Rhetorical Theory Conceptual Framework

Rhetoric is communication with the intent to persuade. In classical rhetorical theory, as developed by Aristotle, effective rhetoric is comprised of ethos, pathos, and logos. Aristotle describes ethos as the character of the speaker (Translated by Kennedy, 1991). In a more modern interpretation and for the context of the teacher-student dynamic, ethos represents the credibility of the teacher as defined by their competence, character, and caring (J. C. McCroskey & Teven, 1999). Before an instructor can use any further persuasion techniques, he must first establish his credibility. When students perceive their instructors as credible, they are motivated to perform well academically (Frymier & Thompson, 1992; Martin, Chesebro, & Mottet, 1997). Pathos is an appeal to the emotions of the audience. Aristotle recognized that decisions are not entirely rational and that "we do not give the same judgment when grieved and rejoicing or when being friendly and hostile" (1991, p. 39). The final element of effective rhetoric is logos, which is an appeal to the logical or rational.

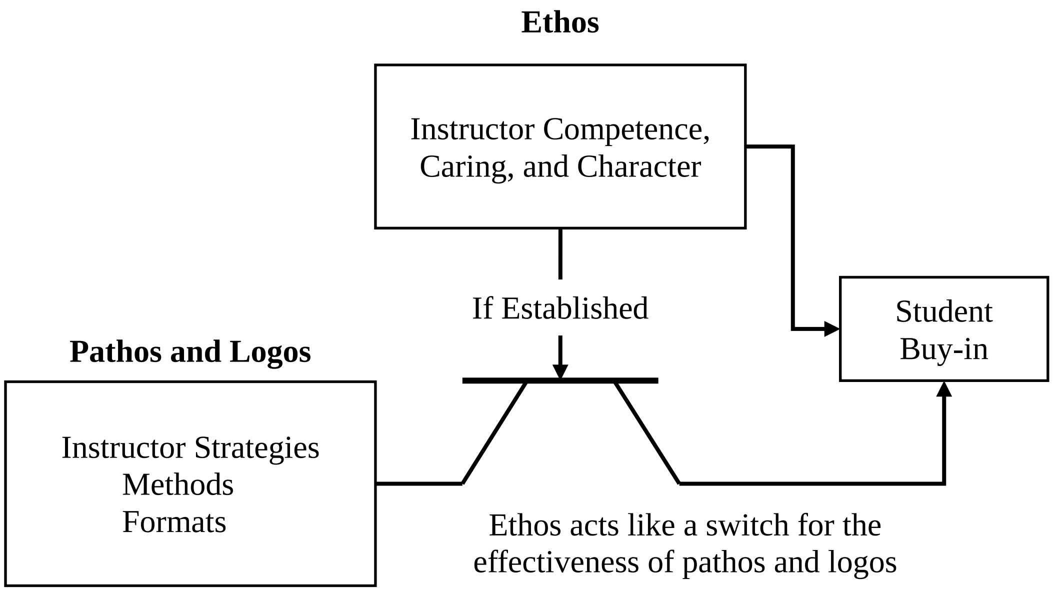


Figure 3. A model of the rhetorical theory conceptual framework in the context of my study.

In the context of my research, the rhetorical theory conceptual framework describes the relationship between the methods instructors might use to get student buy-in and the relationship between those methods and the level of student buy-in for their class, as shown in Figure 3. According to rhetorical theory, instructors must first establish themselves as credible by presenting themselves as someone who is competent, of good character, and caring towards the students. In doing so, instructors cover the ethos aspect of rhetoric. After establishing their credibility, instructors may try to appeal to the emotions of the students (pathos) or present logical arguments (logos) to persuade students to agree with the class format. How exactly instructors may do this is varied and a topic of this research. The success of these arguments will be measured by analyzing student buy-in or agreement levels with the methods or arguments instructors used in their class, as shown in Figure 4.

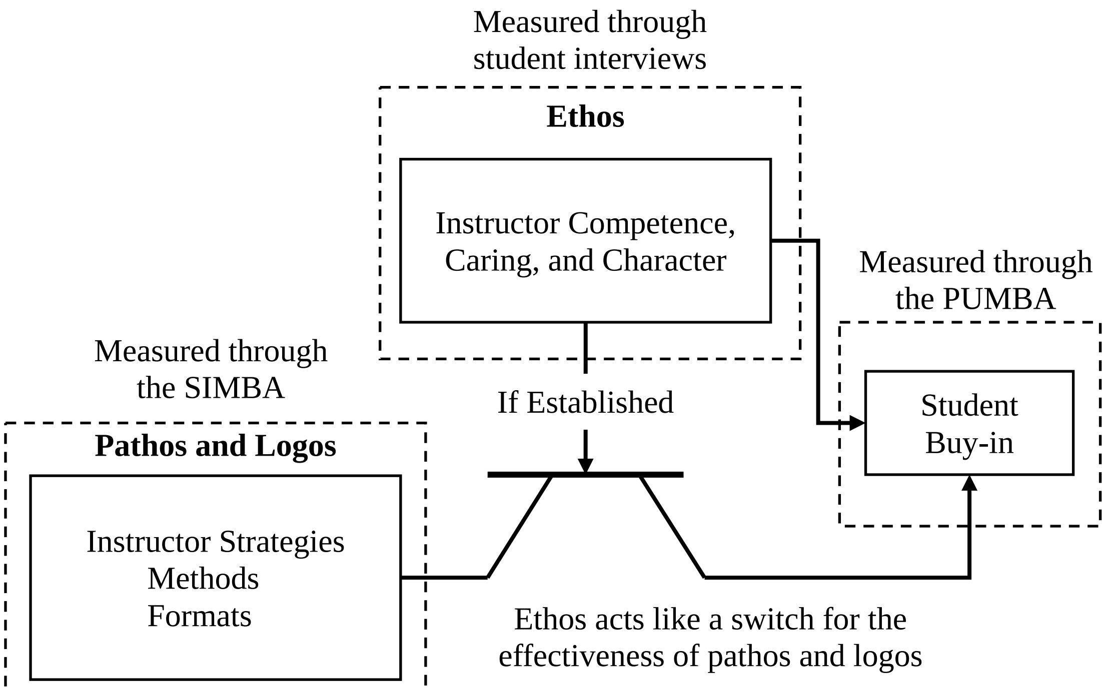


Figure 4. How I measure each component of the rhetorical theory framework.

As an example of looking at instructor methods to gain student buy-in with the rhetorical theory lens (see Table 3), imagine an instructor and her studio physics class on the first day of the semester. The instructor makes an effort to come off as a friendly and authoritative figure. She talks about her experience teaching studio physics and the success she has had. She also wants to establish that the classroom is a place for learning, which means sharing ideas which will sometimes be wrong. Her intent is to make the students comfortable to express their thoughts in the class. These strategies she is using are establishing her credibility (ethos) with the students by showing her character, competency, and caring for the class. She then asks students to think back to a class they have had that they performed well in. As a whole class, they write up aspects of these classes that made them better at teaching the students. She discusses how the studio class might do some of the same things as they had written up. While she is having this conversation, she might be relating the format of the class to how the students should expect a job in their field to operate. These methods are appealing to the students emotionally (pathos). She is making the students relate to the format either by highlighting the aspects of the class that students identify as effective practices or relating the format to their future job opportunities. The instructor may then follow up this discussion by showing a few graphs from education research showing the empirical benefits of group work. Doing so presents a logical argument to the students (logos).

Table 3. An example of rhetoric theory applied to instructor strategies for gaining student agreement.

Instructor Actions	Element of Rhetoric	Explanation
Presents herself as friendly and authoritative	Ethos	The instructor is coming off as caring and credible.
Discusses her experiences and successes teaching the studio class	Ethos	The instructor is providing evidence of her credibility.
Establishes that students should feel comfortable with sharing their thoughts.	Ethos and Pathos	The instructor is showing that she cares for the students' comfort and is appealing to their emotions by comforting them.
Asks students to reflect on the elements of a successful class and relates that to the studio class.	Pathos	The instructor is relating the class to the students' experiences.
Relates the class format to students' future job opportunities	Pathos	The instructor is making the class relevant to the students' interests.
Presents education research showing the benefits of group work	Logos	The instructor is making a claim based in factual evidence.

2.4 Expectancy Value Theory

Expectancy value theory states that an individual's effort in an activity is predicted by their expected success in the activity and their value of the activity (Eccles, 1983). Cooper, Ashley, and Brownell tested this theory on student resistance in an active learning classr to determine if student resistance decreases when: 1) students expect success in active learning; 2) students perceive value in participating in active learning; and 3) students perceive little cost in participating in active learning (2017). This framework, particularly the adaptation to student resistance by Cooper et al., is an excellent model of how student agreement to the class format would be beneficial in the studio class. The framework suggests that instructors should try to gain agreement by ensuring students expect success in the class, perceive little cost in participating in the class, and find value in participating. Indeed, Cooper et al. found that when they students they interviewed had positively aligned with this framework, they were less resistant to active learning. In my study, I investigate the strategies instructors can successfully use to ensure that their students positively align with this framework (i.e. agree with the class format) thus resulting in decreased student resistance.

2.5 Summary

I discussed these frameworks to communicate the ways in which I viewed the theory of my research, the methods, and the relationship between the measured variables. My research models the instructional communication framework; I am looking into the instructor strategies (pedagogy) to effectively communicate (communication) to their students why they should buy in (educational psychology) to the class format. On top of that, I am looking into what students buy in to and what that means for their learning. The rhetorical theory conceptual framework describes how I am relating the variables of my research, instructor strategies, and their connection to the outcome, student agreement. The conceptual framework supports the idea that instructor strategies can affect student buy-in to the class format because the strategies are aspects of rhetoric and student buy-in or agreement is essentially a measure of the success of that persuasive argument. Expectancy value theory then suggests that if students agree with the class format, they will be less resistant to the research-based instructional strategies implemented in the class.

CHAPTER THREE: LITERATURE REVIEW

In this chapter, I review the existing literature that is foundational to my research. To begin, I describe some different studio implementations, highlighting what makes them different from each other and from a traditional introductory physics class. In doing so, I describe how studio classes have incorporated reformed interactive instructional strategies that students might resist due to their departure from the traditional teaching style. The differences between the studio implementations highlight the varied approaches one can take to teach students in an interactive manner and how the different implementations evolved to balance student and institutional needs in a physics class. Next, I discuss the frequency of instructor usage of RBISs and review the literature that asserts student resistance is a barrier to the implementation of RBISs. I then review the surveys that have been made to explore aspects of student resistance through student expectations and desires. With these surveys, I will describe how they are insufficient to answer my research questions and necessitate the creation of new surveys. Finally, I describe a collection of materials and recommendations for instructors to frame their reformed class on the first day of class. I emphasize that although this collection is extremely useful, the materials are not empirically proven to be beneficial.

3.1 Studio Implementations

As defined earlier, studio physics classes combine the multiple elements of an introductory physics class into one setting. In general, studio classes place a large emphasis on student group work and interaction while limiting lecture from the instructor. In the following descriptions of studio implementations, a measure called "normalized gain" is used to quantify the success of a class. One calculates normalized gain from student scores on a conceptual test given at the beginning ("pre") and end ("post") of the semester. It is defined as a student's increase in score on the conceptual test between pre- and post-test divided by the maximum possible increase (see Equation 1) (Hake, 1998). For introductory mechanics courses, the most common concept test is the Force Concept Inventory, or FCI (Hestenes, Wells, & Swackhamer, 1992). For introductory electricity and magnetism courses, the concept test is often the Conceptual Survey of Electricity and Magnetism, or CSEM (Maloney, O'Kuma, Hieggelke, & Van Heuvelen, 2001). In traditionally taught introductory physics courses, Hake found the normalized gain to be 0.23 ± 0.04, placing these classes in the "low-g" or low-gain realm, defined by a normalized gain less that 0.3 (1998). Hake also studied the normalized gain of "interactive engagement" classes that make substantial use of methods that promote learning through interactive activities that provide immediate feedback for students from peers or instructors. In these classes, the normalized gain achieved was 0.48 ± 0.14. Most of these classes fell in the "medium-g" realm (normalized gain between 0.3 and 0.7) with some classified as "low-g". Hake define "high-g" as courses with a normalized gain greater than 0.7.

$$Normalized\ Gain = \frac{post\ test\% - pre\ test\%}{100\% - pre\ test\%} \quad (1)$$

The three studio implementations that I will discuss were chosen to exemplify common qualities and differences between studio methods. I begin with one of the first studio implementations, Workshop Physics, which I chose to highlight researchers' first efforts at reforming the physics class format and to provide a baseline studio implementation to compare others to. Next, I discuss the Collaborative Learning through Active Sense-making in Physics (CLASP) implementation as it is an example of how to teach a reformed physics class without completely eliminating lecture. Finally, I describe the Student Centered Active Learning Environment with Upside-down Pedagogies class because it is currently one of the most popular studio implementations. These are certainly not the only implementations one may use to teach a studio class. Other studio implementations include studio physics at the Rensselaer Polytechnic Institute (Wilson, 1994), Technology Enabled Active Learning (TEAL) at MIT (Breslow, 2010; Dori et al., 2003), and "New Studio" developed at Kansas State University (Sorensen, Churukian, Maleki, & Zollman, 2006).

3.1.1 Workshop Physics

Perhaps the first studio implementation is Workshop Physics created at Dickinson College and first implemented in 1986-87. The creators of Workshop Physics felt that the primary goal of traditional introductory physics courses was to solve physics problems. However, in their studio implementation, they would prefer to focus on students "acquiring transferable skills of scientific inquiry" (Laws, 1991, p. 25). The predominant activities in Workshop Physics are observations, experiments, and discussion. This studio implementation makes extensive use of computers for labs. Specifically, students use microcomputer-based laboratory programs to make calculations and graph data in real-time. Additionally, video software is used to analyze two-dimensional data, such as projectile motions. When collecting real data in class is not possible, simulations are used. These activities are time consuming so to make room for them, the developers chose to eliminate lectures and demonstrations as their effectiveness in teaching students the skills of scientific inquiry were unproven (Laws, 1991). Ultimately, Workshop Physics is an "activity-based, computer-assisted, guided-inquiry curricula" (Laws, Sokoloff, & Thornton, 1999, p. 16).

Initially, Workshop Physics classrooms were adapted laboratory rooms where students and instructors had access to computers and many other scientific devices (Laws, 1991). The rooms held up to 24 students and were staffed by undergraduate teaching assistants on evenings and weekends in addition to normal working hours. The course took place for six hours per week. The developers of Workshop Physics recognized that this environment is not possible in all contexts. As a result, they adapted the pedagogy of Workshop Physics to be incorporated in lecture hall settings by making use of Interactive Lecture Demonstrations (Sokoloff & Thornton, 1997; Thornton, 1997), which combine microcomputer-based lab equipment with demonstrations done in front of a large class (Laws et al., 1999). The instruction of Workshop Physics is based entirely upon observations and experiment. Students make observations of a phenomenon, reflect on it and discuss it with their peers, and develop theories that are then tested via experiment. The instructor's job is to facilitate this process rather than lecture. Students still complete problem solving activities, but unlike traditional physics classes, it is not the focus (Laws, 1991).

Workshop Physics has been successful in increasing students' conceptual knowledge of mechanics. At Dickinson College, where Workshop Physics was developed, students achieved an average normalized gain of 0.40 (Saul & Redish, 1997) which is considered a medium gain and is nearly twice the normalized gain that Hake found for traditional courses (1998). Similarly, when Workshop Physics was implemented at the University of Maryland in addition to their traditional classes, the studio implementation achieved a normalized gain of 0.41 whereas the traditional method achieved a gain of 0.20 (Saul & Redish, 1997).

In the same study, Saul and Redish, found that student attitudes about physics, as measured by the Maryland Physics Expectations (MPEX; see section 3.3.1) survey, remained consistent between the beginning and the end of the semester (Saul & Redish, 1997). This is an improvement over traditional courses, which showed a significant decrease in student expectations over the course. Through student interviews, Saul and Redish found that students believe Workshop Physics helped them learn useful knowledge or skills that they could apply beyond their class. However, the majority of students interviewed believed that a traditional style class would have been a more effective class. Saul and Redish proposed that this might be due to uncooperative groups, saying "When students feel that their group works well, they tend to feel Workshop Physics is effective. If the group has difficulties, students tend to believe they would learn more in traditional instruction." (1997, p. 20)

3.1.2 Collaborative Learning through Active Sense-making in Physics

In the 1996-97 academic year, the University of California, Davis completely replaced their traditional physics course for biology science majors with the Collaborative Learning through Active Sense-making in Physics (CLASP) curriculum (Potter et al., 2014). Like the developers of Workshop Physics, the creators of CLASP wanted to take the focus of traditional physics instruction away from calculations and rote memorization. Instead, they would prefer discussion and argumentation, a focus on concepts, and sense-making. The primary goal of CLASP is to create an environment where students develop their knowledge through discussion and argumentations via a large amount of group work while minimizing lecture. Note that lecture isn't eliminated like it is in Workshop Physics, but only minimized. The type of group activity in CLASP also differs from Workshop Physics. Where Workshop Physics focuses on lab activities to encourage discussion, CLASP group work is open to problem solving, labs, and responding to prompts (via a workbook) which guide students' conversations.

A typical CLASP section includes 25-30 students working at tables of five (Potter et al., 2014). The developers of CLASP find this number of students to be ideal because it is larger enough for productive class discussions and a larger student enrollment beyond 30 students would mean some students did not receive adequate attention from the instructor. The developers of CLASP considered doubling the number of groups and including a second instructor, but found that whole class discussion was not as productive and the class would “feel chaotic” (Potter et al., 2014, pp. 8-9). Students meet twice per week for discussion and laboratory and once per week for a lecture section. This separate lecture meeting, which makes up 22% of the overall class time, is one of the major differences between CLASP and all other studio implementations. The lecture section in the CLASP curriculum is typically used for the first introduction of material to the students. Lecturers do not need to work out problems as this task is something done as collaborative group work during the discussion/lab sections. A typical meeting in the discussion/lab section consists of cycles of group work followed by a short whole class discussion; the cycles culminate in a closing whole class discussion where the instructor ensures that students are leaving the class with the appropriate knowledge.

As measured by the FCI, the CLASP course improves students’ conceptual knowledge more than a traditional course would. Students taking a CLASP course achieved a normalized gain of 0.39 (Potter et al., 2014), similar to the normalized gains for Workshop Physics. Since CLASP was designed for biology students, who often go on to take the Medical College Admissions Tests (MCAT), the developers wanted to make sure that the CLASP course was not hindering their performance on the MCAT. They found that there is no statistical difference in the MCAT scores of students who take the CLASP course and students who had taken the traditional course. Like Workshop Physics, student attitudes about physics, as measured by the MPEX, remained unchanged over the semester (Potter et al., 2014). This is an improvement upon traditional courses where student attitudes typically worsen as the semester goes on.

3.1.3 Student Centered Active Learning Environment with Upside-down Pedagogies

The final studio implementation that I will describe in detail is SCALE-UP, or Student Centered Active Learning Environment with Upside-down Pedagogies. In its infancy, SCALE-UP was the Integrated Math, Physics, Engineering, and Chemistry Curriculum (IMPEC) at North Carolina State University (Beichner et al., 1999). Like most studio implementations, the IMPEC was taught in a single room and students were organized into groups of three. Unlike Workshop Physics that eliminated lecture, or CLASP that included lecture as a separate meeting, lecture was incorporated into the IMPEC classes but kept to a minimum. Much of the class time was spent on making predictions, developing models, collecting and analyzing data using the microcomputer-based lab equipment, and working on semester-long design projects. IMPEC incorporated many hands-on physics activities including those from Workshop Physics. Labs were included in the curriculum but were meant to be short activities directly related to the material being discussed. IMPEC was shown to be extremely successful in increasing students' conceptual knowledge. In the Spring semester of 1996, IMPEC students achieved a normalized gain of 0.42 and in the Spring semester of 1997 they achieved a normalized gain of 0.55. These normalized gains were a statistically significant improvement over the normalized gain of 0.23 for traditional classes found by Hake.

Prior to the first implementation of SCALE-UP, studio classes were taught to up to 36 students at a time which is small compared to the hundreds of students a traditional lecture class can teach. Workshop Physics was designed for up to 24 students (Laws, 1991), CLASP had 25-30 students in a section (Potter et al., 2014), and the IMPEC handled 36 students (Beichner et al., 2007). For North Carolina State University and other large universities, such small class designs are not economically feasible because they require too many teachers and classrooms. Thus, Beichner et al. sought to extended, or “scale up”, their successful IMPEC to a class size of roughly 100 students, resulting in the first iteration of SCALE-UP with the name Student-Centered Activities for Large Enrollment University Physics. When creating SCALE-UP, the developers wanted to adopt numerous research-based pedagogies, such as the University of Washington’s *Tutorials in Introductory Physics* (McDermott & Shaffer, 1998), Workshop Physics (Laws, 1991), and Minnesota’s *Cooperative Group Problem Solving* (P. Heller & Hollabaugh, 1992; P. Heller, Keith, & Anderson, 1997), into their larger classroom setting. Many of the aspects of IMPEC were retained in SCALE-UP, including an emphasis on group work, the incorporation of technology, and short hands-on activities which relate directly to the topic being discussed (Beichner, Saul, Allain, Deardorff, & Abbott, 2000). Like the other studio implementations discussed and a continuation from the IMPEC, lecture is de-emphasized. In fact, in the SCALE-UP FAQ, it is stated that “If you are lecturing for more than 15 minutes, you are probably talking too much” (Beichner, 2011).

By 2007, SCALE-UP, which at that point was called the Student-Centered Activities for Large Enrollment Undergraduate Programs (no longer excluding everything but physics), had become an established studio implementation. In this iteration, the developers of SCALE-UP introduced its reformed studio classroom environment that SCALE-UP is well known for. Some even argue that it is the environment that makes SCALE-UP successful as a sustained studio implementation (Foote, Neumeyer, Henderson, Dancy, & Beichner, 2014). The SCALE-UP environment utilizes 6 or 7 feet diameter round tables (Beichner et al., 2007). They found this size to be just right to fit three groups of three students each at one table without being too large to discourage interaction. Each group at these tables has access to a computer at all times. The rooms contain all the lab equipment needed for experiments and the tables are large enough to handle the experiments. Numerous projectors and screens allow for work and slides to be seen by students sitting anywhere in the class. The final essential element of the SCALE-UP environment is means for students to present their work to the rest of the class. Typically, this is done by having whiteboards along the walls of the room and/or portable whiteboards for each group.

Currently, SCALE-UP stands for Student Centered Active Learning Environment with Upside-down Pedagogies, applying the common term "upside-down pedagogy" to the acronym to highlight that element of the studio implementation method. It has become one of the more common studio style implementations. As of 2014, SCALE-UP was found in more than 300 departments at 189 institutions in 21 countries. Over one-third of those departments are physics (Foote et al., 2014).

The success of SCALE-UP follows the success of the IMPEC. Over a five-year period at UCF, SCALE-UP classes achieved a normalized gain on the FCI of roughly 0.40. Over this same period, traditional introductory physics courses at UCF achieved a normalized gain of 0.20, typical for the type of class (Beichner et al., 2007).

The literature shows that student reaction to SCALE-UP has been generally positive. One student said, "[SCALE-UP] helped me to learn not so much to get an answer but to actually understand concepts." Another student said, "I definitely feel that, compared to traditional, we have a more in depth understanding and knowledge of what is going on." (Beichner et al., 2000, p. 8). Additionally, class evaluations of SCALE-UP are typically positive and even though students felt SCALE-UP is more work, roughly 2/3 of students felt that the extra work was worth the time because they learned more (Beichner et al., 2007).

There are no published attitudinal scores for SCALE-UP from scales like the MPEX. In a recent meta-analysis, Madsen, McKagan, and Sayre collected physics students' attitudinal shifts as reported in 24 studies (Madsen, McKagan, & Sayre, 2015). Many of the classes in these studies were non-traditional, using reformed strategies, focusing on developing expert-like beliefs, or model building. The SCALE-UP studio implementation did not appear in this meta-analysis. However, SCALE-UP is the implementation used for the studio classes at UCF. As measured by student responses to the CLASS (Colorado Learning Attitudes about Science Survey) which measures student attitudes (Adams et al., 2006) and is similar to the MPEX, SCALE-UP students' attitudes about science decreased five percentage points across the semester. In the traditional introductory physics courses taught at UCF, over the same time period (Fall 2011 to Spring 2018) students' attitudes about science decreased six percentage points. As evidenced by these attitudinal shifts at UCF, students in SCALE-UP have a similar decrease in attitudes as students in traditional courses.

3.1.4 Comparing Studio Implementations

Each of the described studio implementations, Workshop Physics, CLASP, and SCALE-UP, share some commonalities and have some differences; see Table 4 for a summary. They all have the primary focus of students working and interacting together while taking the focus away from an instructor lecturing. These studio implementations strive for discussion and argumentation between students as the means for learning. There is also a strong connection to technology in each of the implementations. Workshop Physics makes heavy use of microcomputer-based labs, which are also incorporated into the CLASP and SCALE-UP curriculum. Students still complete problem-solving activities, but this is done at home and/or in groups during class. This "upside-down" pedagogy is so important to SCALE-UP that it even changed its name to reflect that. One final commonality between the three studio implementations is their success in increasing student conceptual knowledge as measured by the FCI. The published normalized gains of these studio implementations ranged from 0.39 to 0.55 which is a significant improvement over the typical traditional class normalized gain of 0.23.

Table 4. A summary of three studio implementations.

Studio Implementation	FCI Normalized Gain	Change in Attitudes Measured by MPEX	Student Response	Class Size	Features
Workshop Physics	0.40-0.41	Remained consistent.	Workshop Physics helped students learn useful skills but they felt groups were uncooperative.	24	No lecture; Lab-based group work
CLASP	0.39	Remained consistent.	Not mentioned in literature.	25-30	Lecture meeting separate from studio; Focus on group work via problem solving, labs, or guided prompts (workbooks) to encourage discussion and argumentation; Periods of whole class discussion throughout the studio meetings
SCALE-UP	0.40-0.55	Decreased over the semester.	Students felt SCALE-UP was more work but worth the extra effort because they learned more.	99	Lecture is minimized; Emphasis on group work, incorporation of technology, and short hands-on activities.

There are several differences among the studio implementations as well. Workshop Physics and CLASP were designed for up to 30 students whereas SCALE-UP is meant for up to 100 students. While all studio implementations seek to minimize lecture, each does so in a different way. Workshop Physics cuts lecture out of the curriculum, CLASP puts lecture in a different meeting time, and SCALE-UP simply encourages instructors to lecture as little as they can. There are also differences in some of the emphasized aspects of the pedagogy. Workshop Physics is largely focused towards observation and experiment leading to discussion. CLASP places a large emphasis on discussion and argumentation of physics concepts between students and the whole class which is spurred on by experiment or guided prompts. SCALE-UP uses a mixture of multiple PER based pedagogies not necessarily emphasizing one specific aspect.

3.2 Barriers to the Use of Research Based Instructional Strategies

3.2.1 Barriers Perceived by Physics Faculty

These studio implementations are intended to make use of many PER based pedagogies or RBISs. However, using these strategies and continuing to use them may be easier said than done. The decision to use RBISs and continue using them can be modeled by the "innovation-decision process" (Rogers, 1995) as shown in Table 5. The first two steps in the innovation-decision process are learning about RBISs and being persuaded of the benefits of RBISs. Note that these first two steps are essentially instructor buy-in to the class format! The final three steps in the process are the decision to use RBISs, implementing them, and then continuing to use them. Henderson and his colleagues (Henderson et al., 2012) surveyed just over 700 physics faculty in the United States to see where they were in the innovation-decision process for 24 RBISs including Context-Rich Problems, Interactive Lecture Demonstrations, Open Source tutorials, and Peer Instruction. Of the approximately 700 respondents, 12% had no knowledge of any RBISs, and 16% knew about at least one RBIS but had never used any. From the physics faculty that had tried at least one RBIS, roughly a third continued to use three or more RBISs, another third continued to use one or two RBISs, and the final third discontinued use. Only half of the physics faculty surveyed were continuing to use RBISs at the time, and, of those faculty, less than half (23% of the total population surveyed) used them to a high degree. Henderson et al., identified the final step of the innovation-decision process as the most "lossy" stating that "work is needed to understand and address the 1/3 of faculty who discontinue use after trying." (Henderson et al., 2012, p. 10).

Table 5. The innovation-decision process applied to instructors using RBISs.

Step in Innovation-Decision Process	Description
Knowledge	Instructor becomes aware of RBISs
Persuasion	Instructor forms a favorable attitude toward RBISs
Decision	Instructor engages in activities in which they can choose to use RBISs
Implementation	Instructor uses RBISs
Confirmation	Instructor seeks reinforcement of the decision to use RBISs

There is a large amount of research investigating the barriers to implementation of RBISs. Many of these studies find that the instructor of the class cites student resistance as a barrier to using their reformed pedagogies. For example, Dancy and Henderson interviewed four physics faculty who they expected would be the kind of faculty to easily adopt PER pedagogies. Although they did not state how many of the four expressed this view, they identified student resistance as one of the most common systemic forces affecting the instructors' practices. One instructor states (Dancy & Henderson, 2005, p. 3),

"What I want to do is to turn the class into a real working session. Where it's just not possible for them to come there and sleep. That may turn off

students and decrease enrollment, they may switch courses. I'm a little worried about attrition. That's another aspect."

This instructor appears to be experiencing student resistance in the form of sleeping during class. He also expresses concern that using an RBIS, such as creating a "working session", would be viewed negatively by the students and possibly lead them to drop the class. Because of these factors of student resistance, the instructor appears hesitant to implement the RBIS.

The issue of student resistance persisted in another interview study with physics faculty. Out of five instructors, three mentioned students not supporting research-based methods. Specifically, students preferred not to interact with each other yet they also came to class unprepared to think independently (Henderson & Dancy, 2007). Later on, Turpen, along with Dancy and Henderson, performed more interviews with physics faculty (N=35) to determine perceived affordances and constraints with their use of Peer Instruction, a RBIS developed by Mazur (1999). They found that 29% of the faculty interviewed discussed difficulty getting student buy-in. For example, one instructor said (Turpen, Dancy, & Henderson, 2016, p. 8),

"I would say occasionally there is student resistance to it. You know, some classes I'll have maybe one person who simply insists that they're not there to do that, so that can be problematic."

Another study by Dancy and Henderson finds that a small proportion of physics faculty surveyed self-report student resistance as a reason for not using more RBISs. Out of roughly 460 physics faculty that discussed what prevents them from using more RBISs, 2.1% reported student resistance as a barrier (Dancy & Henderson, 2010). While this is a small proportion, it still shows that student resistance is affecting instructors. This proportion is also likely a low estimate of the actual proportion of instructors that experience student resistance as a barrier to their teaching. In their survey, this was an essay style question so instructors are likely to only write about the most salient barriers, as evidenced by the two most common responses to this question: lack of time and lack of familiarity with RBISs. Had all instructors been presented with student resistance as a possible barrier, then they may have found a larger proportion of faculty indicating such.

Others have recognized student resistance as one of the difficulties in implementing the reformed pedagogies that define studio classes. In a newsletter to the faculty, Belcher, one of the developers of TEAL at MIT, a studio implementation that merges lecture, problem solving, and labs in two two-hour sessions plus a one-hour problem-solving session (Breslow, 2010), discussed some of the major missteps in moving TEAL from a prototype class to an official, large-scale studio implementation. Among the missteps was the lack of training for students to work collaboratively. Students did not understand the purpose of group work, and so some groups did not work well together (Belcher, 2003). Students described group work as "at best annoying and at worst counterproductive." (Breslow, 2010, p. 25) Considering that group work is a primary goal in many studio implementations, this problem is important to remedy. As described in the newsletter, student resistance to group work appeared because they did not understand its purpose. Achieving student buy-in to group work solves this problem because by definition buy-in is understanding and agreeing with why the class format is what it is. Belcher also describes a similar issue in regard to experiments in TEAL. He states that students did not find them useful, which is another buy-in related problem.

Yerushalmi et al. describe how perceived student resistance can affect physics faculty's practices. They interviewed six physics faculty to determine their teaching values, perceived constraints, and how these affect their instruction (Yerushalmi, Henderson, Heller, Heller, & Kuo, 2007). They find that some of their teaching values led to practices not necessarily aligned with research-based practices. For example, the instructors valued clear and efficient communication of physics. As a result, instructors avoided wordy problems and problems placed in unrealistic contexts. Their value of efficiency and clarity also meant that the feedback they provided students when grading was very short; any solutions given to their students reflected an expert-like solution that may be difficult for the students to understand.

The instructors' perceived limitations to their instruction also influenced their in-class instructional practices; specifically, instructors tried to meet their perceived student expectations. So, the instructors would consider if a problem would yield an easy analysis for students or if a solution was too long or complicated for fear of repelling students. The instructors appeared to be weary of challenging the students, especially wanting to avoid a confrontation with the good students. The authors suggest that it was because of these perceived constraints that the instructors' practices ultimately reflected practices that met the students' desires instead of the students' needs (Yerushalmi et al., 2007).

3.2.2 Perceptions of Student Resistance Beyond Physics Faculty

The problem of student resistance is not exclusive to physics nor faculty. One study of physics GTAs and Learning Assistants (LAs; essentially undergraduate TAs) in SCALE-UP classes found student resistance to be an impediment for them as well (DeBeck, Settelmeyer, Li, & Demaree, 2010). Both GTAs and LAs found it difficult to involve students who did not buy in to the class activities. A problem for LAs, specifically, is that they may be forced to defend the SCALE-UP pedagogies in other classes they may share with the SCALE-UP student they teach. Thus, LAs would hesitate to use the pedagogies if they knew they would later have to defend it.

GTAs from other STEM (Science Technology Engineering and Mathematics) fields have experienced student resistance. Biology GTAs who increased discussion of the nature of science in their labs experienced student pushback (Bautista, Schussler, & Rybczynski, 2014). For example, one GTA mentioned that his students wanted to get in and out of the labs as quickly as possible. Another GTA said that he received negative reviews about the nature of science discussions. The authors state that most GTAs perceived that the students did not value the nature of science discussions. In other words, the GTAs felt that students did not buy in to those discussions. Additionally, a survey of engineering GTAs found that some of those who felt unsuccessful in their use of active learning cited student resistance as a reason why they were unsuccessful (Pinder-Grover, 2013). For example, when one GTA asked the students to work in groups on a more complicated question, the GTA said that the students "just sit there and don't do anything." (Pinder-Grover, 2013, p. 11)

Faculty in engineering also report student resistance as a barrier to the implementation of RBISs (Froyd, Borrego, Cutler, Henderson, & Prince, 2013). In a survey completed by 122 undergraduate engineering faculty, they were given a chance to describe barriers to each of 12 different RBISs. Out of 1464 possible opportunities (122 faculty * 12 RBISs) to cite a barrier, instructors cited student resistance nearly 200 times. One response said that students complained about attendance being more important when an RBIS was used. Another said that "stronger students pushed back" on implementing RBISs (Froyd et al., 2013, p. 397).

3.2.3 Summary

As summarized in Table 6, student resistance is an issue experienced by faculty, GTAs, and LAs across disciplines. In many cases, instructors specifically cited student resistance as a barrier to the implementation of RBISs. Some of these studies connected student resistance as a student buy-in issue, recognizing that to solve the problem students must understand and agree with why their instructor is using the RBISs (Bautista et al., 2014; Belcher, 2003; DeBeck et al., 2010; Turpen et al., 2016). Although there are many studies identifying this as an issue, there are relatively few on how to solve the problem.

Table 6. A summary of the studies citing student resistance as a barrier to reformed instruction.

Study	Population	Barrier	Outcomes from Barrier
Dancy and Henderson (2005)	Physics faculty	Student resistance to turning the class "into a real working session."	Decreased enrollment; students switching courses
Henderson and Dancy (2007)	Physics faculty	Student not interacting with each other and coming to class unprepared to think independently.	-
Turpen, Dancy, and Henderson (2016)	Physics faculty	Student resistance to Peer Instruction	Students insisting "they're not there to do that..."
Dandy and Henderson (2010)	Physics faculty	Student resistance in general	-
Belcher (2003) and Breslow (2010)	Physics students at MIT	Student resistance to group work	Students described group work as annoying and counterproductive; students did not work collaboratively
Yerushalmi et al. (2007)	Physics faculty	Instructor expectations of student resistance	Poorer feedback when grading; weariness to challenge students
DeBeck et al. (2010)	Physics GTAs and LAs	Low student buy-in	-
Bautista, Schussler, and Rybczynski (2014)	Biology GTAs	Student resistance to discussion of the nature of science	Students rushing through labs; negative reviews of class discussion
Pinder-Grover (2013)	Engineering GTAs	Student resistance in general	-
Froyd et al. (2013)	Engineering faculty	Student resistance in general	-

3.3 Previous Surveys Investigating Elements of Student Buy-in

Researchers have developed several surveys to explore elements of student buy-in and the ways in which instructors attempt to gain student buy-in. However, these surveys do not measure all elements of student buy-in, or not in a manner that suggests ways to improve student buy-in. In this section, I describe these surveys and some of their findings. I will also highlight how each survey is insufficient to answer my research questions, detailing specific deficiencies that are resolved in my two surveys.

3.3.1 Maryland Physics Expectations Survey

The first step in gaining student buy-in is getting students to understand what they should expect of the class. One of the first surveys developed to investigate student expectations in introductory physics courses is the Maryland Physics Expectations (MPEX) Survey (Redish, Saul, & Steinberg, 1998). The MPEX is a 34-item, Likert scale survey given during the first week of the semester and at the end of the semester. The items on the MPEX were chosen as items "that have an effect on how students interpret and process the physics in the class" (Redish et al., 1998, p. 4). The items cluster to probe six dimensions of learning physics: independence, coherence, concepts, reality link, math link, and effort. Students are given a prompt, such as "I read the text in detail and work through many of the examples given there" (which is part of the effort dimension) and respond on a five-point Likert scale from strongly disagree (1) to strongly agree (5). The percentage of responses in which a student responds similar to or different than an expert determines that student's score for the MPEX.

The developers of the MPEX first gave their survey at six institutions, three large flagship universities, a smaller university, a small liberal arts university, and a two-year college (Redish et al., 1998). In this distribution, they found that 50-60% of the student responses to the MPEX aligned with the expert response. At the end of the semester, all institutions showed a decrease in the proportion of responses that agreed with the expert response. Also, students reported feeling like they did not put in as much effort as they expected they would at the start of the semester. The number of responses that disagreed with the expert response increased from the beginning to the end of the semester. Considering that students could respond neutrally on the Likert scale ("neither disagree nor agree"), it could have been possible that student responses shifted from agreeing with an expert response to neutral. However, this was not the case; students' expectations and attitudes went from aligning with an expert to opposing an expert-like response. Similar results are found with the similar Colorado Learning Attitudes about Science Survey (CLASS) (Adams et al., 2006). The CLASS functions like the MPEX and also measures student attitudes about doing and learning science. The results of the CLASS are consistent with the MPEX in that, over a course of a semester, student attitudes become less aligned with an expert's attitudes.

The MPEX was one of the first surveys to look at students' attitudes towards introductory physics. However, it explores student expectations in a rather broad sense. It does not specifically measure student expectations of the class format but more so the class environment and learning culture. Certainly, these are two aspects that students should agree with, but this is not the type of buy-in that I am interested in. Another drawback for the MPEX as it applies to my research is that it does not ask for student preferences. A student may agree that they "read the text in detail and work through many of the examples given there", but they may not agree that doing so is the best way for them to learn. It is possible that a student may do that because they feel it is necessary to succeed in their course but ultimately is not the most effective way for them to learn physics. In such a case, the student would not buy-in to the amount of reading or problem solving required for the course. However, one cannot determine that with the MPEX because it does not question the student's preferences. So, two key features missing in the MPEX that are necessary in my surveys are: (1) asking about specific actions in class instead of the learning culture; and (2) asking about student preferences.

3.3.2 Pedagogical Expectation Violation Assessment

The Pedagogical Expectation Violation Assessment (PEVA) (Gaffney et al., 2010) measures students' expectations and experiences with a reformed physics class. Unlike the MPEX or the CLASS, which compare student expectations and attitudes to "expert" attitudes, the PEVA compares students' expectations of their class with their experiences of the class. Gaffney et al. were interested in student expectations of how the class was taught (e.g. frequency of lecture, group activities, reading, etc.), in addition to elements of the class environment or learning culture (e.g. how often missed classes would harm the students' learning, how frequently students will need to memorize equations). Gaffney et al. felt that differences between students' expectations and their experiences in a reformed class may make the class more difficult for students. When the difference between what the students expect and what they experience (called an expectancy violation) is perceived as negative, it can result in negative reviews of the class (McPherson, Kearney, & Plax, 2003). To determine how effective instructors were at aligning student expectations with the class format, students took the PEVA immediately upon entering the class and also after they had been oriented to the style of the class. The PEVA was also used at the end of the semester to measure their experiences so that this can be compared with their initial expectations.

The first version of the PEVA contained 13 base items that probed student expectations about course instruction (Gaffney et al., 2010). Many of these items are actions that happen for the class, such as lecture, group discussions, reading, and labs. Some of the items, like memorizing equations and the frequency of an "amphitheater-style classroom" still relate to the class environment or learning culture. Students rank the frequency of each of these items on a Likert scale for their class from very infrequently (1) to very frequently (7). The student responses to each item are averaged and compared to the averages for that same item at the different points in the semester.

The PEVA was given to students taking SCALE-UP physics courses at three universities (Gaffney et al., 2010). The results showed that students initially (before orientation) expected a largely traditional type of class. This is evidenced by the large expected frequency of lecture, reading, lecture hall style classroom, and having separate lab sections. The researchers also found that instructors, via an orientation, could successfully change student expectations to be more in line with a reformed class format. In terms of what students experienced compared to their expectations, there was a mixture of results. In some cases, students did experience what they expected after orientation. The authors referred to this as a success; this shift towards alignment was found for the "interacting with the instructor" action and the "discussing work with classmates" action. There were some items in which the instructor successfully shifted initial expectations in the right direction, but not to the extent that students would eventually experience. For example, initially, students expected "labs separate from class" to occur rather frequently, averaging 5.4 on the Likert scale. After orientation, students expected this action to occur slightly infrequently, averaging 3.0. Ultimately, students experienced labs separate from the class infrequently, with a reported average frequency ranking of 1.9. In other cases, the shift in expectation overshot what students would eventually experience. This occurred with lecture, where students initially expected it to occur frequently, ranking just over 6 on the Likert scale. After orientation, they expected it to occur less often at a 4.8 on the Likert scale. They experienced lecture to occur somewhere between these two expectations. The last possibility, which the authors called a failure, is when student experiences are opposite to their shifted expectations. This occurred with the item “Missed classes would be harmful to my learning”. Student expectations were shifted towards this occurring more frequently, meaning missed classes would be more harmful. However, they reported at the end of the semester that this happened less frequently than their initial expectations, meaning missed classes were less harmful than they initially expected.

In the first version of the PEVA, one could determine if a student’s expectations were violated and whether they experienced more or less of an action than what they expected. However, it is important to know whether or not a student perceived an expectancy violation as positive or negative. In a later design of the PEVA, added to each item was a question asking the student to rank their displeasure or pleasure with the amount of time spent on that item (Chini et al., 2013). With this version of the PEVA, researchers found that SCALE-UP students experienced a negative expectancy violation more often than students in a traditional physics course. Also, more frequent negative expectancy violations were negatively correlated with overall course satisfaction (Gaffney, Gaffney, & Chini, 2013). For example, if students more often experienced a negative expectancy violation regarding solving physics problems individually, they would have a lower overall satisfaction with the course. These studies are representative of instructional communication research. The PEVA surveys students’ expectations and expectation violations (cognitive processes) which are then related to their satisfaction with the class, findings representative of educational psychology research. They also found that student expectations shift after being oriented to the class indicating that there is some element of how the instructor communicated the class format to the students that affected their expectations, findings representative of pedagogy and communication research.

Overall, it appears that the PEVA has accomplished what it was made to do, investigate the effect of positive or negative expectancy violations on course satisfaction. However, some questions remain unexplored. When a student's expectations are negatively violated, what would they prefer? The PEVA inquires about student satisfaction of their experiences by the end of the semester, but how satisfied are students with their expectations at the beginning of the semester? Does this affect their performance in class? To answer these questions, my survey asks about student preference at the beginning and end of the semester. Because the research with PEVA demonstrated that student expectations are affected by orientation from the instructor, my research targets student expectations after this orientation.

3.3.3 Student Response to Instructional Practices Instrument

Recently, a large group of education researchers from across the country also recognized the need to investigate how students respond to reformed instruction (DeMonbrun et al., 2017). Like the motivation for my own research, they were driven by the common concern among instructors of student resistance inhibiting their use of RBISs. To investigate student resistance, these researchers created the Student Response to Instructional Practices (StRIP) instrument to measure student response to instruction and the strategies instructors used for their instructional practices. With the StRIP instrument, they investigated the relationship between student response to instruction and: (1) students' expectations of different types of instruction; (2) students' reported frequency of different types of instruction; and (3) the instructor strategies used (Nguyen et al., 2017).

The StRIP instrument has four sections that measure the types of instruction used, the instructor strategies for using in-class activities, student response to instruction, and student characteristics (Nguyen et al., 2017). There are several subscales within each measure, as shown in Table 7.

Table 7. Scales and subscales measured with the StRIP instrument.

Scale	Subscales
Types of Instruction	Passive Lecture
	Active Lerning lecture
	Group Based Activities
	Self-Directed Activities
Instructor Strategies for Using In-Class Activities	Explanation
	Facilitation
Student Responses to Instruction	Value
	Positivity
	Participation
	Evaluation
Student Characteristics	Expected Grade
	Prior Experiences with Active Learning

The StRIP survey is taken at three points in the semester: the start of the semester, two weeks into the semester, and the end of the semester (Nguyen et al., 2017). At each instance, student response to instruction is measured. The first two times the students complete the StRIP survey, they are asked about the types of instruction they *expect*. When taking the survey at the end of the semester, students respond about the types of instruction they *experienced*. Additionally, students report the instructor strategies for using in-class activities at the end of the semester. For the “Student Response” and “Instructor Strategies for Using In-Class Activities” measures, students respond on a five-point Likert scale to indicate the frequency with which each item occurred, from less than 10% to greater than 90%, with 20% point jumps between possible responses (Nguyen et al., 2016). When responding to the frequency of instructional practices, students' responses corresponded to a rate of instances per semester or week, from "never" to "more than once a week". The instrument also asked students how often these practices would occur in their ideal course. Again, they responded on a Likert scale from "Much less" (1) to "Much more" (5).

To determine how expectations, experiences, instructor strategies, and student characteristics affect student response to instruction, Nguyen et al. gave the StRIP to engineering students in four universities in Spring 2015 (Nguyen et al., 2017). They created different models to predict the outcome of each factor of student response: value, positivity, participation, and evaluation. The final models are a result of iteratively adding predictors (expectations or experiences of instruction, instructor strategies, and student characteristics) from the different administrations of the survey. Table 8 summarizes the results of these final models, which can be found in Table 10 of Nguyen et al. (2017). Overall, only a few expectations or experiences of the course instruction affected student response to the course. Expectations of active learning lecture negatively impacted student response, while expectations of passive lecture positively affected student response. Although this appears to paint a picture of students favoring the traditional lecture style, students did respond positively to experiencing group based activities. Instructor strategies for using in-class activities, specifically explaining the purpose of the activities, were common significant predictors of positive student response. Explaining the purpose of the activities increased student value, positivity, and evaluation. Also, when the instructor assisted in facilitating these activities, student positivity increased.

Table 8. A summary of significant predictors to student response to instruction as measured by the StRIP survey (Nguyen et al., 2017).

Student Response Factor	Significant Predictors
Value	Expectations of passive lecture two weeks into semester
	Experiences of group based activities at end of semester
	Instructor explanation strategies
Positivity	Experiences of group based activities at end of semester
	Instructor explanation strategies
	Instructor facilitation strategies
Participation	None
Evaluation	Expectations of active learning lecture at start of semester*
	Instructor explanation strategies

Most relationships are positive, meaning a higher value for the predictor corresponds to a higher value for the student response factor. Only "expectations of active learning lecture at start of semester" (indicated by "*") was a negative predictor.

The development of the StRIP survey is a large step forward into the investigation of student resistance to reformed instruction. Although the StRIP survey does appear to measure buy-in in some sense, it does so in a rather broad way. The factors of student response apply to the class as a whole instead of to specific activities in the class. With my research, I investigated student buy-in both overall and to specific activities and aspects of a class. Additionally, the instructor strategies for using activities are also not specific to certain activities, only in general. The instructor strategies for using in-class activities group into either explaining or facilitating; the StRIP survey does not have the capability to determine specifics about these strategies and how they relate to student response. In my research, I explored the specific topics of discussion and the means of having the discussion to investigate the most effective ways an instructor can gain student buy-in. Lastly, although they have the capability to do so, the developers of the StRIP survey have not published on student preference when analyzing student response. Student preference is integral to the calculation of student buy-in and is discussed in my research.

3.3.4 First Day Framing Materials

The previous surveys discussed in this chapter address student attitudes, expectations, and responses to reformed instruction. Only the StRIP survey measured the effect that instructors might have on students' affective response. Even then, the instructor effect is measured as two broad concepts that are not suggestive of specific practices or examples of how to get students to buy in to reformed instruction. An instructor interested in improving student agreement with their reformed class may find from the StRIP survey that they should explain and facilitate their class activities to get better responses from the students. However, these surveys do not provide explicit feedback about which topics to cover with students and how to cover them. It is here that an instructor may turn to the Science Education Initiative's (SEI's) materials for framing an interactive engagement class ("Framing the Interactive Engagement Classroom, Boulder: Science Education Initiative at the University of Colorado, Boulder," 2013).

The SEI has compiled a set of "strategies that instructors use to engage students in active learning classes" (Chasteen, 2017, p. 1). These strategies come in the form of written descriptions and PowerPoint slides submitted by instructors across the country. The materials have been organized into an extensive set of research supported guidelines for engaging students in active learning. They begin by discussing the factors that promote student engagement: expectations, metacognition, motivation, and class community. The first part of the guideline is to set student expectations. The SEI collection describes why setting expectations is important and then provides suggestion on how to set clear expectations. For example, one suggestion is to highlight the shortcomings of traditional lecture. They offer several ideas on how to do so based on the materials provided by instructors across the country. While research heavily supports the concept of the necessity to set clear expectations, in several cases, the effectiveness of the specifics on how to do so (e.g. highlighting the shortcomings of traditional lecture) is not research supported.

Following the factors that promote student engagement, the SEI framing guidelines discuss how to apply those factors in specific class situations. For example, on the first day, they recommend making explicit connections between the class and the students' lives. Later on in the semester, to reinforce interactive engagement, they suggest having students reflect on their learning. Both of these strategies were suggested by other instructors who have found them useful. However, that does not mean they are proven to increase student engagement or agreement with the course. Many of the specific practices suggested by the SEI are anecdotal accounts of what instructors have found useful.

The SEI framing materials are an extensive resource for instructors interested in getting students engaged or to buy-in to their reformed course. It provides instructors with the numerous concepts and ideas that go into student engagement. The ideas are backed by research as being important to getting or keeping students involved in the class. The framing materials provide instructors with many example strategies on how to implement these ideas in their class. However, many of these specific strategies are only anecdotally shown as useful. There is still room for research to empirically show if these specific strategies do indeed lead to an increase in student buy-in to reformed classes.

3.3.5 Summary

Research has begun to address students' affective domain in regard to reformed instruction. Physics education researchers have developed several surveys that address students' attitudes and expectations. The first of these may be the MPEX, which measures how students interpret the process of physics in class (Redish et al., 1998). The MPEX, however, looks at students' attitudes towards the class environment more than it does so for the class format. Since my interest is in student buy-in towards the class format, the MPEX is not appropriate.

The PEVA (Gaffney et al., 2010; Gaffney et al., 2013) is a step in the right direction because it does ask students about specific class actions that make up the format of the class. Using the PEVA, Gaffney et al. were able to show that student expectations can shift towards alignment with reformed pedagogy when the instructor takes the time to orient the students. Also using a revised form of the PEVA, one can determine if students reacted positively or negatively when their expectations were not met (Chini et al., 2013). However, the PEVA does not allow me to determine what students would rather have from their class if their expectations were negatively violated. Thus, I cannot determine the extent of their disagreement nor their preference.

The StRIP instrument (Nguyen et al., 2016) was designed to measure how students respond to reformed instruction on the factors of value, positivity, participation, and evaluation. It also measures students' expectations of a class, their preferences, and the strategies their instructor used for in-class activities. The StRIP instrument has shown that, in general, students do not act unfavorably towards reformed instruction (Nguyen et al., 2017). However, student response is not necessarily a measurement of buy-in. Student response could be thought of as a result of student buy-in or as independent from buy-in. Students could possibly disagree with the class format and thus respond poorly to the reformed instruction. It may also be possible that students disagree with the class format but still think the instruction they are receiving is good enough and so react positively to it. So, although it may appear at first that a measure of student response is a measure of student buy-in, they are not necessarily one in the same. The StRIP instrument makes measurements at the class level; a person using the StRIP survey cannot determine student response to the specific activities that occur in the class. It would be beneficial for my research to know exactly what activity students do or do not buy in to so instructors can better prepare to gain buy-in by discussing the specific activities students disagree with most.

The final survey was a call for materials from the SEI to collect and disseminate strategies for introducing and implementing RBISs (Chasteen, 2017). These strategies have been organized into a set of guidelines instructors can follow to help them teach a reformed class. While the concepts behind the guidelines, like setting clear expectations, are supported by research, the specific suggestions on how to do that, like describing the disadvantages of a traditional class, are not necessarily research supported. With my surveys, I hope to prove the effectiveness of specific strategies for increasing student buy-in, many of which are recommended strategies found in the SEI materials.

There is one final deficiency that appeared in every survey, aside from the SEI call for materials. Every response on each of these surveys were in the form of students ranking on a Likert scale. Likert scale responses are often reported with the average and standard deviation of the rankings. Averages and standard deviations are appropriate for use with interval data, where there is equal spacing between adjacent values. For example, “height” can be represented as interval data because there is one unit (e.g. inch or centimeter) between each successive value. Likert scale data, though often treated like interval data, is actually ordinal. Ordinal data are categorical data that have a certain order to them. For example, infants (recorded as 1), toddlers (recorded as 2), children (3), teens (4), adults (5), and elderly (6) are categories that have a sensible order to them. Note, that the spacings between these ordered categories are not uniform; the difference between infants and toddlers is a year or two but the difference between adults and elderly can be tens of years. It would be misleading and inappropriate to treat this data as interval data ranging from one to six and report the average, yet this is exactly what is frequently done with Likert scale data.

Likert scale data is also often used for subjective rankings. One will commonly see a survey prompt to indicate their agreement on a scale from strongly disagree (1) to strongly agree (5). In the context of the StRIP survey, students are asked if in their ideal class, an action would occur “Much less” (1) or “Much more” (5) than what they expect of their actual class. As these are subjective rankings, what one student imagines as “much less” may be what another student imagines as simply “less”. Even though they may have the same concept of how much less they want, they may call it different things. Thus, again, it would be inappropriate to average this type of data because the values do not necessarily mean the same thing to every person. In my surveys, students will be asked to respond using a more objective interval type scale. In the cases where I use a Likert scale response, the data will be treated in a way that does not assume equal intervals between values. Also, the responses on the Likert scale will not have different subjective meanings for each student taking the survey.

CHAPTER FOUR: PUMBA AND SIMBA

In this chapter, I describe the two surveys I have created to quantify student buy-in and to identify the methods and formats instructors use to get student buy-in. For each survey, I begin by providing a general overview of the content of each survey and how it is administered. Then I discuss the development process, outlining the major decisions made as I refined the surveys from their first versions to their final versions. Many of the changes to both surveys were made as a result of my findings from validity interviews conducted with students or instructors, as described in the following section. Finally, I explain how agreement is calculated from the student responses on the student survey.

In chapter three, I identified several features that should be addressed in my surveys to best answer my research questions. These features were not present in some or all of the prior surveys in the area of student attitude or response to instruction. By learning from the past surveys and addressing these features in my survey, I have created an instrument that effectively measures student buy-in and the methods used to generate buy-in. As a reminder, these features are:

1. *The surveys should measure specific activities of the class format instead of the learning environment or culture.* I have defined the class format as the importance placed and time spent on class activities. While student buy-in to the learning environment or culture is something to pursue, it is not the focus of this study.
2. *The student survey should measure student preferences.* This allows me to know whether students agree with their class and lets instructors know how to adjust their message to encourage greater student buy-in.

3. *The surveys should be given after students have been oriented to the class.* The PEVA shows that student expectations shift after an orientation. I want to know how well students agree with what they are told will be the format of the class during orientation.
4. *The student survey should be able to measure buy-in to the class as a whole and to individual activities of the class.* This feature looks similar to feature 1; however, the StRIP survey addresses feature 1 but not feature 4. The StRIP survey asks students about specific activities of the class but their calculation for the student response to those activities is only to the class as a whole.
5. *The instructor survey should measure instructor methods to get student agreement for each class activity.* One of the goals of this research is to provide practical advice to instructors on how to gain student buy-in. In order to do this, I need to measure specific strategies.
6. *The surveys should not treat Likert scale data as equal interval type data.* Treating Likert scale data as interval type data assumes equal spacing between the levels, which is not a valid assumption.

4.1 The PUMBA

4.1.1 Overview

I developed the Perceptions of Undergraduates Measure of Buy-in Achievement (PUMBA; see Appendix A) survey to be taken by students enrolled in a studio style course. Students take the survey online, usually outside of class time, and is typically completed in five to ten minutes. Students are asked to complete the PUMBA at the beginning of the semester, after they have been oriented to the format of the class. Taking the PUMBA at this point measures student buy-in with the instructor-described class format and addresses feature 3. Students are also asked to take the survey at the end of the semester, within the last two weeks of class, to measure student buy-in at the end of the semester. The versions of the PUMBA given at the beginning and end of the semester differ slightly in their wording. When referring to their actual studio physics class, the beginning-of-the-semester PUMBA is written in future tense. This version of the PUMBA measures what students expect the class will be like. The end-of-the-semester PUMBA is written in present tense. This version measures what their actual studio physics class is like in an average week instead of what it will be like.

The PUMBA asks students to compare the class format for their actual class and for a hypothetical class that is the most effective at teaching them physics. I am defining the class format as the importance placed on class activities and the time spent on these class activities. Table 9 defines the in-class activities, and Table 10 defines the out-of-class activities represented in the PUMBA. I chose the in-class and out-of-class activities based on my experiences and observations of studio classrooms. The class activities are meant to cover all activities that may happen for a studio class. Student and instructor interviews (see sections 4.1.4 and 4.2.3) provided evidence that the activities described in Tables 9 and 10 do indeed cover all activities for studio classes. There are four measures of student agreement: agreement with 1) the importance of in-class activities, 2) the importance of out-of-class activities, 3) the time spent on in-class activities, and 4) the time spent on out-of-class activities.

Table 9. In-class activities and their definitions.

Activity	Definition
Lecture	The instructor is talking to the whole class.
Quiz/Test	Formal graded assessment.
Step-by-step labs	Labs in which you are told what to do, what to measure, and what to calculate.
Investigative labs	You are given a research question and you must determine what to do, measure, and calculate to answer it.
Other small-group work	Any work that students do with the help of their classmates that has not yet been covered by the other listed activities.
Individual work	Students are working individually without the help of their classmates.
Student presentation	Students present their work to the entire class.
Viewing/engaging in physical demonstrations or simulations	Students are taking part in, or watching others engage in, a brief activity with the purpose of being able to observe a physical phenomenon.
Whole class discussion of scientific concepts	Discussing scientific concepts with the whole class. Students contribute to the majority of the conversation.

Table 10. Out-of-class activities and their definitions.

Activity	Definition
Reading/annotating assigned text	Reading or annotating texts that your instructor has assigned to you.
Reading beyond what is assigned	Reading texts that your instructor has not assigned to you.
Completing assigned conceptual/problem solving homework	Completing conceptual or problem-solving homework that your instructor has assigned to you.
Completing extra conceptual/problem solving work	Completing conceptual or problem-solving work that your instructor has not assigned to you.
Writing lab reports	Completing a lab report as homework for a lab activity done in class.
Working on group projects	Working on projects with a group out of class. This does not include doing homework, reading, or group studying.
Using other online resources or viewing instructional videos	Using online resources to supplement your learning. Viewing any videos with the purpose of gaining knowledge appropriate for class.
Getting help from instructor/TA/tutor	Getting additional help from someone else, not including classmates.

4.1.2 How the PUMBA Measures Student Buy-in

The functionality of the PUMBA was inspired by my earlier work investigating GTA buy-in to reformed instructional strategies (Wilcox et al., 2016). The survey used in that study asked GTAs to indicate how much class time they felt they were supposed to spend on certain teaching actions in the classroom (e.g. explaining, open dialogue, observing the room, etc.) based on the course design. Additionally, the GTAs were asked to indicate how much time they felt was best to spend on these same teaching actions. If, for the same action, their response was different for these two prompts, then the GTA disagrees with how much time they feel they should be spending on that action. For example, a GTA might indicate that based on the design of the mini-studio, only a little time should be spent explaining. However, they might believe that it would be best if much of the time in the mini-studio was spent explaining. Because this GTA answered differently, they do not agree with how they believe they are meant to teach based on the design of the course when it comes to explaining. The sum total of disagreements across all teaching actions surveyed provides an overall measure of the GTA's buy-in to how they are meant to teach.

The PUMBA operates in a very similar fashion as the GTA survey. Both surveys essentially measure how something is or is meant to be as perceived by the survey taker, and what the survey taker would rather have. The difference between these two concepts is the measure of buy-in or agreement. For this research, instead of teaching actions like the GTA survey, I want to measure agreement with a class format. I define the class format as the importance placed on and time spent on in-class and out-of-class activities instead of the learning and classroom culture, addressing feature 1. It may be important to measure student buy-in to a classroom culture, but it is not the focus of this research.

4.1.3 Description of the Final Version of the PUMBA

The PUMBA begins with a section that asks students, "For which [class activities] did your instructor explain its role for this course, such as how it will work and/or how much time will be spent on it?" The options here are similar to the in-class activities with two changes. First, "step-by-step labs" and "investigative labs" are combined into "lab group work". Second, an extra option, "out-of-class work", is added. These options match what instructors see in their survey, the SIMBA. The small differences are a result of how the options in the SIMBA were chosen and will be explained in the SIMBA section (4.2.2). After describing which feature of the class format the instructor discussed, students are asked which of those discussed activities the instructor tried to get them to agree with or understand the importance of. These two sections are essentially a miniature version of the instructor survey, taken from the students' perspective and can work as a check to the instructor responses on their survey. Differences between the students' responses here and their instructor's responses on the SIMBA may indicate that either students were not paying attention to their instructor or that the instructor was not clear in the purpose of their discussion.

Next, students begin the main portion of the PUMBA which measures their agreement with the class format. I first present students with all of the class activities and their definitions as shown in Tables 9 and 10. Then, I ask students to imagine their most effective physics class. This class is described as "a hypothetical class that is designed so that it is the best class possible for teaching you physics." This hypothetical class represents a student's ideal learning experience. It is not necessarily meant to represent their desires; if students were asked for what they wanted, they might not respond with anything at all. The hypothetical most effective class (HMEC) is supposed to make students reflect on what they know about themselves and how they learn best.

Students next respond to two sections surveying their agreement to the importance placed on the in-class and out-of-class activities. I ask students to compare the importance placed on each activity in their actual class to the importance they would place on that activity in their HMEC. Specifically, I ask them to complete this sentence, "In my <u>most effective</u> class, compared to my actual class, this <u>in-class</u> activity would be..." Students then select one of three choices: 1) "Less important than my actual class", 2) "Equally important as my actual class", or 3) "More important than my actual class". If students respond with the "equally important" choice, that shows that they agree with the importance placed on that activity. If students respond with "less important" or "more important" I know what they would prefer, thus the survey addresses feature 2. In asking if students would simply prefer more or less importance, I make no assumptions about how much more or less. I am not treating Likert scale responses as interval type data, and so I am conforming to feature 6. Students respond in this format for importance placed on both in-class and out-of-class activities.

I placed these sections before the two sections regarding the time spent on these activities so that students would have already answered about importance before they responded about time. Thus, when responding to the time agreement sections, they should recognize that the time sections should be thought of separately from importance. To reinforce this concept, the prompt of each of the four buy-in sections of the survey includes the statement, "Do not confuse importance with time. Something could seem important to the class but not occur very often."

After responding to the importance agreement sections, students move to the in-class time agreement section. For each in-class activity, students are asked to indicate the proportion of in-class time that will be or is spent on that activity, once for their actual class and then again for their HMEC. As these activities cover everything that might happen in a studio class, without any overlap, the sum of the proportions must add to 100%. Asking for students to respond for their actual class and their HMEC allows me to learn how students would prefer to spend their class time in comparison to how their time is spent in their actual class. This addresses feature 2, which is that the survey should measure student preferences.

The final section of the student buy-in portion of the PUMBA asks students about the time spent on out-of-class activities. This section functions similarly to the importance agreement sections. Students are asked to compare the time they would spend on each activity for their HMEC to the time they feel they are *intended* to spend on each activity for their actual class. I emphasize “intended” because how students actually spend their time outside of class is up to them. However, I want to compare what they think is best for them to what they think they are supposed to be doing, regardless of if they are actually doing that. So, students finish the sentence, "For my most effective class, *compared to my actual class*, this out-of-class action would spend..." with one of four choices: 1) “less time than my actual class intends”, 2) “equal time as my actual class intends”, 3) “more time than my actual class intends”, and 4) “I have no feeling as to how much time my actual class intends for me to spend on this activity.” I included the final option in case students were never given an idea about how much time they should be spending on out-of-class activities. In such a scenario, they cannot determine if they would prefer less/equal/more time in their HMEC.

At the end of the PUMBA, students answer several demographic questions, including gender, race, ethnicity, age, and class standing (e.g. Freshman, Sophomore, etc.). In addition to the more typical demographic questions, students also indicate if they have taken a studio style class before the current semester. I collected this demographic information to explore its effects on student buy-in. This is important to do because, for example, a student’s gender might adversely affect their agreement with the course. Some previous studies have found that student gender alone does not affect course evaluations (Feldman, 1977). However, when student gender is combined with the gender of the instructor, there are differences in teacher evaluations (Feldman, 1993; Potvin, Hazari, Tai, & Sadler, 2009). Especially in physics, Potvin et al. find that female students rate female instructors statistically worse than they rate male instructors, a trend not found in biology or chemistry. Male physics students also rate female instructors statistically worse than male instructors and this trend is found in biology and chemistry. Past research has also shown that class standing typically does not have an effect on student ratings of a course (Heilman & Armentrout, 1936; Stewart & Malpass, 1966). Although some of these demographics may not have an effect on course or teacher evaluations, they still might affect student agreement with the course. Thus, it is still necessary to record student demographic information.

In summary, the PUMBA begins by asking students what their instructor discussed about the class format and which activities the instructor tried to get the students to agree with. Following this, students compare the importance placed on in-class activities between their actual class and their HMEC. Then, they do the same for out-of-class activities. Next, students indicate the proportion of class time that is or will be spent on various in-class activities for both their actual class and their HMEC. The final agreement section has students compare the intended time spent on out-of-class activities for their actual class with the time they would spend on those activities for their HMEC. Finally, students respond to demographic questions.

4.1.4 Revisions to the PUMBA

The PUMBA underwent several major revisions over the course of a year of development. Starting from the skeleton of a survey for GTA buy-in to teaching methods, the PUMBA evolved into what you can now find in Appendix A. Throughout the development, some issues became clear after trial administrations of the survey, while some issues and their solutions arose from validity interviews with students. For more information on the validity interviews, see section 4.1.5.

4.1.4.1 Inclusion of Importance

Early in the development of the PUMBA, while the survey was still in the pilot stage, there was a concern that students might confuse the amount of time spent on an activity with its importance. Specifically, if an action seems to be important for the class, then due to its importance a student might incorrectly indicate that action as occurring very frequently. This may not always be the case in actuality. For example, an instructor might express the importance of demonstrations in the class, but demonstrations may only occur for brief periods in an average week. However, a student in this class may understand that demonstrations are important, and so she may unknowingly inflate the amount of time she states was spent on demonstrations.

To counteract this, I included with the time sections questions about the importance of each activity. Up to version 2, students would respond to the time spent on an activity and the importance placed on that activity side by side. Doing this should make it very clear to the students that they should separate the ideas of time and importance. Beginning with version 3, I separated the time and importance questions, placing the importance questions at the start of the survey and the time questions following. I did this to facilitate a check through the survey program which would ensure that students' responses to the time spent on in-class activities added to 100%.

In the early versions of the survey, I checked whether students were assigning a large proportion of class time to activities because they seemed important even if the activities may not have taken much of the class time. The survey had students rank the importance of each activity on a scale from 0 (Not Important) to 3 (Very Important). I used the student responses to this section in combination with their responses towards the proportion of class time spent on an activity to test if students were assigning a large amount of time to activities because they seemed important. Figure 5 graphs the importance ranking against the time spent on an activity for all nine in-class activities for 334 students. The graph reveals that students assigned high importance to activities that occur for a larger proportion of the class. This makes sense as a class wouldn't spend much time on something that is not important. Most importantly, the graph does not show that high importance equates to a large amount of class time. Many different proportions of class times were connected to high importance rankings, meaning that when students are asked about the importance of an activity separately from the time spent on that activity, the two concepts are also separated in the student's mind. One final thing to note from Figure 5 is that students assigned importance to activities that they stated did not occur in their class. Although I did not want students to think that activities that occurred infrequently may not be important, it makes logical sense that if an activity does not occur at all, it must not be important. Figure 5 is evidence that students were not making this logical connection. This would lead to what I would later call the "time-importance mistake" which I discuss in section 5.4.2.

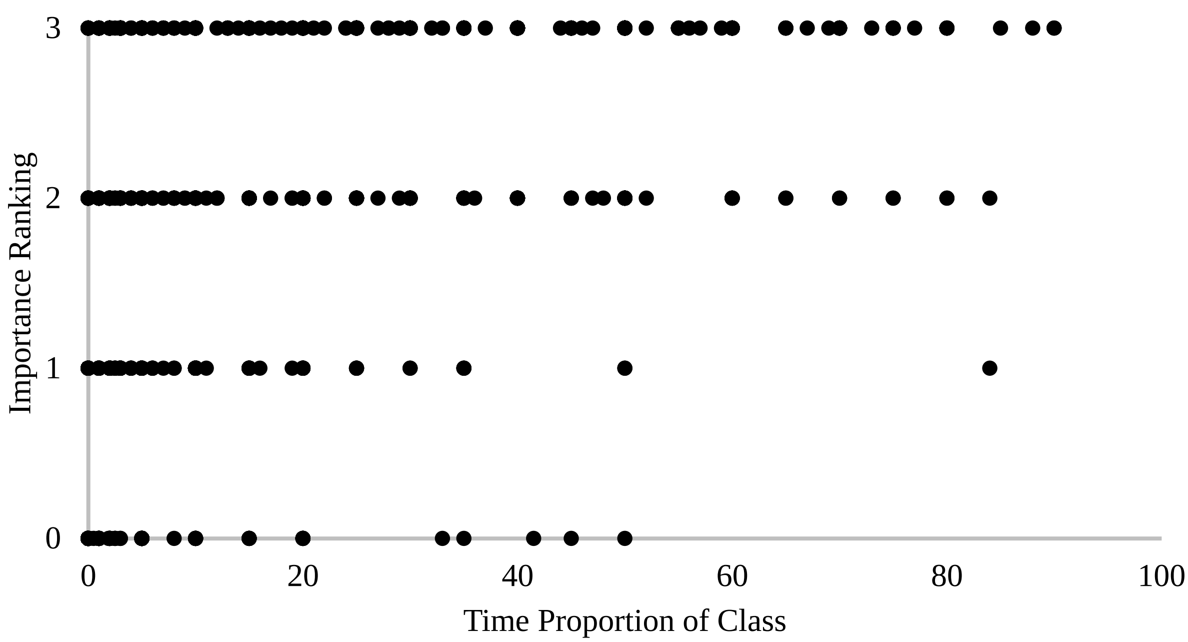


Figure 5. Importance ranking of an activity versus its indicated proportion of class time based on results from an early version of the PUMBA.

4.1.4.2 Evolution of the Response Type

The out-of-class time section of the PUMBA went through several iterations in an attempt to determine the best response method for students taking the survey. Unlike the in-class time section, the out-of-class time section cannot be answered with proportions. Proportions work for the in-class time section because there is a set total amount of time spent in class each week. Students are asked to respond based on an average week of class, which is roughly six hours of class time. However, the time spent on out-of-class activities each week is variable; some students may spend more time on out-of-class activities than others. If students were to respond to this section with proportions, I would be uncertain as to how much time that proportion represents. This would especially lead to difficulty when comparing the proportion of time spent on out-of-class activities for the student's HMEC and what they feel their instructor intends, e.g., 10% of out-of-class time in the HMEC may be 15 minutes where as 10% of out-of-class time for what the instructor intends may be 30 minutes. In this example, the student would spend less time on an activity than what they feel their instructor intends, but the proportion response type makes it seem as though the student agrees with the time spent since they are both 10%. To avoid this issue, I could not use proportions for the out-of-class time agreement section.

Naturally, the next logical option is to ask students to respond in terms of hours or minutes spent on each out-of-class activity. Students responded this way to the out-of-class time section up to version 2. Several difficulties arose with this answering method. First, students found it cognitively taxing to figure out in hours how much time they spent on each activity. When given the option to respond via rankings, percentages, or minutes on this section of the survey, students often preferred rankings. One student, when asked why he preferred rankings, said, "The lack of math would be a huge component." I wanted to avoid such challenging tasks because I did not want to fatigue students as they took the survey, in fear that they might start responding with nonsense answers to quickly finish. Another issue with this type of response method that arose from a validity interview was an overestimation of the time in hours spent on the out-of-class activities. One student, when discussing that section of the survey, realized in hindsight that her responses were much higher than reality. She did not realize as she was filling out the survey how much total time spent on out-of-class activities she was ending up with. This indicates that her responses to individual activities were inaccurate due to overestimation. Lastly, I worried that students might have difficulty assigning a specific value to how much time they were meant to spend on these activities for their actual class. Since out-of-class activities are not necessarily controlled by the instructor like in-class activities are, it is hard to determine what exactly is the intended time for students to spend on each activity.

The last response option for this section is a less precise ranking type of answer. Students mentioned having an easier time indicating vague rankings like “a small amount of time” working on lab reports or “a large amount of time” working on conceptual/problem solving homework. This should be easier for students as they should have some feeling as to roughly how much time they should be spending on out-of-class activities. They also should know how much time they would spend on these activities in their HMEC. Thus, a Likert scale response may be acceptable for this section of the PUMBA. The difference between the ranking for an activity for their actual class and the ranking for the same activity for their HMEC is equivalent to their agreement with the time spent on that activity. However, in deference to feature 6, I do not want to use the magnitude of the difference in Likert scale responses as a direct measure of agreement. At best, I can safely say whether a student would have less, equal, or more time spent on an in-class activity in their HMEC as compared to their actual class. I simplified the answer choices for this section to reflect how I would analyze the responses, giving students the option to say if that activity would occur less, equally, or more often in their HMEC than their actual class. I also give the students an option to say they have no feeling towards the amount of time they are intended to spend on an activity. In such a case, a student cannot accurately determine if they would spend less, equal, or more time on that activity. These response options began with version 3 and remain in the final version of the PUMBA.

Responses to the importance placed on class activities began as Likert scale response where students would indicate an activity being not important (0) to very important (3). When making revisions to the second version of the PUMBA, I realized that this violated feature 6. So, I changed the response options for the importance sections to match the new response options for the out-of-class time section. Students would compare the importance placed on an activity in their HMEC to their actual class by stating if it would be less/equally/more important.

4.1.4.3 Addressing the "Time-Importance Mistake"

One of the unexpected difficulties revealed from the validity interviews was students not knowing how to respond if an activity in the survey did not occur for their class. Strangely, multiple students stated that they would spend less time on an out-of-class activity, or place less importance on activities that they stated did not occur in their actual class. Students who made this mistake on the survey often noticed their error in the interview. However, since they did not realize the mistake until reviewing their responses with me and not while they took the survey, I could only expect that more students would make the same mistake as they complete the survey. To combat this, beginning in version 4, I added a statement to those three sections saying, "If an action seems that it will not occur in your actual class, consider it to have no importance in your actual class (i.e. your response should be either 'Equal…' or 'More…')" or "If your actual class does not intend for you to spend any time on an action, consider it to have no time spent on that action in your actual class (i.e. your response should be either 'Equal…' or 'More…')". While I expected this to solve the problem, I later realized that it did not. This mistake, that numerous students made when taking the survey, and how I accounted for it is discussed in detail in the "Data Cleaning" section of Chapter Five (5.4.2).

4.1.4.4 Adjusting Order of In-class Activities

In the third version of the survey, I realized that students were not responding to the amount of time spent on individual work like I would have expected based on my observations of studio classes. On average, students at one university at the end of the Spring 2017 semester were reporting individual work as occurring in 11% of their actual class. However, in all of my observations, individual work occurred very infrequently. I believe the reason for the discrepancy between student responses and my observational experience came down to a misunderstanding of the activity definitions as a result of the order of the activities. Originally, the in-class activities were ordered as: lecture, individual work, whole class discussion, quiz/test, step-by-step labs, investigative labs, student presentation, viewing or engaging in physical demonstrations or simulations, and other small-group work. There was little justification for the ordering of the activities aside from the first and last. I placed lecture first because I expected it to be one of the more frequent activities. I placed other small-group work last because I wanted it to be a catch-all. Any small-group work that a student did that wasn't captured by any of the earlier activities should have been considered this activity. This should have included group problem solving on worksheets or clicker questions. I realized that students may have been interpreting "individual work" as work that is *not* whole class work (e.g. group problem solving is not work done in front of the whole class). Most of what I had considered "other small-group work" may have been considered by the students as individual work (i.e. not whole class work). To address this, I rearranged the ordering of the activities to that shown in Table 9. "Other small-group work" follows immediately after the two lab activities because those are the only activities that are small-group work that wouldn't be considered "other small-group work". "Individual work" must come after "other small-group work" to avoid the issue of misunderstanding its meaning. After making these changes, student responses to the proportion of class time spent on individual work dropped to 6% as measured at the end of the Fall 2017 semester, which is more in line with what I expected based on my observations of previous studio classes.

4.1.5 Validity Interviews

As I was developing the PUMBA, I conducted individual interviews with students to ensure that they understood the survey and that I was correctly interpreting their results. In some cases, I was able to recruit students who had already taken the survey as a part of their class. At other times, when I wanted the students' insights in the middle of a semester or over the summer, I recruited students to participate in an interview who had not already completed the survey. At first, in the Summer of 2016, I interviewed six students to test test pilot versions of the PUMBA. Beginning in the Fall of 2016, I had the first official version of the PUMBA ready to distribute to classes. Starting with the first official version, and with each subsequent version, I found that three student interviews for each version provided me with a good idea of what I needed to change for the next version. Figure 6 illustrates the timeline of the revision process. Although students from multiple universities took the survey during its development, all students who participated in a validity interview took a studio class at UCF. This is primarily due to easy accessibility and availability of students at the same university in which I was developing the survey. Participants were compensated with $10. I continued the revising and interviewing process until I felt confident that students were correctly understanding the PUMBA without difficulty and that I was correctly interpreting their responses. In total, 15 students participated in a validity interview.

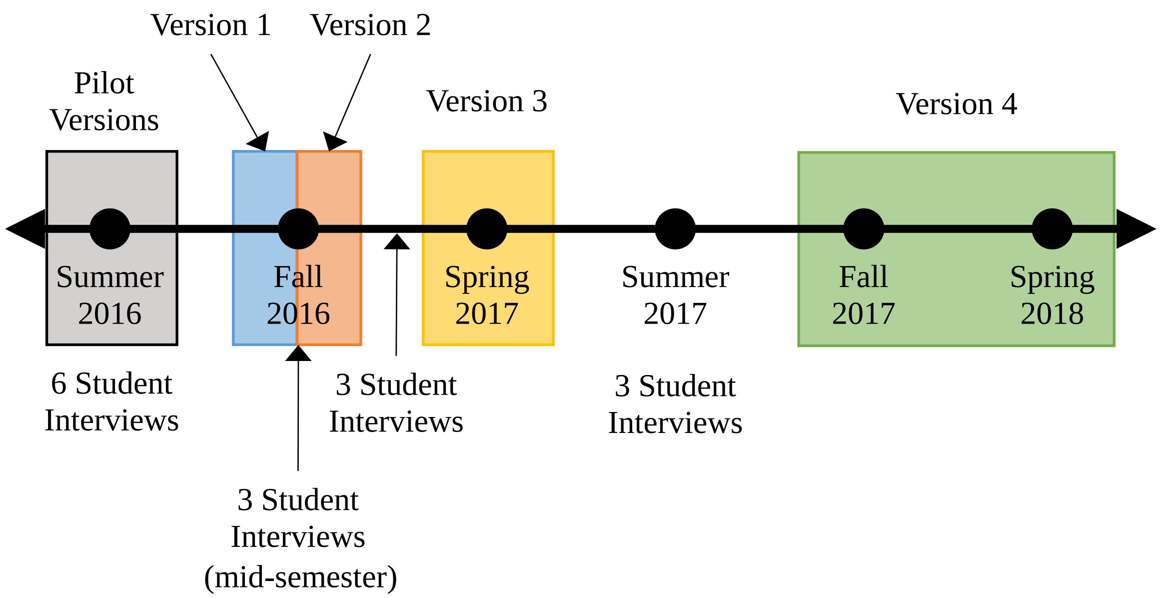


Figure 6. The PUMBA revision and validity interview timeline.

4.1.5.1 Types of Validity

The purpose of the interviews was to find evidence for validity based on the content of the survey and the relationship between the survey results and the student's self-described agreement with the class format. Reeves & Marbach-Ad state the goal of establishing validity "is to accumulate over time a comprehensive body of relevant evidence to support each intended score interpretation within a particular population" (2016, p. 2). In general, validity is "the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests" (AERA, APA, & NCME, 1999, p. 9). Although my survey is not necessarily a "test", the need to establish validity applies to any sort of item that results in a score (Reeves & Marbach-Ad, 2016).

Finding evidence for test validity based on its contents refers to the items in the survey and their relation to the construct being measured (AERA et al., 1999). In other words, this measure of validity checks if what is in the survey actually measures what I claim the survey measures. For example, items in a survey about a person's favorite Italian food should all consist of types of Italian food, not Chinese food. The survey items should fully represent the domain of the construct surveyed.

Although content-based validity is necessary to a survey's overall validity, it is not sufficient. As Messick points out, in finding evidence for content-based validity, test-scores are irrelevant so content-based validity alone is insufficient proof towards the validity of the test as a whole (1989). In other words, the test may be covering the right topics and ideas, but the interpretation of the results could be off base. Thus, I must also find evidence for validity based on the survey scores' relation to other variables. Typically, this is done by comparing scores to a measure of a similar construct, called "convergent validity", or measures of a different construct, called "discriminant validity" (Reeves & Marbach-Ad, 2016). Test scores should correlate with measures of a similar construct and show little to no correlation with measures of a different construct. To validate the results of my survey, I searched for evidence of convergent validity through interviews in which I compare agreement as measured by the survey (survey scores) with the student's self-described agreement (the measure of a similar construct).

When reporting the evidence of validity, I will refer to the interviews of students who took the third version of the survey. It is with these students that I found sufficient evidence for validity. Any evidence of validity found before version 3 would be for a version of the survey that is not representative of the final version of the survey. The final version only differs from version 3 in two aspects, neither of which fundamentally changes the survey. One change between version 3 and 4 was to include a sentence to prevent the "time-importance mistake". This change only serves to clarify the survey. The second change was a reordering of the in-class activities. Neither change would affect how a student understands the questions, if anything it would make it easier to respond to the survey. Thus, evidence of validity for version 3, which is fundamentally the same as the final version, is evidence of validity for the final version.

4.1.5.2 Evidence of Validity Based on Survey Content

In the context of my survey, which measures agreement with the class format, obtaining evidence of validity based on the survey's content means that the survey items should fully represent the class format as I am measuring agreement with the class format. As I have defined the class format as the importance placed on and time spent on class activities, the survey should cover all class activities. A common means of collecting evidence of content validity is by submitting the test items to expert review (Reeves & Marbach-Ad, 2016; N. L. Webb, 2006). Experts review the items and determine their relevance to the domain being measured, specifying any items that are missing or repeated. For my survey, the experts on the class format are studio instructors and students in studio classes. Instructors and students know the various activities that have or will occur in their class, so I asked both instructors and students if any activities were missing from the survey during their validity interviews for their respective surveys.

I began searching for evidence of content validity with question 8 of the student validity interview protocol (Appendix B) which asks students to explain in their own words what each question is asking them to establish whether the student understood the survey as I intended. I typically fixed discrepancies between how the student interpreted a question and how I intended for them to understand the question by a simple rewording of the question or by adding emphasis to overlooked parts of the question. For example, some students were considering the time spent writing lab reports both out-of-class and in-class when responding to the agreement with time spent on out-of-class activities section. As a result, I added emphasis to the “out-of-class” and “in-class” wording in the prompts.

Questions 10 and 12 of the interviews ask students to describe the class activities and if they felt that any class activities were missing or overlapping in the survey. These two questions provided evidence of content validity. Nearly all students agreed that the activities listed in the survey covered everything they do for their studio class. There were two instances from validity interviews on the first version of the survey where a student felt that something was missing from the list and both cases resulted in an addendum to the definition of an already existing class activity. A class at UCF makes use of Perusall (Mazur, King, Lukoff, & Miller), a program for annotating and interactively commenting on textbook readings, for homework. Originally, one of the out-of-class activities was simply “reading assigned text”. A student noted that with Perusall, a significant amount of time is spent coming up with a comment, question, or answer in the Perusall program and that this would not technically be considered “reading”. Another student noted, with regard to the original “viewing instructional videos” activity, that they used other online resources when studying besides videos. For example, this student mentioned that they often will turn to Wikipedia or HyperPhysics (Nave, 2000) for help. In both cases, I updated the original activity names and definitions to include the missing activity identified by the students. By version 3 of the PUMBA, each student interviewed agreed that the list of activities in Tables 9 and 10 cover all activities for their class and do not overlap. For instance, one student said, when asked if she felt any activities were missing in the survey, “I think it’s pretty comprehensive.”

Question 13 states, "To respond to several of the questions you need to have some idea of the importance of various activities for your actual class. What were your ideas about this based on?” The interviews revealed that students have developed a sense of importance for each activity based on their own experiences and on the instructor’s descriptions of the class at the beginning of the semester. One student, referred to as student 3A (third survey version and first interview), said in response to question 13 that her feelings about the importance of each activity were based on “How [the instructor] explained it at the beginning with like, the syllabus and also like, from like the in-class things, like what [the instructor] devotes the most time to. I feel like that’s what [the instructor] feels like is the most important aspect in a way.” Another student, 3C, based her ideas about the importance on her own perceptions as well as her instructor’s explanations. In response to the question student 3C said, “What I felt. And I based what I felt how important these are based on the explanation [the instructor] has given and like why we’re doing things. [The instructor] does explain that sometimes…and what also actually works and what the people around me say and what helps them the most.” Each student had developed some conception of the importance of each activity in their actual class. These conceptions were usually based on the instructor’s description of the class, the student’s experience with the class or previous classes, or both.

From the student interviews, I did find evidence that the different class activities did cover all activities that could occur in a studio class. I also found evidence that students were correctly understanding the questions and were correctly using conceptualizations of the importance of class activities to respond to the importance agreement sections of the survey. With this evidence, I felt confident that the PUMBA had content-based validity.

4.1.5.3 Evidence of Convergent Validity

I did not refer to the survey in the first half of the interviews because I used these questions to get a sense of how well the students agreed with the class format, providing a measure of student agreement that the survey scores could be compared to for convergent validity. I designed these questions to probe the student's opinion about the class. For example, I asked "How has taking the studio course been different than you expected?" Learning about their expectations and how the class compared to their expectations could reveal aspects of the class they are satisfied or unsatisfied with, suggesting aspects of the class they may agree or disagree with. I also used students' descriptions of their expectations of the class to validate their survey responses to questions about the importance of and time spent on class activities for their actual class. By asking the student what they like and dislike about the class (questions 3 and 4), I determined what specific activities I should expect to see high or low agreement within the survey. The final questions of the validity interview ask students to explain any differences or similarities between their responses for their most effective class and their actual class. These questions also work to confirm the reasoning behind the students' responses. By the students explaining why the responded how they did, I can ensure that I am interpreting their survey responses correctly.

For examples of how the student responses to the third version of the survey align with how they describe their agreement with the class activities, see Table 11. From these examples, you can see that the survey is accurately measuring the students' agreement with the individual activities. Thus, the survey scores of students' agreement correlates with a similar construct of student agreement measured a different way (through interviews). This is evidence of convergent validity.

Table 11. A summary of students' stated agreement with specific activities and how their survey matches their response.

Student	Statement	Summary	Survey Response
3A	The conceptual things often times for me personally are more difficult. So, I think like, just, how the class is set up where the labs also integrate with what we're learning in class kind of, uhm, when you do that and like visualize it you can kind of like understand what's going on.	Agrees with how labs work in class	Agrees with the importance of labs and the time difference is only off by 2 or 3 percentage points.
3A	I'm like usually not a fan of group work but I think that like this definitely I feel like I am understanding better throughout. Just being able to like talk through it.	Finds group work helpful for her understanding.	Agrees with small group work importance and would only prefer 5 percentage points less time spent on it.
3A	Usually, I like to...I'm more of an independent worker I guess, but...I think sometimes I would rather do the problem myself first and then talk about it.	Feels that there is not enough individual work.	Prefers 10 percentage points more individual work.
3B	[Talking about small group work] So that's really helpful because you talk with people and like there, you know, [the instructor] and the TA and the LA, and you can be like, "I got this far but I have no idea what happens next."	Finds group work helpful because of interaction with students and instructors	Agrees with importance and time of small group work.
3B	It's been really nice to have like the lab and lecture together.	Satisfied with labs and lecture.	Agrees perfectly with labs and lecture importance and time.
3B	Ideally, I would like to get...having to go to SARC [an out-of-class help center], you know, something that's less important because I understand what's happening in class like so much more.	In her ideal class she would not need to get help outside of class because she would already understand the material.	Out-of-class help is less important.
3C	She does emphasize getting help from her and the TAs and the tutor a lot and I just haven't gone. So, I just thought that was less important than what she was making it seem to be.	Student does not think that out-of-class help is as important as the instructor makes it out to be.	Out-of-class help is less important.

In the final round of validity interviews, for question 7 I asked students to assign a number from 0 to 100 corresponding to their overall agreement level. I wanted to see if this number was similar to the overall agreement level measured by my survey (see Table 12). From the table, you can see that the survey does reflect the students' stated agreement level for students 3A and 3B. Student 3A said her in-class activity agreement was 80 out of 100. Her survey scores for the two in-class sections were 75 and 85. Student 3B ranked her overall agreement as a 90 but generally disagreed with out-of-class activities. Her in-class scores do agree with her overall agreement (88 and 95) and her out-of-class scores do indicate low agreement (25 out of 100 for each). The survey scores do not match well with student 3C's stated agreement level. However, as we went through her responses, she justified her disagreements with individual activities. This is why the survey does not ask students directly how well they agree with the class overall. By forcing students to respond to individual activities they must carefully consider their agreement, which should give a more accurate measure.

Table 12. A summary of students' stated agreement compared to their survey measured agreement.

Student	Stated Agreement	In-Class Importance Agreement	Out-of-Class Importance Agreement	In-Class Time Agreement	Out-of-Class Time Agreement
3A	Agreement with in-class activities ranked as 80 out of 100	75	50	85	29
3B	Overall agreement ranked as 90 out of 100. For out-of-class activities she said, "It's as good as it's gonna get" with a negative connotation implying low agreement.	88	25	95	25
3C	Overall agreement ranked as 100 out of 100.	55	25	87	N/A (too many "no feeling" responses)

4.1.6 Summary of the Development of the PUMBA

With 15 student interviews over the course of a year, I searched for evidence for validity of the PUMBA based on content and on the survey scores' relationship to the students' agreement as revealed in their interview responses. Through the interviews, I learned several ways in which I could improve the PUMBA. I summarize the changes I made to the PUMBA throughout its development, including the changes due to student interviews, in Table 13. The validity interviews have ensured that the class activities cover everything that could occur in their studio course. From the student responses to the validity interviews for version 3 of the PUMBA, I am confident that they are interpreting the questions and response options correctly, and also that I am correctly interpreting their survey answers. Although only UCF students were involved in the validity interviews, the features of the survey (i.e. class activity options and response methods) are applicable to a range of studio style methods. Thus, it should be safe to administer the PUMBA to students of other studio style classes outside of UCF.

Table 13. Changes to the PUMBA throughout development and the motivations for those changes.

Changes to the PUMBA	Version in Which the Change First Appeared	Motivations
Added to Definition for “Online Help” and “Reading” Activities	2	Students suggested that "annotating text" and using resources like "HyperPhysics" were not represented in out-of-class activities.
Condensed Prompts and Added More Emphasis	2 and 3	S2B: “Maybe you should bold certain things or like bullet points rather than paragraphs because like students, they're not going to read the paragraphs."
Added 100% In-Class Time Check	2 and 3	32% of students in version 1 did not total the in-class time proportions to 100%.
Added Attention Checks	3	To ensure that I do not keep responses from students rushing through or being dishonest.
Placed Importance Section Before Time Section	3	To ensure students separate the ideas of importance and time and allow for the 100% in-class time check.
Changed Importance and Out-of-Class Time Sections to “Less/Equal/More” Responses	3	To simplify the response method and conform to feature 6.
Added "No Feeling" Response	3	To allow students to have an option indicating that they cannot make a comparison because they do not know how much time they are intended to spend on an out-of-class activity.
Added Student Perceptions of Instructor Strategies Section	3	If students did not perceive or cannot recall the instructor strategies, the strategies may not have an effect on their buy-in.

Changes to the PUMBA	Version in Which the Change First Appeared	Motivations
Reordered In-Class Activities	4	Student responses to the survey indicated confusion between "individual work" and "other small group work".
Added Time-Importance Mistake Prompt	4	S3A for on saying “less important” for group projects: “I guess it was just because like we didn't do it. So, I guess like my correct response should have been like, "I have no feeling as to how much my actual class intends for that one." S3B: “Well there is some things that we didn't do in class so I was like, ‘What am I supposed to do?’...So I was like do I put equally important as my actual class?”

4.1.7 Reliability

The AERA, APA, and NCME *Standards for Educational and Psychological Testing* defines reliability as referring to the consistency of measurements “when the testing procedure is repeated on a population of individuals or groups” (1999, p. 25). In essence, a reliable measure should yield similar results across multiple administrations to the same population. A measurement’s reliability can be quantified by reliability coefficients which classically fall under one of three categories (AERA et al., 1999):

1. “Alternate-form coefficients”, which are derived from the administration of similar but different forms of an instrument.
2. “Test-retest coefficients”, which are derived from the administration of the same instrument at different points in time.
3. “Internal consistency coefficients”, which are derived from the relationship among scores on items within the instrument.

A more modern reliability coefficient is the “generalizability coefficient” which allows the developers of an instrument to attribute error to specific sources, like the number of items in the instrument or the population using the instrument (Brennan, 2010). The unique aspect, and what may make the generalizability coefficient preferable over the classical reliability coefficients, is that one can generalize the effect of these sources of errors to determine what is needed to create an adequately reliable instrument. For example, an instrument with 20 items may yield a generalizability coefficient of 0.70. By knowing how the number of items contributes to the calculation of the generalizability coefficient, one can then calculate the number of items needed to reach an acceptable generalizability coefficient. So, one may be able to calculate that 25 items in the instrument would obtain a more acceptable 0.80 generalizability coefficient.

Not all reliability coefficients are appropriate for every instrument. Test-retest reliability has come under criticism due to the fact that respondents on a test likely change between administrations. Respondents may become sensitized to the matter of the instrument and other influences, such as their mood, may affect how they respond (Crocker & Algina, 1986; Golafshani, 2003). For the PUMBA, it is not reasonable to assume that students’ buy-in to the class will remain static throughout the semester. For example, the instructor may change their teaching style to better appease the students’ desires or may have a discussion that convinces students to agree with the format; in both cases, student agreement will likely increase. Even receiving a grade on a minor quiz or clicker question could affect how students agree with the course at a specific point in time. Test-retest reliability is an inappropriate measure of reliability for the PUMBA because I cannot assume that student agreement should remain consistent over time.

The other reliability measures are also inappropriate for the PUMBA, due to the fact that it is a formative test as opposed to a reflective test. Formative and reflective tests differ in the relationship between the construct being measured, called the latent construct, and the observable items responded to (Fornell & Bookstein, 1982). For a reflective test, the latent construct causes something that can be observed (i.e. responses to items on a test). On the other hand, for a formative test, the test items (what is observed) comprises the latent construct.

For example, the Force Concept Inventory (FCI) is a reflective test. The latent construct measured by the FCI is the student's ability with the concepts of force. Their ability with force concepts determines their performance on items of the FCI, each of which probe the students' ability with force concepts. If their ability with force concepts increases or decreases, that should be reflected on each item. A student who does well on the FCI does so *because* they have high ability with the material. Doing well on a test does not *create* high ability with course material. Thus, the FCI is a reflective test.

On the other hand, suppose there is a test measuring the latent construct of overall customer satisfaction with an automotive mechanic. The test may have three items: satisfaction with the price, satisfaction with the speed of service, and satisfaction with the quality of service. The three items comprise the latent construct: overall satisfaction. A customer may be completely satisfied with their automotive mechanic *because* they are satisfied with each of the items. Thus, this is a formative test; the items form the construct. If any item in the customer satisfaction test were to increase, then their overall satisfaction increases. However, an increase in overall satisfaction does not necessarily mean an increase in each item comprising their satisfaction (Diamantopoulos & Winklhofer, 2001).

The PUMBA is a formative survey; the student agreement construct is formed by their agreement to individual class activities. Student agreement (the overall construct) increases because agreement increases with one of the class activities (the items in the survey) . An increase in overall student agreement does not necessarily mean that the student should agree more with every class activity as would be the case in a reflective test. Diamantopoulos and Winklhofer (2001) operationally define a formative test as one where “a concept is assumed to be defined by, or to be a function of, its measurements” (Bagozzi & Fornell, 1982, p. 34). In the case of the PUMBA, student agreement as I have defined it is the agreement a student has with the importance of and time placed on class activities. So, I created the PUMBA as a formative test.

Due to the nature of a formative test, alternate-form reliability coefficients are impossible. To have an alternate form of the PUMBA, I would need to change or remove items that create student agreement. However, that fundamentally changes the meaning of the student agreement score. As Bollen and Lennox stated in regard to formative indicators (items), “omitting an indicator is omitting a part of the construct” (1991, p. 308). The same holds true when considering changing indicators in a formative test. Since the PUMBA cannot have an alternate test, this form of reliability is not applicable.

Internal consistency coefficients and generalizability coefficients are not applicable to the PUMBA for the same reason; both rely on correlations among the items of the test, which is not a feature of a formative survey. The most common traditional measure of internal consistency (Hogan, Benjamin, & Brezinski, 2000) is the coefficient α (Cronbach, 1951) also called Cronbach’s α. Other measures for internal consistency exist, like the split-halves test and the Kuder-Richardson test (often referred to as KR20 or KR21) (Kuder & Richardson, 1937). The split-halves test measures reliability by determining the correlation in scores for two halves of a test. The Kuder-Richardson test is an improvement upon the split-halves test as it determines the average correlation between all possible split halves of a test. One of the disadvantages of the Kuder-Richardson test is that it may only be used when items are scored as either correct or incorrect. Cronbach's α improves upon the Kuder-Richardson test because items may receive any score, and, if the items are scored dichotomously (i.e. correct or incorrect), then Cronbach's α simplifies to the Kuder-Richardson test (Cronbach, 1951). Each of these traditional internal consistency measures rely on correlations. The split halves test is a measure of the correlation between two halves of a test and the Kuder-Richardson test measures the average of all possible split half correlations. The calculation for Cronbach's α, as shown in Equation 2, relies on larger correlations to result in a larger coefficient.

$$\alpha = \frac{n}{n-1}\left(\frac{2\sum_{i=1}^{n}\sum_{j=i+1}^{n} r_{ij}}{n+2\sum_{i=1}^{n}\sum_{j=i+1}^{n} r_{ij}}\right) \tag{2}$$

In Equation 2, n is the total number of items on the test and r_{ij} is the correlation between item i and item j. As correlations between items increase, the term in the parentheses increases approaching a value of 1, thus increasing the value of Cronbach's α. In order to achieve a high reliability coefficient using Cronbach's α, there must be large correlations between items.

In classical test theory, the domain where Cronbach's α resides, measurement error (what leads to unreliable tests) is considered random variation, not associated to any specific source (N. M. Webb, Shavelson, & Haertel, 2006). In generalizability theory, measurement error is connected to specific sources, like differences in item difficulty or differences in testing occasion. In generalizability theory, there is an internal reliability coefficient called the generalizability coefficient. However, the generalizability coefficient is not appropriate for the PUMBA for the same reason as Cronbach's α; it relies on correlations between items.

These reliability coefficients work well with reflective tests where variance in total test scores cause variance in item scores (Diamantopoulos & Winklhofer, 2001). In fact, this connection between variance of total test scores, referred to as the "composite" in Cronbach's paper, and item scores, referred to as "subtests", is the foundation of Cronbach's α (Cronbach, 1951).

> *"Under the assumption that the variance due to common factors with each subtest is on the average equal to the mean covariance between subtests, α indicates what proportion of the variance of the composite is due to common factors among the subtests."*

However, in formative tests, causality flows opposite to that of reflective tests; variation in item scores cause variation in total test scores. There is no basis to assume correlations between items. Referring back to the formative survey example of satisfaction with an automotive mechanic, one person may be completely satisfied with the speed of service and nothing else while another person is completely satisfied with the cost and nothing else. These two people would have the same total satisfaction scores but the items would have no correlation. Thus, internal consistency reliability measures, which depend on high correlations, are not appropriate for formative tests.

Although no reliability measure is appropriate for formative tests, it does not mean that the PUMBA is unreliable; rather, it means that reliability is not an applicable construct for the PUMBA. It is like trying to check the tire pressure of a wooden wheel. It may be a wheel but tire pressure is irrelevant for its function. A formative survey may be a survey but reliability is irrelevant for its function. Even though it cannot be measured if responses to the PUMBA would be consistent, I have evidence from student interviews that the responses are valid and representative of the students' agreement with the course at the time the survey is completed.

4.1.8 Calculating Student Agreement

Referring back to the operational definition model of a formative test, the construct is a mathematical combination of its measurements (Bagozzi & Fornell, 1982; Diamantopoulos & Winklhofer, 2001). In the PUMBA, there are four latent constructs: agreement with 1) the importance place on in-class activities, 2) the importance placed on out-of-class activities, 3) the time spent on in-class activities, and 4) the time spent on out-of-class activities. The measure of each construct depends on student agreement with the parts that comprise it. For the importance agreement constructs and the agreement with the time spent on out-of-class activities construct, the mathematical definition is shown in Equation 3. $SA_{ICI/OCI/OCT}$ is the student agreement measure for in-class importance (ICI), out-of-class importance (OCI), and out-of-class time (OCT). The overall student agreement for those constructs are equal to the proportion of activities which a student responded with "equally". For example, a student who indicates four in-class activities as being equally important in his HMEC as it seems to be in his actual class will receive an in-class importance agreement measure of 44 (i.e. 4 "equally" responses out of 9 activities). If a student agrees with the time spent on six out-of-class activities, their agreement measure for the time spent on out-of-class activities will be 75 (i.e. 6 "equally" responses out of 8 activities).

$$SA_{ICI/OCI/OCT} = 100 * \frac{Number\ of\ "Equally"\ Responses}{Total\ Number\ of\ Activities} \tag{3}$$

A small correction may be needed for student agreement with out-of-class activities. Students are given the option to state that they have no feeling regarding the time that the instructor intends for them to spend on an out-of-class activity. In such cases, the student cannot state if they would spend more or less time on that activity in their HMEC. To account for this in Equation 3, I find the proportion of out-of-class activities in which the student responded with "Equally…" out of the number of activities for which a student had a feeling for the amount of time they are intended to spend on it. In other words, the denominator, "Total Number of Activities", becomes "Total Number of Activities – Number of 'No feeling' Responses". This is essentially calculating student agreement ignoring activities that a student was uncertain about. However, I did not want to calculate student agreement based on too few activities because I felt that this might adversely affect the agreement score in comparison to other students; ignoring two out of eight activities, or 25% of the possible out-of-class activities, is too many to ignore as it increases the impact of a disagreement with one activity on the overall agreement measure. So, if a student had "No feeling" for two or more activities, they did not receive a score for their agreement with the time spent on out-of-class activities. Doing so eliminated 14% of the scores, across all distributions of the survey, for out-of-class activity time agreement. Ultimately, for the remaining students, I calculated agreement for the time spent on out-of-class activities based on either seven or eight activities.

The calculation of student agreement with the time spent on in-class activities differs from the other agreement measures because of its unique response method. Since students provide proportions for the time spent on an activity both in their actual class and their HMEC, student agreement for this construct is calculated more precisely (see Equation 4). The calculation for student agreement with in-class activities essentially adds together the overlap in the amount of class time spent on each activity between the student's HMEC and their actual class. For example, if a student says their HMEC would spend 50% of the class time on lecture and their actual class spends 30%, the overlap would be 30 percentage points. If they say that 35% of their actual class is spent on small group work and they would prefer 15% in their HMEC, the overlap is 15%.

Mathematically, the overlap is the minimum of the two responses for one activity. The overall agreement score is the sum of the overlaps for each activity, so the final score represents how much of the total time spent in their HMEC overlaps with their actual class. Like the other overall measures, this measure ranges from 0-100 with larger values indicating greater agreement.

$$SA_{ICT} = \sum_{i=1}^{9} Minimum(HMEC_i, Perceived\_Actual_i) \quad (4)$$

SA_{ICT} represents student agreement with the time spent on in-class activities. "i" indexes the nine in-class activities. $HMEC_i$ and $Perceived\_Actual_i$ represent the student's response to activity i for their HMEC and their actual class. I used "Perceived_Actual" to emphasize that this is not an objective measure of the time spent in their actual class. This is what the student feels is the proportion of class time spent on activity i. "Perceived_Actual" is a somewhat subjective measure but that does not decrease the validity of the agreement measure because their HMEC response is equally subjective. No matter if the student's response for "Perceived_Actual" is close to reality or not, the important matter is whether they wanted more or less of whatever they feel is happening in class.

These calculations for student agreement provide a measure of the agreement with the overall class format. However, to satisfy feature 4, the PUMBA needs to provide the means to determine student agreement with each class activity, which is done fairly simply. For all four agreement measures, I only need to determine the proportion of students who agree with an activity. For example, 20 students in a 30-student class may agree with the importance placed on student presentations. Thus, I would report to the instructor that 67% of their class agrees with the importance placed on student presentations. With the PUMBA, I can do more than report the proportion of students who agree with an activity; I can also say what students prefer in their most effective class. For example, 20% of the class may indicate student presentations would be less important in their HMEC, 67% may agree with the importance, and the final 13% may place more importance on student presentations in their HMEC. This information may help instructors adjust their discussion with their students to get more students to buy-in to student presentations. Regarding student agreement with the time spent on in-class activities, the PUMBA provides information on how much more or less of an activity students would have in their HMEC. When reporting these results, not only do I tell the instructor the proportion of the class who would rather have less lecture, for example, I also can say that the students want on average 10 percentage points less lecture. Again, this information allows the instructors to tailor their message to the students. Knowing the extent of the disagreement with an activity should help instructors know how much effort they may need to put in to get student buy-in.

4.2 The SIMBA

4.2.1 Overview

I created the Survey of Instructor Methods for Buy-in Achievement (SIMBA) to be taken by studio instructors (see Appendix C). The survey is taken online and typically requires five to ten minutes to complete. Instructors take the SIMBA at the beginning of the semester, after the first class but before the end of the second week, and at some point during the final two weeks of the semester. These time frames coincide with the times in which students take their survey, the PUMBA.

The SIMBA is a simple instrument to collect information about what instructors have done to explain the course format to their students. There are no latent constructs, no deeper meaning or idea, being measured by the SIMBA; its purpose is to reveal what instructors have done in their class. The SIMBA must be given after instructors have explained the class format to their students so that instructors indicate what did happen in class, instead of what they plan on doing in class. I want to explore what the instructors do in class that affects their students' agreement with the class, so plans, especially those that don't come to fruition, are unhelpful. At the end of the semester, instructors are asked what they have done since the last time they took the survey (i.e. excluding the first day particularly) to explain the course format to their students.

In the SIMBA, what the instructors do to explain the course format is broken down into three parts: 1) The specific activity the instructor is discussing with the students; 2) the ways in which the instructor tries to get students to agree with an activity; and 3) how the discussion is taking place (e.g. lecture (talking to the students), class discussion/activity, readings, etc.). As the instructors navigate the SIMBA, they should be asking themselves the following questions, in order:

1. Did I discuss with the students how much of an activity for class they should expect, how it will work, and/or how I expect them to behave during that activity? (Yes or No).
2. If yes, what did I do, if anything, to get the students to agree that adhering to this expectation is the best way to learn?
3. Through which format(s) did I have this discussion about those expectations?

The three-part thought process aligns with the three parts of an explanation about the course format. The instructors should go through this thought process for each activity. If an instructor did not discuss an activity, they can skip to the next activity. If they did discuss an activity, they can choose from various methods they might use to get students to agree with an activity. An instructor can choose any number of these methods including none at all. Following this, instructors must choose at least one discussion format. There must be at least one format chosen because if no formats were used, it implies there was no discussion and thus they did not actually talk about that activity. When taken online, the sections regarding the methods and formats used for a specific activity do not appear unless the instructor has indicated that they did discuss that activity. See Appendix C for a copy of the SIMBA, in which an instructor has indicated that they discussed lecture, but no other activities.

4.2.2 Development

In early versions of the SIMBA, what are now referred to as "activities" were called "expectations". There were two types of expectations: expectations the students should have of the class and expectations the instructor has of the students. These were what the instructor might discuss to get the students to understand the class format. All of the expectations the students should have of the class regarded specific class activities like group work, lecture, and student presentations. These "expectations" still survive as class activities an instructor might discuss to describe the course format. The other set of expectations, those about what the instructor expects of the students, regarded student behavior like their engagement in activities. However, the expectations organized as such made it difficult to analyze how student agreement is affected by instructors' discussions of the class format. The analysis on the effect of instructor methods to gain student agreement was done using a hierarchical multiple linear regression. Student agreement to an activity was used as the outcome variable with the different instructor methods regarding that activity as the multiple predictors. Methods used for discussing behavior expectations could apply to any activity and so would have to be included as predictors along with the predictor methods for the specific activity. This made the model too complex due to the large number of predictor variables.

Fortunately, I realized that when discussing what the instructor expects of the students, it should be with regard to a certain activity. For example, “You should be fully engaged with the group work” or “Students who do well take notes during lecture” are some things an instructor might tell their students that they expect of them. In both cases, the behavior expected of the students is in reference to a specific activity of the class (e.g. group work or lecture). Thus, a discussion about an activity might involve talking about the instructor’s expectations of the students. So long as I include that in the definition of a discussion, I only need to ask instructor what activities they discussed with their students (see the first question in the instructor SIMBA thought process). So, what were originally called “expectations” are now “activities”, and a discussion of an activity means to discuss what the students should expect of the activity and what the instructor expects of the students regarding that activity.

I chose the activities to align with the activities in the PUMBA. These activities, and their definitions as given in the SIMBA, are shown in Table 14. Instead of having two separate lab activities like the PUMBA, the SIMBA combines step-by-step labs and investigative labs into an all-encompassing lab activity. The PUMBA distinguishes between a more engaging investigative lab and the less engaging “cookbook” labs because students might prefer one type over the other and need the option to express that. No such distinction is necessary for the instructor survey. If the instructor, when describing the course format, discusses both lab types in terms of one being preferred over the other, than this is captured with the “comparison between reformed and traditional classes” option in the follow-up question. Otherwise, the labs are discussed in general as the labs for the class. Thus, there is no need for a differentiation between the two lab types as a class activity in the SIMBA. The other difference between the activities on the PUMBA and SIMBA is that all specific out-of-class activities are combined into one. I did this so as to not make the SIMBA too cumbersome in length.

Table 14. The SIMBA class activities and their definitions.

Activity	Discussing an activity means talking about how much of that action students should expect, how they should expect it to typically operate in class, and how you expect them to behave for that action.
Lecture	The instructor is talking to the whole class.
Quiz or tests	Students are completing a formal graded assessment.
Lab group work	Students are working in groups to conduct a lab.
Non-lab group work	Any work that the students do with the help of their classmates that has not yet been covered by the other listed aspects.
Individual work	Students are working individually without the help of their classmates.
Student Presentations	Students present their work to the entire class.
Demonstrations or simulations	Students are taking part in, or watching others engage in, a brief activity with the purpose of being able to observe a physical phenomenon.
Whole class discussion	Discussing scientific concepts with the whole class. Students contribute to the majority of the conversation.
Out-of-class work	Activities that students do for class but not during the class meeting time.

If an instructor indicates that he or she discussed a specific activity, two follow up questions appear. The first follow-up question asks what the instructor talked about to get students to agree with that activity. This question will be referred to as the “methods” section. The second follow-up question asks how the discussion took place and will be referred to as the “formats” section. I developed the responses for the methods section primarily from the Science Education Initiative’s (SEI) materials for framing an interactive classroom ("Framing the Interactive Engagement Classroom, Boulder: Science Education Initiative at the University of Colorado, Boulder," 2013), which is discussed in section 3.3.4. These materials exhibit numerous different methods an instructor might use to describe their reformed course.

To find all of the methods an instructor might use to get students to agree with a course format, I read through the materials provided by the SEI. I searched for emerging common examples of methods among the PowerPoints, written responses, and testimonials. Ultimately, I found eleven different examples of specific discussion actions or topics that may work to convince students to agree with the course format. The eleven examples became the eleven response options in the methods section of the SIMBA. However, after an initial administration of the surveys, it was found that these were too many options; any analysis on the connection between instructor methods and their students’ buy-in were too complex due to the large number of predictors. I then combined several sets of response options into broader method options, reducing the number of choices from eleven to six. The methods were combined based on their commonalities. Two options discussing the advantages of reformed classes and the disadvantages of traditional classes were combined into a “Comparison between reformed and traditional classes” since one option implies a comparison to the other. There were four original methods that focused on the students: comforting students, relating the class to the students’ interests, discussing student accountability, and asking the students to reflect on their own experiences. Each of these methods appeal to the students affect (their feeling or emotion) and thus can be described as a pathos argument in the rhetorical theory of argument (Aristotle, 1991). As each of these methods use the same type of argument and all appeal to the student, I combined them into the singular "student focused discussion" method. The final two combined methods were a discussion of how people learn and an appeal to research data or results. These are both arguments based on fact and logic and so are considered as a logos argument in the rhetorical theory (Aristotle, 1991) and so I combined them into one "evidence based discussion" method. The final six methods and the eleven original methods that comprise them are outlined in Table 15.

Table 155. The SIMBA instructor methods and the definitions of the original methods that comprise them.

Method	Definitions of original methods.
Advantages of engagement	The advantages of engaging in this class. A mention of how being engaged will be good for the students.
Comparison between reformed and traditional classes	The advantages of using reform based instructional strategies (RBIS) in class. Such strategies include but are not limited to: Peer Instruction, an emphasis on group work, a de-emphasis on lecture.
	The disadvantages of traditional type of class which typically employ teacher-centered approaches with passive student learning.
Student focused discussion	Encouraging and comforting students. Letting them know it is okay to be wrong or make mistakes in this setting.
	Discussing how this type of class is relevant to their interests or future job.
	Discussing how student engagement or lack thereof can help or hinder their classmates.
	Students are asked to think back on their own experiences and to reflect on what they know of themselves.
Extended example	An extended example to demonstrate what you are trying to say. The example used should have been planned into the discussion, referred to multiple times in the discussion, or both.
Evidence based discussion	Discussion of how people learn.
	Appealing to research data or results.
"Try it and you will see."	Stating that the students will understand or agree with the class from experiencing how the class works. An instructor may use this if they don't intend to have much discussion on other agreement topics but rather have the students experience the class to see its benefits first hand. There must have been an explicit mention of this to the students for it to be considered a topic discussed.

The formats, which are the physical means in which the instructors discuss the class format, were also primarily developed from emergent themes in the SEI materials. Instructors might simply tell the students what the format is and why they should agree with it, in which case the format would be “lecture”. If the instructor asks the students to reflect on their prior learning experiences, this is considered a “class activity”. Instructors may convey the information about the class format through other media, either text or audio/video. Throughout the semester, an instructor might begin a reformed activity with a reminder to the students about their discussion at the start of the semester explaining why they are doing such an activity. Lastly, an instructor might convey the idea of "this is how you should be acting for this activity" by rewarding appropriate behavior. This format is commonly seen when giving a small amount of credit to students who volunteer to do a presentation. The list of discussion formats and their definitions are found in Table 16.

Table 16. The SIMBA formats and their definitions

Format	Definition
Lecture	Talking at the students. The instructor the primary voice with little student input.
Class discussion/activity	Discussion between students or discussion guided by instructor but with heavy input from students.
Reading/syllabus	Students are given something to read.
Audio/Video	Using an audio recording or an audio/video to convey the information.
Reminders throughout the semester	Explicit references throughout the semester to prior discussion about the style of the class.
Rewarding appropriate behavior	Explicitly incentivizing appropriate behavior that matches what is expected in class. DO NOT select this if it is an incentive just for having a right answer.

The final section in the SIMBA asks instructors what they perceive is the proportion of class time spent on in-class activities as defined in Table 9. The in-class activities cover every action that might happen in a studio classroom and so the sum of the proportions should equal 100%. The responses to this section of the SIMBA is compared to their students' responses to the similar section in the PUMBA. When comparing responses from the surveys given at the beginning of the semester, the comparison can measure how well the students understand what the class should be like (i.e. the instructor's responses). For each student, the sum of the absolute differences between the instructor's and student's response for each item, divided by two and then subtracted from 100, provides a value for the proper understanding of the class format. The understanding calculation is similar to Equation 4, except in this case the comparison is between the student's and the instructor's perceived actual time spent (see Equation 5). The results of this calculation should be interpreted with caution; the responses come from two different people who will perceive identical things slightly differently. What the instructor might interpret as 20% of the class time, a student might call 25% of the class time. Though, when averaged across a class, there may be some overall trend that appears in the understanding data within the noise that results from the subjectivity of the measure.

$$Understanding = 100 - \frac{\sum_{i=1}^{9} |Student\_Perceived\_Actual_i - Instructor\_Perceived\_Actual_i|}{2} \quad (5)$$

4.2.3 Validity interviews

Throughout the development process of the SIMBA, I interviewed studio instructors to make sure they could understand the survey and that I correctly interpreted their results. With the instructor interviews, I obtained evidence of validity based on content, the survey responses' relation to other variables, and on the response process. Obtaining evidence of validity based on the response process means finding the fit between what is being measured and the nature of how the response is given (AERA et al., 1999). For example, an instrument measuring critical thinking skills should include items that require critical thinking to respond to. For the SIMBA, the ways in which instructors respond describing how they try to achieve student buy-in should reflect the actual ways in which they discuss class activities and why students should agree with it.

I interviewed a total of four studio instructors, some on multiple occasions. Each instructor taught at UCF. In several cases, the validity interview followed an administration of the SIMBA and PUMBA for the instructor's class. For some interviews, the instructor completed the survey during the interview, describing their thoughts as they took it. The questions asked for the instructor validity interviews can be found in Appendix D.

4.2.3.1 Evidence of Validity Based on Response Process

The first two questions, and question nine, of the instructor validity interview protocol helped me make an easy to understand survey. Typically, instructors did not misunderstand the questions on the SIMBA. From the responses to these questions, and the discussions that followed, I began to recognize a thought process that instructors used when interpreting the questions. In an early version of the SIMBA, if a certain class activity was discussed, instructors would be asked what they talked about to get students to agree with the expectations for that activity. They would then be asked what other methods they used when discussing the expectation or when trying to get students to agree with that expectation. My original idea was that there were expectations of an activity to be discussed, topics instructors might talk about to get students to agree with an expectation, and methods or strategies that an instructor might use when having these discussions. Throughout the validity interviews, I realized that there is a simpler mental model that instructors use when describing how they try to get students to agree with the class format. This mental model is the thought process that is described at the beginning of the SIMBA. I updated the questions of the survey to better reflect that mental model, which goes towards providing evidence of validity based on the response process.

4.2.3.2 Evidence of Validity Based on Content

Questions three through eight on the instructor validity interview protocol aim to provide evidence of validity based on the content of the survey. For each part (class activities, agreement methods, and formats), instructors first explain to me their understanding of what each response option means. I found that instructors were interpreting each option as I had intended. For example, on instructor correctly describes in-class demonstrations as, “typically, it would be either me, or the TA, or the LA, um, has some kind of physics apparatus, and we are demonstrating a particular concept through using- using that apparatus.” Another instructor described the “advantages of engagement” method based on how she used it, saying,

> *“I think what I do is just say like this is, uh, the type of activity, this is why it's useful and this is sort of what you'll get out of it and this is why we do it in small groups and this is why we do it, like we do working things out on the whiteboards as opposed to just on your piece of paper and then you can sort of compare and contrast.”*

A third instructor correctly describe the “lecture” format saying, “If I said, 'Okay ...’ You know, even if I'm like presenting, uh, evidence. So, I'm just saying, ‘Hey, look, here's this thing. It works. This is why it works. Please believe me. Give it a shot.’” These are just a few examples of instructors describing their understanding of each activity, method, and format. From the interviews, I found evidence that instructors were correctly understanding the response options on the SIMBA.

There were multiple instances where instructors were uncertain of the meaning of a response option at first, but when reminded that they could use their computer mouse and hover over the text of the response option for more information, it would always resolve their problem. Since I will not be there to remind instructors of this when they take the survey, I added reminders of this functionality of the survey with larger, bolded, and underlined font at multiple points in the SIMBA. In addition to the instructors' interpretation of each response option, I asked if they felt anything was missing from the survey or overlapped with other response options. Cases where an instructor felt as if something they have done might overlap with two response options, and so indicated in the survey, were always situations in which I agreed the instructor should have indicated both options. For example, one instructor tells his students that he has had previous students come back to him from their traditional style, second-semester physics course, and tell him that they enjoyed his reformed, first-semester physics class and learned more from it. This instructor felt, and I agreed, that this is both an example of a comparison between traditional and reformed classes and an evidence-based argument. For the SIMBA, it is okay for something to be represented in two or more options. It is not okay for something done in class to not be represented in the survey. Fortunately, no instructor felt that there was an activity, agreement method, or format missing from the SIMBA. Thus, I have evidence of validity based on content.

4.2.3.3 Evidence of Convergent Validity

The final question of the instructor validity interviews asked instructors to describe some of the things they do in class to get students to agree with the course format, and how they indicated that in the survey. The purpose of this question is two-fold; it allows me to make sure I can correctly interpret the responses on the survey and it provides evidence of validity based on the survey responses relation to other variables (i.e. their description of the course format discussion). Due to time constraints, instructors only provided a few examples in response to this question. By the final version of the survey, the instructor's descriptions of what they did in class matched how I would have expected them to respond on the survey based on that description. For example, one instructor described having his students listen to a story from National Public Radio (NPR) that is about a medical school using a flipped classroom and a professor who wrote a chapter on lecturing in a prominent teaching textbook. The instructor said, “[The story] would be something that I use to set up students’ expectations for the fact that this class is going to be less lecture, and more, um, applications based.” The instructor was describing what he did that led him to select the “comparison between traditional and reformed classes” method and the “Audio/Video” format when discussing lecture expectations. I agreed that what he told me should have prompted him to choose those response options. Another instructor, when asked to explain how he uses the “discussion about students” method regarding small group work, says, “Trying to level with the students by telling them, ‘look, physics is hard. The reason that we have you do these small group things, and that we assign homework, is that you have the opportunities to mess up in a non-stressful environment so that you can then succeed on the exams…So the reason that we are doing all these things is so that you have the opportunity to make mistakes now instead of later.’” Again, based on that description, I agreed that he used the “discussion about students” method for the “small group work” activity. As shown in these examples, and is true in the many other examples obtained from interviews, the way the instructors respond on the SIMBA are befitting of the scenarios the instructors described.

Through the instructor interviews, I have found evidence of validity based on content, response process, and the survey responses’ relationship to other variables. I feel confident that instructors are interpreting the SIMBA correctly, and that I am correctly interpreting their responses. Like the PUMBA, reliability is also not a factor for the SIMBA. Unlike the PUMBA, the reason that reliability is not measured with the SIMBA is simply that the SIMBA does not measure a latent variable and thus there is no need for reliability. All the SIMBA does is survey what the instructors have done in their class and how they perceive the class time is spent on various in-class activities.

4.3 Summary

In this chapter I described the two surveys I developed to collect the majority of the data for this study. The student survey, the PUMBA, asks students to imagine a hypothetical most effective physics (HMEC) class. Then, the PUMBA asks students to compare the importance and time placed on various in-class and out-of-class activities between their HMEC and their actual class. The difference between the HMEC responses and the actual class responses measures how well students agree with each of the activities and with the class format overall. I continued on to describe the major revisions to the PUMBA as I developed it. Revisions included: incorporating the importance placed on activities in addition to the time spent on activities, fine-tuning the response method (rankings, proportions, time in minutes, etc.), and changing the order in which activities were presented to the students. I then discussed how I validated the PUMBA through student interviews and how reliability measures were inappropriate for this type of survey. Finally, for the PUMBA, I described how I calculated agreement.

Next, I describe the instructor survey, the SIMBA, which surveys the strategies instructors used to discuss the class format with their students. Specifically, the SIMBA asks instructors which class activities they discussed, which methods they used for the discussion, and through which means the discussion took place. I continue on to describe how I collected evidence of validity for this survey through interviews with instructors.

CHAPTER FIVE: METHODS

In this chapter, I describe the methods with which I carried out the research. First, I briefly describe the design of the study and provide a timeline of the research. Next, I provide a description of the participants in my research studies, including: students who took the PUMBA; students I interviewed; instructors who took the SIMBA; and instructors I interviewed. I also describe the institutions represented by these students and instructors. Then, I outline the interview protocols used for the second set of interviews I conducted, which I used to learn about students' and instructors' experiences and opinions of their studio classes. I describe the methods used for coding and analyzing these interviews. Lastly, I provide an overview of the analysis methods I will used on the survey data.

5.1 Study Timeline

Like many other studies using the instructional communication framework (Conley & Ah Yun, 2017), I primarily collected data through surveys. Also, similar to other instructional communication research, the population surveyed is comprised of undergraduate students and their instructors. I distributed the SIMBA and PUMBA at the beginning and end of the Fall 2017 and Spring 2018 semesters. Figure 7 illustrates the survey distribution timeline within a semester. The timeframe in which a student or instructor could take their beginning-of-the-semester survey started after the first class and ended at the end of the second week of class. Students must take the surveys after they have been oriented to the class format, thus the survey cannot be taken before the first class. The two-week timeframe allows instructors and students ample time to complete the survey without extending too far into the semester when the class format may begin to change. Students and instructors also have two weeks to complete their surveys at the end of the semester. Two weeks prior to the last day of class, the SIMBA and PUMBA open to the instructors and students. The final day students could complete the PUMBA was the last day of class in the Fall semester, and due to time constraints, the day before the final exam in the Spring semester.

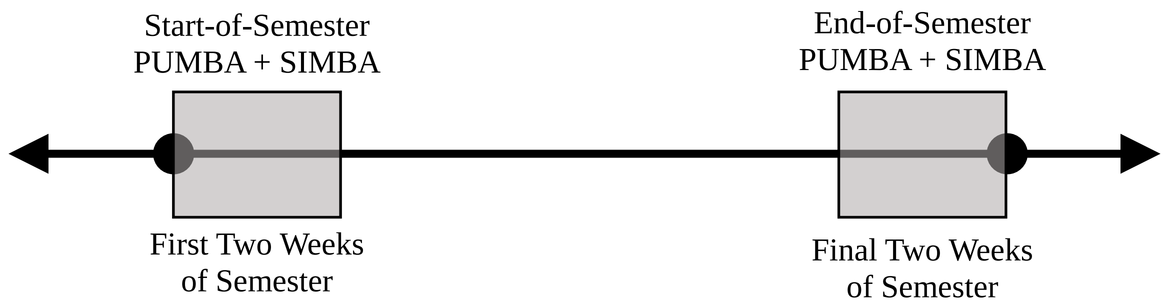


Figure 7. The distribution timeline of the surveys within a semester.

5.2 Participants

In Fall 2017, nine instructors teaching twelve classes participated in this research. In Spring 2018, five instructors teaching six classes participated. See Tables 17 and 18 for information about the instructors and their classes. Table 17 also outlines the few instructors who did not fully participate in taking or distributing the surveys. Table 18 summarizes the student demographics of each class, which are gender, race/ethnicity (well represented or underrepresented in STEM fields), age, studio experience, and college experience. Students whose race/ethnicity is well represented in the STEM fields are those who are White non-Hispanic or Asian. All other races and ethnicities are considered underrepresented in STEM fields. The five instructors to participate in the Spring also participated in the Fall. No instructor taught a mechanics class in the Fall and an E&M class in the Spring, so there should be no overlap in student populations. Some instructors from the Fall semester did not continue to participate in this research because they were no longer teaching a studio class. I contacted other instructors to recruit them to the study but was unsuccessful in gaining any new instructors.

You will notice that the nine instructors have several commonalities. Specifically, most are men and all that chose to indicate their race/ethnicity were White non-Hispanic. Additionally, all instructors were less than 40 years old (this information is not reported by instructor to preserve the confidentiality of their identities). Due to the many similarities between the instructors, any results obtained from the instructors may not be generalizable to those outside the young, predominately male, White non-Hispanic instructor demographic.

Table 17. The instructors who participated in this study. Studio experience recorded at the start of the Fall 2017 semester.

Instructor	Gender	Studio Experience (Semesters)	Racial/Ethnic Background
1	Woman	1	White non-Hispanic
2	Man	0	White non-Hispanic
3	Man	4	White non-Hispanic
4	Man	8	Preferred not to answer
5	Woman	2	White non-Hispanic
6	Man	4	White non-Hispanic
7	Man	1	White non-Hispanic
8	Man	0	White non-Hispanic
9	Man	6	White non-Hispanic

Table 18. Summary of the classes participating in the study.

Class ID	Instructor	University	Course	Number of usable student responses (pre/post/matched)	Survey Participation
Fall 2017					
1	1	A	Algebra EM	34/48/19	Full participation.
2	2	A	Calculus Mechanics	49/29/16	Full participation.
3	3	A	Algebra Mechanics	60/56/41	Full participation.
4	4	A	Algebra Mechanics	44/4/4	No SIMBA at the end of the semester.
5	5	B	Calculus Mechanics	16/18/9	Full participation.
6	5	B	Calculus Mechanics	8/14/2007	Full participation.
7	6	C	Calculus Mechanics	24/40/17	No SIMBA at the end of the semester.
8	7	C	Calculus Mechanics	40/46/30	Full participation.
9	7	C	Calculus EM	36/44/28	Full participation.
10	8	D	Calculus EM	22/41/9	Full participation.
11	9	D	Calculus Mechanics	25/-/-	No SIMBA or PUMBA at the end of the semester.
12	9	D	Calculus Mechanics	39/-/-	No SIMBA or PUMBA at the end of the semester.
			Total Number of Students	397/340/180	
Spring 2018					
13	1	A	Algebra EM	67/61/49	Full participation.
14	2	A	Calculus Mechanics	39/44/28	Full participation.
15	4	A	Algebra Mechanics	51/29/20	Full participation.
16	7	C	Calculus Mechanics	50/44/32	Full participation.

Class ID	Instructor	University	Course	Number of usable student responses (pre/post/matched)	Survey Participation
17	8	D	Calculus Mechanics	30/27/14	Full participation.
18	8	D	Calculus Mechanics	24/29/16	Full participation.
			Total Number of Students	261/234/159	

In total, there were 397 usable student responses to the PUMBA at the beginning of the Fall semester and 340 usable student responses at the end. There were 180 matched responses from the beginning to the end of the semester resulting in a total of 557 unique students who had taken a version of the PUMBA at some point in the Fall. I recruited students were through an email sent by their professor, by an announcement made by their professor during class, or both. Often, instructors offered a small amount of extra credit to incentivize students completing the survey.

For the Spring 2018 semester, there were 261 usable student responses at the beginning of the semester and 234 usable responses at the end of the semester. For this semester, to incentivize students taking both surveys, increasing the number of matched responses, I asked instructors to only offer credit to students completing the PUMBA at the beginning and end of the semester. Ultimately, 159 students took the survey at the beginning and end of the Spring semester, meaning 337 unique students took the PUMBA at some point. Although the proportion of students to complete the survey at both points in the semester was higher for the spring semester (47% in the Spring versus 32% in the Fall), fewer participating classes resulted in a similar total number of matched students.

The demographic information of the students in each class is summarized in Table 19. Overall, there were slightly more men than women. About 70% of the students identified with a race/ethnicity that is well represented in STEM fields, either White non-Hispanic or Asian. The average student age was 20 years old, about 20% of the students had taken a studio style class before, and a little more than 40% of the students are in their third or more year of college.

During and after the Spring 2018 semester, I interviewed instructors and students about strategies to gain student agreement and their experiences and opinions of the studio format. The interview protocol will be discussed in detail in the next section. Five instructors participated in the interview: instructors 1, 2, 4, 7, and 8. I interviewed instructors 1 and 7 after the Spring semester with questions regarding both the Fall and Spring semester. The other instructors I interviewed at the beginning of the Spring semester with questions regarding the Fall semester and beginning of the Spring semester. After the Spring semester, I interviewed nine students from these instructors' Spring classes. Ultimately, I interviewed two students from each instructor's class, except I only interviewed one student from instructor 8's class.

Table 19. Demographic breakdown of the students in each class.

Class	Number of Students	% Men	% Women	% Well Represented in STEM (White non-Hispanic or Asian)	% Underrepresented in STEM	Average Age	% With Studio Experience	% With 3+ Years of College
Fall 2017								
1	63	16%	84%	56%	37%	22	27%	100%
2	62	74%	24%	69%	27%	20	3%	40%
3	75	24%	76%	56%	39%	21	12%	83%
4	44	36%	61%	57%	34%	22	9%	73%
5	25	56%	44%	52%	48%	21	20%	52%
6	15	100%	0%	80%	20%	20	27%	13%
7	47	60%	40%	79%	15%	19	4%	11%
8	56	73%	27%	89%	11%	20	14%	13%
9	52	73%	25%	77%	15%	19	48%	29%
10	54	72%	28%	89%	11%	20	70%	37%
11	25	60%	36%	92%	4%	19	16%	20%
12	39	62%	38%	87%	10%	18	21%	8%
All Fall Classes	557	55%	45%	72%	24%	20	23%	45%
Spring 2018								
13	79	30%	70%	63%	25%	22	28%	82%
14	55	64%	36%	64%	29%	19	4%	15%
15	60	27%	73%	47%	43%	22	12%	80%
16	62	69%	29%	81%	16%	19	8%	10%
17	43	63%	37%	98%	0%	19	30%	16%
18	38	76%	24%	84%	16%	20	26%	8%
All Spring Classes	337	52%	48%	70%	23%	20	18%	41%

5.3 Interviews

To support the findings of the SIMBA and PUMBA, I conducted interviews with instructors and students. I used the findings from the interview to triangulate the survey responses and to explore the possible reasons for a shift in student agreement. I conducted one-on-one semi-structured interviews, either in person or on the phone, which would typically last up to one hour.

5.3.1 Instructor Interviews

The instructor interview protocol can be found in Appendix E. When creating the instructor interview protocol, I had several topics I wanted to explore with the instructor. First, I asked instructors a few questions to ensure they understood their survey, the SIMBA. Although the SIMBA was previously validated, these questions could reveal any new understandings that might affect how their answers should be interpreted. Question 3, which asks instructors to explain something they did in class and how that led to a response on their survey, not only verifies my interpretations of the instructors' responses, but also gets instructors to start remembering the various ways in which they tried to generate student agreement.

Next, I asked several questions to explore the instructor's perceptions of student resistance. Many instructors cite student resistance as a barrier to the implementation of research-based instructional strategies (Dancy & Henderson, 2005; DeBeck et al., 2010; Froyd et al., 2013; Henderson & Dancy, 2007). As a beginning exploration into instructor perceptions of student resistance, for question 4 I asked instructors how much resistance they expected when they first taught a studio course and what about the class format they thought students would resist. Further, I asked if these expectations have changed and what led to any changes. My final questions about perceived student resistance asked about student resistance to the class format that the instructor has experienced (as opposed to expected) and if student resistance changes over the semester. .

Following the student resistance portion of the interview, I asked the instructors about the methods they used to generate student buy-in. Beginning at question 8, I asked instructors to describe some of the ways in which they attempt to get student buy-in. To help the instructors respond to this question by jogging their memory, I provided them a summary of their past responses to the SIMBA. On this summary, I pointed out their specific responses for which I would like them to describe what they did in class that led to that response on the SIMBA. The responses which I asked the instructor to elaborate on were methods and formats that were not commonly used among the other instructors. For example, few instructors indicated they used an evidence-based argument or used an audio/video format to have the discussion. So, for those that did use that method or format, I asked them to describe what they did. Next, I asked the instructors which methods and formats, regarding which activities, they felt were successful or unsuccessful in getting student buy-in. Then, the instructor and I would spend some time looking at their students' responses on the PUMBA. I would specifically point out activities for which the instructor lost or gained a large amount of agreement and ask if they had any ideas about why agreement changed for that activity. I would also point out activities that were consistently high or low in student agreement, and ask the instructor if they had any ideas about students' agreement with those. The final question in this section focused on where the instructors learned about the methods or formats they used to get buy-in. With this information, I could have a set of resources for ideas on how to gain student buy-in that I could provide to other instructors.

The last topic I explored was changes in instruction. I asked the instructor if, in either semester, they had made changes to their class format within a semester and about the role of student resistance in the decision to make those changes. Their response to this question also indicated whether a change in instructional style was a possible cause for a change in student agreement over the semester.

5.3.2 Student Interviews

The student interview protocol can be found in Appendix F. In general, the questions in the student interview asked about student experiences and opinions of the class as well as student perceptions of what the instructor did or could do to increase student buy-in. The first three questions probe the student's experiences at the start of the semester. First, I asked what the student expected from the class and how the instructor explained the class format to them. I also asked if the instructor did anything to get the students to agree with the class format and if the interviewee felt those strategies effective at getting them to agree with the format. These questions triangulate the student's responses on the PUMBA, particularly to the first set of survey questions which ask the student which activities their instructor discussed with them and which of those their instructor tried to get them to agree with. Also, in question 3, (Did your instructor do anything to get you to agree with the format at the start of the semester? Did this help you agree with the class format?) the second part begins to qualitatively explore which instructor methods affect student buy-in. The next question, which asks if the class ended up being different than what they were initially told to expect, provides an opportunity for students to describe class activities that did or did not meet their expectations and if they were happy or unhappy about their class experience not meeting their expectations (i.e. a positive or negative expectancy violation like that in Gaffney et al. (2010). The following three questions asked students to identify activities they agreed or disagreed with and any activities in which their agreement changed over the semester. Follow-up questions to the student's responses asked students to explain why they agreed or disagreed with the activities. Question 8 asked if the instructor did anything throughout the semester to get the interviewee to agree with the class format. This is another check of the student's responses on their survey.

Questions 9 through 11 explore the effectiveness of the instructor's actions to get student buy-in and what might work, from the student's perspective, to get student buy-in. First, students were given the opportunity to express their own ideas of what might convince them to buy in to the class. Afterwards, students were given the list of instructor methods and formats and asked which of those options they felt would gain their buy-in. Finally, the last three questions explored effective argumentation for student buy-in using the lens of rhetorical theory, which claims that instructors must establish their credibility before they can persuade the students to buy-in to the class format, they must first establish their credibility. McCroskey and Teven identified a credible source as someone who appears competent, caring, and of good character (1999). To see if the instructor has established their ethos, as Aristotle called it, I asked the students if they perceived their instructor as competent, caring, or of good character, and if this changed over time. Lastly, I asked which of the other two forms of argument, pathos (appealing to emotion) or logos (appealing to logic) is more likely to convince a student of something. These types of argumentation can be loosely applied to the methods instructors may use to get student buy-in. For example, the "discussion about students" method is largely a pathos argument and the "evidence-based discussion" method is a logos argument.

5.3.3 Interview Analysis and Inter-Rater Reliability

My goals through these interviews were to determine the strategies instructors actually used to explain the class format and to gain student agreement, from both the instructor and student perspective. I also aimed to determine which methods were effective at influencing student agreement and which methods could be effective. To do this, I gave instructors and students the opportunity to describe for themselves which methods or formats are or would be effective, and I also allowed them to choose from the pre-made list of methods and formats (Tables 15 and 16). When students were not referring to the pre-made list of methods and formats, a degree of inference must be made on my part when interpreting their responses and which methods or formats they are classified as. Classifying the responses to certain instructor methods or formats or other common response themes is known as "coding".

I began the coding process by reading the interview transcripts and recording the ideas or themes of common responses. Many of the responses did fit into one of the already developed and validated "codes" (i.e. the instructor methods and formats found in Tables 15 and 16). There were several additional themes found among the student responses that students used to explain how instructors tried to explain the class or get their agreement, or to explain what affected their agreement at the start of the semester or agreement change throughout the semester. These themes were developed into five new codes defined in Table 20. One code is an additional instructor method in which the instructor discussed the general benefits of a class activity without referring to anything that might be considered another method code (e.g. without specifically referring to engagement or comparing the benefits to a traditional class). The other four new codes are additional influencers on student agreement that are not necessarily an aspect of a discussion or activity an instructor can use to try to gain student buy-in. The first two new codes are based on the students' experiences either before the semester or developed through the course of the class. Student agreement may be affected by student experiences prior to the studio class or through the successes or failures in their current class. It was also common for students to say that the instructor had no effect on their agreement, so a code was created to reflect this response. The final new code covers student agreement influenced by a student's trust or faith in the instructor. A student may trust the instructor's choices for the class format leading the student to agree with however the instructor describes the format.

Table 20. New codes for student interview analysis.

Student Interview Code	Definition	Example Quote
Method-Benefit	Generally stating that some aspect of the class will be beneficial to the students. Do not code if the same statement is coded as another method (e.g. advantages of engagement, comparison between reformed and traditional).	"Our instructor, uh, very much explained that it would be very beneficial to go to the TA or a tutor…"
Prior Knowledge/ Experiences/ Opinions	Agreement based on what students know of themselves, based on past experiences in other classes, or based on previously developed opinions of what makes an effective class. Code this only if the knowledge was obtained before the semester.	"So obviously I don't want to like, judge [the instructor's] own like, style of teaching, because she's been teaching for a while. But a lot of [my agreement] is also on my own personal opinion on how I learn."
Current Class Experience	Agreement based on experiences with their current class (i.e. whether they feel an aspect of the class is working for them or not based on their experiences/opinions of it).	"Near the end, we didn't do [reading beyond what is assigned], so because we didn't do that, I had to put equal [importance]."
No Instructor Affect	An explicit mention that the instructor did not influence their agreement with an aspect of the class format.	Interviewer: "Was any of [your agreement] affected by anything your instructor did at the start of the semester?" Student: "Uh, no, but that's just because I didn't really have a- a good feel for the class yet."
Trust/Faith in Instructor	Agreement with the class format because the student trusts that the instructor knows what they are doing, is looking out for the students' best interests, etc. Typically, if a student trusts in the professor, the instructor only needs to describe how the class will work to gain agreement with the class.	"I thought [the instructor] was a nice person, and educated. So I mostly just followed it because he kn-knows how to teach, and he's trying to help us."

Applying these new codes and the previously established instructor method and format codes required interpretation of the interview responses. To ensure that my interpretations of interview responses are consistent with how others would interpret the responses, I must achieve inter-rater reliability (IRR) with another person when applying these codes to the interviews. IRR is “a way of quantifying the degree of agreement between two or more coders who make independent ratings about the features of a set of subjects.” (Hallgren, 2012) Doing so ensures that I am applying the codes in a reliable way (Rose, Spinks, & Canhoto, 2014). After developing the codes, I followed the IRR procedure recommended by Burnard (1991) and Graneheim and Lundman (2004). First, I trained a colleague on the codes using exemplary quotes from several interviews. Once my colleague felt confident that she understood the definitions and applications of the codes, we separately analyzed one interview and then discussed the differences between our codes to come to a consensus. After coding an interview, we quantitatively measured our coding agreement using Cohen’s Kappa (Cohen, 1960), which is a measure of agreement between two raters that accounts for chance agreement. A Cohen’s Kappa value of 0.8 or more is considered an acceptable level of agreement between two raters (Landis & Koch, 1977). I achieved a Cohen’s Kappa of 0.81 with my colleague, indicating that we have an acceptable level of agreement on the application of these codes and that I am applying these codes in a reliable way. Once I achieved IRR, I coded the remaining interviews.

5.4 Data Collected from the SIMBA and PUMBA

5.4.1 Variables

Tables 21 and 22 list the variables measured on the PUMBA and SIMBA, respectively. For both surveys, each variable is collected at the beginning and end of the semester (except for student expected final grade, which is only collected at the end) so time is another variable. From the PUMBA, there are four overall agreement scores, measured for each student, regarding agreement with the importance or time spent on in-class or out-of-class activities. For each class, there is a value for the proportion of students who agree with each individual activity in terms of the importance or time spent on that activity. Also, the PUMBA quantifies how much more or less of some in-class activity students want. At the beginning of the PUMBA, students state which activities their instructor discussed with them and which of those the instructor tried to get their students to agree with. Student responses to these two questions I refer to as “perceived discussion” and “perceived agreement discussion” variables. Each of these quantities are measured at the beginning and end of the semester, so time is a variable as well. The SIMBA yields 117 dichotomous variables that consist of which methods or formats the instructors used regarding which activities. For example, one of these 117 variables is whether or not an instructor discussed the advantages of engagement when explaining the role or expectations for small group work. Another variable would be whether or not an instructor used a class discussion or activity when discussing the role or expectations for whole class discussion. There are nine activities and twelve methods or formats. Including a variable that is simply whether or not an activity was discussed at all results in 9 x 13 (or 117) variables.

In both the SIMBA and PUMBA, the respondent provides demographic information, such as gender, race/ethnicity, age, and studio experience. For each student, on the end-of-the-semester PUMBA, I ask for their expected final grade. If it was available, I received from the instructor their students’ scores on concept tests, like the FCI (Hestenes et al., 1992), CSEM (Maloney et al., 2001), and attitudinal surveys like the CLASS (Adams et al., 2006).

The independent variables, or “predictors”, are organized into a two- or three-level nested hierarchical structure, as illustrated in Figure 8. For my data, the top level contains variables corresponding to the class, such as the instructor’s demographics and the methods or formats the instructors used to discuss class activities. The second level consists of variables corresponding to students, such as student demographics. When appropriate, the lowest level contains the time of measurement variable.

Table 21. The variables collected from the PUMBA.

Variable or Predictor	Description	Source	Data Type	Values
Overall In-class Activity Importance Agreement	Total agreement with the importance placed on in-class activities.	Calculated from importance agreement for each in-class activity.	Continuous	0-100
Overall Out-of-class Activity Importance Agreement	Total agreement with the importance placed on out-of-class activities.	Calculated from importance agreement for each out-of-class activity.	Continuous	0-100
Overall In-class Activity Time Agreement	Total agreement with the time spent on in-class activities.	Calculated from time agreement for each in-class activity.	Continuous	0-100
Overall Out-of-class Activity Time Agreement	Total agreement with the time spent on out-of-class activities.	Calculated from time agreement for each out-of-class activity.	Continuous	0-100
Individual In-class Activity Importance Agreement (for each of 9 activities)	Importance agreement for each in-class activity.	Proportion of in-class activities for which the student responded with "Equally important".	Continuous	0-100
Individual Out-of-class Activity Importance Agreement (for each of 8 activities)	Importance agreement for each out-of-class activity.	Proportion of out-of-class activities for which the student responded with "Equally important".	Continuous	0-100

Variable or Predictor	Description	Source	Data Type	Values
Individual In-class Activity Time Agreement (for each of 9 activities)	Time agreement for each in-class activity.	Proportion of in-class activities for which the student responded with "Equally important".	Continuous	0-100
Individual Out-of-class Activity Time Agreement (for each of 8 activities)	Time agreement for each out-of-class activity.	Proportion of out-of-class activities for which the student responded with "Equally important".	Continuous	0-100
In-Class Activity Time Difference (for each of 9 activities)	Difference in proportion of class time spent on each activity between a student's HMEC and their actual class.	HMEC - Perceived_Actual	Continuous	0-100
Student Perception of Discussed Activities	The activities for which a student perceived that their instructor discussed regarding the class format.	Response on the survey	Categorical	Each class activity
Student Perception of Discussion to Agree with Activities	The activities for which a student perceived that their instructor discussed why they should agree with it.	Response on the survey	Categorical	Each class activity
Student Gender	Gender that the student identifies as.	Response on the survey	Dichotomous	Woman (0) or Man (1)
Student Race/Ethnicity Representation	Race and ethnicity that the student identifies as. "Well-represented" race/ethnicities in STEM fields are white non-Hispanic and Asian.	Response on the survey	Dichotomous	Under-represented (0) or Well-represented (1) in the STEM fields
Student Age	Age of the student.	Response on the survey	Continuous	18+

Variable or Predictor	Description	Source	Data Type	Values
Student Studio Experience	If the student has taken a studio class prior to the current semester.	Response on the survey	Dichotomous	No studio experience (0) or Any studio experience (1)
Student class standing	Years of college experience.	Response on the survey	Dichotomous	First- or second-year (0) or third or more year (1)
Expected Final Grade	The grade the student expects to receive for their class (only measured at the end of the semester).	Response on the survey	Categorical	A, B, C, D or lower

Table 22. The variables collected from the PUMBA.

<table>
<tr><th colspan="3">Variable or Predictor</th><th>Description</th><th>Source</th><th>Data Type</th><th>Values</th></tr>
<tr><td>Discussed</td><td rowspan="14">For each of…</td><td rowspan="2">Lecture</td><td rowspan="14">The methods and formats used when discussing each activity. 117 total variables.</td><td rowspan="14">Response on the survey</td><td rowspan="14">Dichotomous</td><td rowspan="14">Method/Format was used (1) or not used (0)</td></tr>
<tr><td>Methods</td></tr>
<tr><td>Advantages of engagement</td><td>Quiz</td></tr>
<tr><td>Comparison between reformed and traditional</td><td rowspan="2">Labs</td></tr>
<tr><td>Discussion about students</td></tr>
<tr><td>Evidence-based discussion</td><td rowspan="2">Non-lab group work</td></tr>
<tr><td>Extended Example</td></tr>
<tr><td>"Try and see"</td><td rowspan="2">Individual Work</td></tr>
<tr><td>Formats</td></tr>
<tr><td>Lecture</td><td rowspan="2">Student Presentations</td></tr>
<tr><td>Class discussion/ activity</td></tr>
<tr><td>Reading/ syllabus</td><td rowspan="2">Demonstrations</td></tr>
<tr><td>Audio/ video</td></tr>
<tr><td>Reminders throughout the semester</td><td>Whole class discussions</td></tr>
</table>

Variable or Predictor			Description	Source	Data Type	Values
Rewarding appropriate behavior						
Proportion of class time spent or will be spent on activity (for each of 9 activities)			The proportion of class time the instructor plans to spend on each activity (at the start of the semester) or did spend on each activity (at the end of the semester).	Response on the survey	Continuous	0-100
Instructor Gender			Gender that the instructor identifies as.	Response on the survey	Dichotomous	Woman (0) or Man (1)
Instructor Race/Ethnicity Representation			Race and ethnicity that the instructor identifies as. "Well-represented" race/ethnicities in STEM fields are white non-Hispanic and Asian.	Response on the survey	Dichotomous	Under-represented (0) or Well-represented (1)
Instructor Age			Age of the instructor.	Response on the survey	Continuous	18+
Instructor Studio Experience			Number of semesters that the instructor has taught a studio class prior to the current semester.	Response on the survey	Continuous	0+

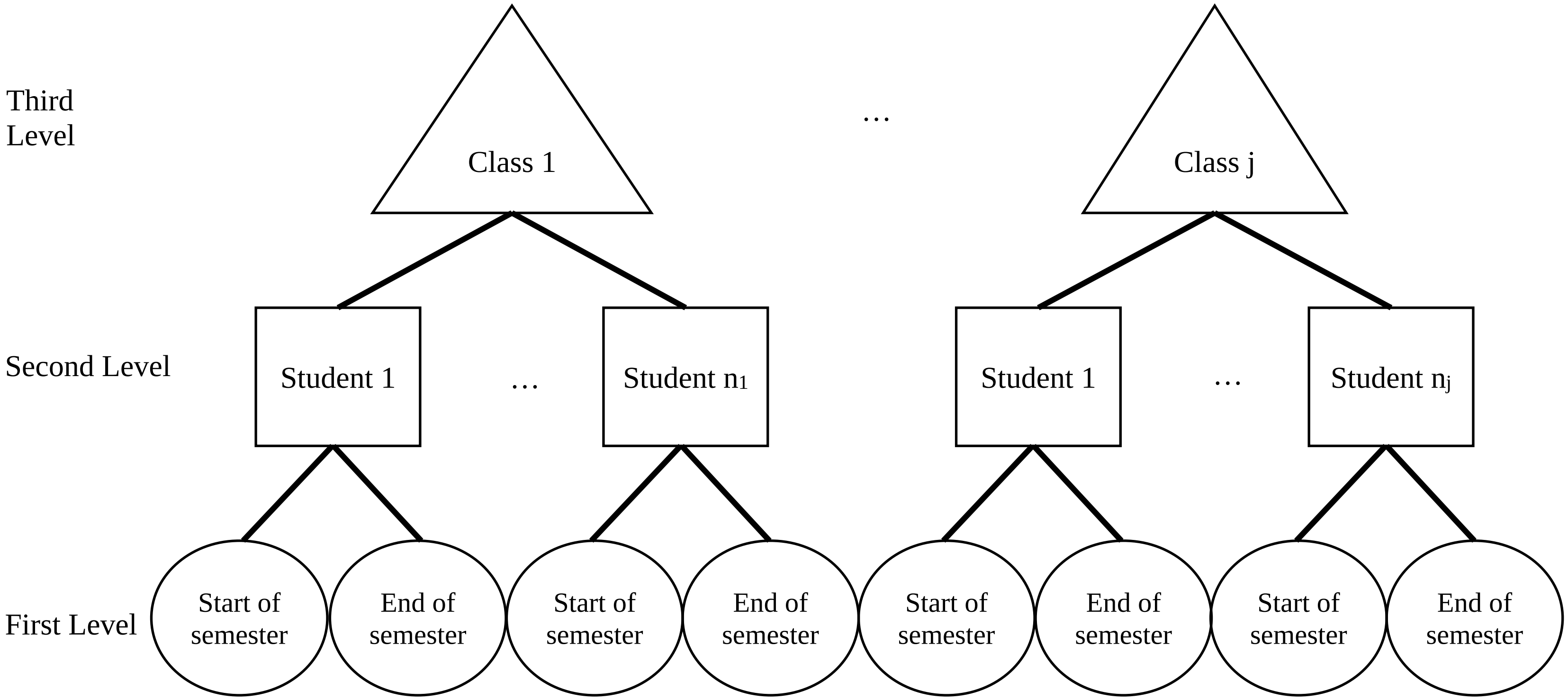


Figure 8. Illustration of a three-level nested hierarchical organization of variables.

5.4.2 Data Cleaning

Although there were 1426 instances of a student beginning the PUMBA in the Fall 2017 semester alone, only 737 of those possible responses remained as useable for analysis. Many responses went either unfinished or were completed after the two-week deadline had passed. When eliminating those responses, there were 1066 completed surveys for the Fall 2017 semester and 687 completed surveys in the Spring 2018 semester. These surveys then went through several checks before being deemed as a useable response, as shown in Table 23; the checks were applied in the order in which they appeared in the survey. First, I removed the students who did not consent to participating in the research. For the Fall 2017 semester, this was the filter which eliminated the most students. Second, I eliminated responses in which the student indicated that they did not fully read the definitions of the class activities. For each survey administration, only about 5% of students who consented stated they did not read the definitions; I removed these students' responses because I cannot be certain that they understood the activities. Next, I placed two attention checks in the middle of the PUMBA. One is placed in the out-of-class importance agreement section and asks students to respond with "Equally Important". The second is placed in the out-of-class time agreement section and asks students to respond with "More time". If a student is randomly responding in order to finish the survey and get the credit, there is a chance that the student will incorrectly respond to these checks. If students failed on either of these responses, I eliminated them from the pool of useable responses. In all survey administrations, the attention checks eliminated about 10% of the respondents who had passed all prior checks. For the Spring semester, the attention checks were the largest eliminating factor. Finally, I checked how long it took students to complete the survey. If students completed the survey in less than five minutes, I eliminated their responses. While students can complete the survey quickly, with the median survey time between 9 and 10 minutes for students who pass all checks, a respondent who completed the entire survey in five minutes would not have taken enough time to carefully read through and consider a response to all survey items. Considering that completing the PUMBA was often incentivized with a small amount of extra credit, it is possible that students rushed through the survey simply to receive the credit.

Table 23. The number of students remaining after each filter is applied.

Beginning of Fall 2017	Cumulative number of remaining responses	% of initial	End of Fall 2017	Cumulative number of remaining responses	% of initial	Total Fall 2017	Cumulative number of remaining responses	% of initial
Finished	557	100	Finished	509	100	Finished	1066	100
Consent	469	84	Consent	455	89	Consent	924	87
Read Definitions	444	80	Read Definitions	438	86	Read Definitions	882	83
Attention Checks	410	74	Attention Checks	384	75	Attention Checks	794	74
5+ minutes	397	71	5+ minutes	340	67	5+ minutes	737	69
Beginning of Spring 2018	Cumulative number of remaining responses	% of initial	End Spring 2018	Cumulative number of remaining responses	% of initial	Total Spring 2018	Cumulative number of remaining responses	% of initial
Finished	359	100	Finished	328	100	Finished	687	100
Consent	331	92	Consent	304	93	Consent	635	92
Read Definitions	313	87	Read Definitions	289	88	Read Definitions	602	88
Attention Checks	281	78	Attention Checks	257	78	Attention Checks	538	78
5+ minutes	260	72	5+ minutes	238	73	5+ minutes	498	72

During the Spring 2018 semester, I realized that students were still making the "time-importance mistake" (first discussed in section 4.1.4.3) despite my inclusion of a prompt explicitly addressing the issue. I first encountered this mistake in the validity interviews for the PUMBA and addressed it in the survey before beginning the official distribution of the PUMBA for data collection in Fall 2017. However, it was not until Spring 2018 that I realized that based on the survey results, students continued making the mistake throughout the official distributions of the PUMBA. In the section of the PUMBA where students respond based on the importance of in-class activities, some students say that certain activities that they later indicate did not happen would be less important in their most effective class than it seems it will be or that it was in their actual class. If an activity did not or will not occur in class, then logically it would have no importance. Thus, the students are indicating that in their HMEC, an in-class activity would have less importance than the no importance it has in their actual class, which cannot be. After encountering the mistake in the validity interviews, I added explicit instructions to avoid it in each section of the PUMBA. For example, in the in-class importance agreement section of the survey I explained, "If an activity does not occur in your actual class, consider it to have no importance in your actual class (i.e. your response should be either "Equal…" or "More…")." Unfortunately, 54% of the students who passed every other check made this mistake on at least one in-class activity; 23% of the students that passed every other check made this mistake for two or more in-class activities. This is concerning because it is an indicator that students did not understand this aspect of the survey. Fortunately, with the in-class activities, I can determine which specific activities each student made this mistake on, and adjust the calculation of student agreement with the importance placed on in-class activities in a similar manner to the adjustment made to the calculation for student agreement to the time spent on out-of-class activities.

Students most commonly made the mistake on the “student presentation” activity; 86% of across all distributions of the PUMBA and 79% of cases where they student only made the mistake once. This occurred especially often on the surveys taken at the beginning of the semester; “student presentation” was the mistaken activity in 88% of the cases for students that made the mistake once. I am highlighting the students that made the mistake once to demonstrate that if I were to allow in the calculation for student agreement to the importance of in-class activities for one mistake, then I would be ignoring one activity for the vast majority of these students. Ignoring one activity is the practice I used when calculating agreement with the time spent on out-of-class activities for students that had no feeling as to how much time they are intended to spend on one activity. Using the exact same approach as that adjustment, I only calculate the student agreement score for the importance placed on in-class activities for students that did not make the mistake or only made the mistake one time. Students who made the mistake two or more times did not receive a score for this agreement measure because I did not want to calculate an agreement score on too few activities. The fewer activities that are included on the calculation for an overall agreement measure, the more impact a disagreement has on the calculation. Making this correction to the calculation, the agreement scores on this measure are similar between students who did not make the mistake and the students who made the mistake once. Without the correction to this measure, these two groups of students differed significantly ($p<.001$) as measured by a t-test. With the correction, there is no significant difference between these two groups of students ($p=0.085$)

I checked to see if students in three groups responded differently to other portions of the survey: 1) those who didn’t make the mistake; 2) those who made the mistake once; and 3) those who made the mistake on two or more activities. For example, as shown in Table 24, I explored how these groups of students scored on the four agreement measures (after all corrections to the calculations). In Table 25, you can see which student mistake groups (labeled as “1”, or “2+” mistakes) scored similarly as the students who made no mistakes (labeled as “0” mistakes). To make these comparisons, I transformed students’ overall agreement scores to within-class z-scores. Z-scores transform a set of numbers into a set with a mean of zero and standard deviation of one. In doing this, I could then compare the difference in agreement levels based on the student mistake group across all classes together because the z-scores eliminate any class level differences in agreement level. Thus, any differences in agreement level that remain once class level differences are accounted for must be due to making the time-importance mistake on the survey. I compared the “1” and “2+” groups to those that made no mistakes because I assume those students who made no mistakes were correctly understanding the survey, and thus I can use their responses without hesitation. (This assumption is supported by my validity interviews.) I find that for in-class importance the “0” and “1” groups of students responded similarly, due to the correction to that calculation (before the correction these students had statistically significant different agreement levels). Remember that the “2+” students were removed from this calculation so they could not have appeared as similar for this measure. For the out-of-class importance agreement measure, neither the “1” nor “2+” groups responded similarly to the “no mistake” group. Both groups of students that made the mistake received a statistically significant lower score (as measured by the post-hoc Dunn’s test to a Kruskal-Wallis test) than the students who didn’t make a mistake. I suspect that this is because these students are making the same mistake they made for the in-class activities, but due to the nature of the survey, I do not know which out-of-class activities they are mistaken on and so cannot correct for it. Somewhat alarmingly, not all groups of students responded similarly to the time agreement sections of the PUMBA. I expected these sections to be unaffected by the mistake because the causality of the mistake flows from saying no time is spent on an activity meaning you cannot say it is less important. However, it appears there are some differences in scores on the time sections based on the group of students. A plausible explanation could be that if a student disagrees with the class so much that they indicate an activity would have less than zero significance in their most effective class, then this level of disagreement could appear in the other measures. However, because I know that these students misunderstood the survey at one point, I cannot be certain that these differences between their agreement level and the non-mistaken students' agreement level is real or just a product of misunderstanding the survey. So that I do not make claims based on possibly invalid data, I take a conservative approach and eliminate the responses that are different from the "no mistake" group because I am uncertain if the difference is due to something real or a misunderstanding.

Table 24. Average z-score for the overall agreement measures by mistake group. Number of students in each group is in the parentheses.

Survey Distribution	Mistake Group	In-Class Importance	Out-of-Class Importance	In-Class Time	Out-of-Class Time
Beginning Fall 2017	0	0.003 (213)	0.207 (213)	0.165 (213)	0.077 (184)
	1	-0.006 (120)	-0.186 (120)	-0.228 (120)	-0.026 (100)
	2+	N/A	-0.336 (65)	-0.120 (65)	-0.197 (59)
End Fall 2017	0	0.076 (138)	0.352 (138)	0.080 (138)	0.243 (111)
	1	-0.106 (104)	-0.162 (104)	0.000 (104)	-0.082 (91)
	2+	N/A	-0.324 (98)	-0.113 (98)	-0.222 (88)
Beginning Spring 2018	0	0.032 (136)	0.166 (136)	0.086 (136)	0.016 (129)
	1	-0.059 (75)	-0.155 (75)	0.040 (75)	0.049 (66)
	2+	N/A	-0.218 (50)	-0.294 (50)	-.122 (43)
End Spring 2018	0	0.119 (93)	0.426 (93)	-0.039 (93)	0.339 (84)
	1	-0.178 (62)	-0.359 (62)	0.069 (62)	-0.094 (57)
	2+	N/A	-0.219 (79)	-0.009 (79)	-0.313 (74)

Table 25. Groups of students available for analyses using the four overall agreement measures.

Useable groups (0, 1, or 2+ mistakes)	**In-class Importance**	**Out-of-class Importance**	**In-class time**	**Out-of-class time**
Beginning Fall 2017	0, 1 (N=333)	0 (N=213)	0, 2+ (N=277)	0, 1 (N=284)
End Fall 2017	0, 1 (N=242)	0 (N=138)	0, 1, 2+ (N=340)	0 (N=111)
Beginning Spring 2018	0, 1 (N=211)	0 (N=136)	0, 1, 2+ (N=262)	0, 1, 2+ (N=239)
End Spring 2018	0, 1 (N=155)	0 (N=93)	0, 1, 2+ (N=234)	0, 1 (N=141)

I take this approach for each of the other measures in the PUMBA. I determined which groups of students responded similarly to those who did not make the mistake for each PUMBA-based measure that I used. Before running any analysis that involved a variable collected from the PUMBA (like agreement with the time spent on lecture), I first ran an ANOVA or a non-parametric Kruskal-Wallis test using PUMBA-based measure as the dependent variable and the groups of students as the categorical independent variable. If the data violated the assumptions for the ANOVA (homoscedasticity, for example) then I would use the Kruskal-Wallis test. A Tukey post-hoc test for the ANOVA or a Dunn post-hoc test for the Kruskal-Wallis told me which groups were similar and different from the "no mistake" group. Because I used z-scores for these comparisons, I have accounted for class level differences. If, while accounting for class level agreement differences, there is still a difference in agreement between the student groups, I know it is due to a misunderstanding in the survey and not due to the class. As these differences are due to a misunderstanding of the survey, the agreement levels of these groups (the groups found different from the students who did not make the mistake) are not representative of the students' actual agreement level. I remove these responses from analyses because I know that they are invalid measures of a student's actual agreement level. I then ran the main analysis using only the students who scored similar to the "no mistake" group.

5.5 Analysis Methods

5.5.1 Describing Student and Instructor Responses

I explored the student responses to the PUMBA and the instructor responses to the SIMBA to address research questions Q1 and Q3, which ask what students agree with and how instructors try to get agreement. For the PUMBA responses, I use descriptive statistics to describe the aspects of the class format (importance placed and time spent on in and out-of-class activities) that students do or do not agree with at the beginning and end of the semester. Also, I aggregate all students in each semester to explore trends across all classes. Next, I describe the specific activities that students agree and disagree with. For the time spent on in-class activities, I report how much more or less of an activity students would have in their HMEC if they did not agree with the time spent on that activity.

From this investigation, an interesting phenomenon appeared where students' perceptions of the time that would be spent on an in-class activity differed depending on whether or not they wanted less of that activity. Students were separated into three groups for each activity: 1) those who would prefer less time spent on that activity in their HMEC compared to their actual class; 2) those who agreed with the time spent on that activity; and 3) those who would prefer more time on that activity. For every student, I calculated the class-based z-score (the number of standard deviations their response is from their class's average) for their expected or perceived time spent on each activity in their actual class. Using the z-score allows me to analyze all students from every class at once, without having to worry if one class spends more or less time on an activity than the others. Had I not used z-scores, if there was a class with many students in one of the agreement groups, then the perceptions measure for that agreement group would be biased towards that one class. With the three groups of students as the categorical independent variable, I used either an ANOVA or a Kruskal-Wallis test, depending on the normality and homoscedasticity of the data, to explore for which in-class activities this phenomenon occurs. To provide a possible explanation for this phenomenon, I compared student expectations for the time spent on in-class activities with what the instructor stated would be the time spent on those activities. This comparison is a measure of students understanding, or misunderstanding, of the class format. I describe the results of this analysis on student understanding and how it relates to the differing perceptive phenomenon at the beginning and end of the semester.

Next, I explored the instructor responses to the SIMBA. I describe the activities that instructors commonly discuss with their students to introduce them to the class format and the strategies instructors use when having that discussion. I do this for instructor responses at the beginning and end of the semester. Also, by examining the data, I explore if instructors used certain methods or formats more or less commonly for certain activities. Due to the small sample size and lack of independence between the use of methods across activities, it would be inappropriate to use any inferential statistics (like a chi-square goodness of fit test) to see if a method or format was used more or less commonly for certain activities. Thus, I discuss salient trends in the data that suggest avenues for future research.

5.5.2 Hierarchical Models

5.5.2.1 Model Description

The remaining analyses heavily rely on the use of hierarchical linear models, or HLMs, to determine the effect of numerous continuous or dichotomous predictors on an outcome variable. HLMs are similar to ordinary multiple linear regressions but with the added complexity of allowing for intercepts and effects (slopes) to vary between groups of related subjects (i.e. students in a class). HLMs are used with nested, or hierarchical data, like mine. At most, my data falls into three levels, where level 1 is nested within level 2, which is nested within level 3. To explain a hierarchical linear model, I will refer to a three-level HLM. However, for many analyses, I only need two levels. The equations for the two-level model are very similar to the three-level model; removing the third level and third subscripts achieves the two-level model equations. The equation for the lowest level of an HLM appears, at first glance, like an ordinary linear regression (see Equation 6). Each parameter in Equation 6 has three subscripts. For the coefficients "β" and the predictor X, the first subscript denotes if it is an intercept (0) or a regression coefficient (non-zero). For the other parameters, the first subscript denotes the group of the level one variable (e.g. time of measurement). For all parameters, the second subscript corresponds to the level two group (student) and the third corresponds to the level three group (class). β_{0jk} represents the intercept for student j in class k, and β_{1jk} is the regression coefficient, or effect, for predictor X_1 for student j in class k. There may be any number of predictors and corresponding regression coefficients. e_{ijk} represents the error unexplained by the model. The next step is where the beauty and functionality of a hierarchical linear model comes in. β_{0jk} and β_{1jk} are allowed to vary by student, as defined by Equations 7 and 8. When a parameter is allowed to vary by group, it is referred to as "random" (e.g. a random intercept or a random effect). The intercept for student j in class k is equal to the overall intercept for class k, γ_{00k}, plus an error term for each student, U_{0jk}. Likewise, the regression coefficient for predictor X_1 for student j in class k is equal to the overall regression coefficient for class k, γ_{10k}, plus an error term for each student, U_{1jk}. Extending to a three-level HLM, the overall intercept and regression coefficient for class k is then defined by the overall intercept or regression coefficient for all of the data plus another error term (see Equations 9 and 10).

Level 1

$$Y_{tij} = \beta_{0jk} + \beta_{1jk}X_{1jk} + e_{ijk} \tag{6}$$

Level 2

$$\beta_{0jk} = \gamma_{00k} + U_{0jk} \tag{7}$$

$$\beta_{1jk} = \gamma_{10k} + U_{1jk} \tag{8}$$

Level 3

$$\gamma_{00k} = \delta_{000} + V_{0k} \tag{9}$$

$$\gamma_{10k} = \delta_{100} + V_{1k} \tag{10}$$

When analyzing a hierarchical linear model, statistical software will report the parameter estimate and significance tests for the fixed effects and the intercept. Also, unique to HLMs, I receive the estimates and significance tests for the variance of the intercepts and variance of any random effects (i.e. effects that can vary across groups). From the fixed effect estimates, I can determine the overall effect of a predictor. Based on that effect's variance, if I allow it to vary in the model, I can determine if the effect is different among the higher-level groups. Likewise, based on the variance of the intercept, I can determine if the groups are different without the effect of any predictors. For example, I may be able to say that the overall intercept for student agreement is 70, when all predictors are set to their 0 value. If the variance of the intercept is statistically significant, I know that this intercept is different between classes. As an example of interpreting an effect in an HLM, I may find that having experience in a studio class raises agreement across all students by five percentage points. This would be a fixed effect. If I allow the effect of having experience in a studio class to vary across classes, I could find a statistically significant variation in this effect. If so, I know that the magnitude of this effect is different from class to class.

Hierarchical linear models are applicable for analyses in which the outcome is continuous. However, there are some analyses I must run where the outcome is categorical. For instance, when investigating the effect of student agreement on predicted final grade ("A", "B", "C", "D or lower"). In such cases, I used a hierarchical logistic model or a hierarchical cumulative logistic model. The cumulative logistic model is similar to the logistic model but is used when the outcome has more than two categories. These models apply a logit or cumulative logit link to the data. The cumulative logit link determines the cumulative odds that a certain response would result in the outcome. For example, the result of a hierarchical logistic regression using a cumulative logit link applied to the final grade variable would report estimates for N-1 intercepts (where N is the number of outcome levels) that can then be used to determine the probability that a student expects to receive an A, the probability that a student expects to receive an A or B, and so on. The probability of the final intercept must be one since it is the cumulative probability of all possible outcomes, thus it is not directly reported. The calculation used to convert an estimate to a probability of an outcome is shown in Equation 11, where n is the intercept estimate.

$$P_{outcome} = \frac{e^{n}}{1+e^{n}} \tag{3}$$

If I wanted to know the probability that a student expects to receive *only* a B, I would take the difference between the probability that a student expects to receive an A or B, and that of a student who expects to receive an A, as shown in Equation 12.

$$P_B = P_{AB} - P_A = \frac{e^{n_{AB}}}{1+e^{n_{AB}}} - \frac{e^{n_A}}{1+e^{n_A}} \tag{4}$$

When I add predictors into the model, the estimate for the effect of the predictor is added to an intercept estimate and then used in Equation 11. For example, if the included predictor variable was studio experience ("exp"), where 0 = no studio experience and 1 = studio experience, then the effect of studio experience is estimated to be n_{exp}. The probability that a student with studio experience would expect to receive an A in the class would be calculated using Equation 13. However, the probability that a student with no studio experience would expect to receive an A would be calculated without adding the studio experience effect estimate because no studio experience was coded as 0.

$$P_{A,exp} = \frac{e^{n_A+n_{exp}}}{1+e^{n_A+n_{exp}}} \tag{5}$$

For the effect of a dichotomous variable such as studio experience, it is informative to discuss the odds ratio, which describes how the likelihood of an outcome changes between two groups. For example, I could say that the likelihood that a student with studio experience receives an A is 1.5 times the likelihood that a student without studio experience receives an A. The odds ratio (OR) is calculated with the probability of the outcome between the two groups p_1 and p_2 (see Equation 14). p_1 could be calculated from Equation 11 and p_2 could be calculated from Equation 13.

$$OR = \frac{p_2(1-p_1)}{p_1(1-p_2)} \tag{6}$$

Although this discussion has been based on the hierarchical cumulative logistic model, the hierarchical logistic model, used for dichotomous outcomes, is analyzed and interpreted similarly. The only difference in the interpretation of the results between the two logistic models is that for the non-cumulative logistic model, Equation 12 is not applicable.

5.5.2.2 Model Specification

When building a model to explain a set of data, the goal is to find a model that best describes or predicts the data without being too complex (Snijders & Bosker, 2012). In the vocabulary of statisticians, I want to find the most parsimonious model, one that explains the most variance using the fewest predictors. I will be following the model building procedure for two-level hierarchical data outlined by Snijders and Bosker, who recommend using a two-phase approach. In essence, this approach adds predictors to the model one level at a time, so it is easily extended to a three-level model when necessary. This procedure is also recommended by Raudenbush and Bryk (2002), who provide the theoretical background for much of the functionality of the "HLM" software. The two-phase procedure is the method with which the "HLM" software analyzes a hierarchical linear model (Raudenbush & Bryk, 2002; Snijders & Bosker, 2012).

The first step in creating a hierarchical linear model is to create a "null model", which allows for random intercepts (intercepts that vary by group) but contains no predictors. From the null model, I calculate the intraclass correlation coefficient, or "ICC". The ICC is the proportion of the total variance of the data that is explained by differences between groups (see Equation 15). It is defined as the variance of the random intercepts, τ_{00}, divided by the total variance of the data or the sum of τ_{00} and the remaining unexplained variance, σ^2.

$$ICC = \frac{\tau_{00}}{\tau_{00}+\sigma^2} \tag{7}$$

The ICC indicates the maximum amount of variance that exists at the group level and that could be explained by group level variables. Thus, an ICC near 0% (meaning group level variables explain virtually none of the total variance in the outcome variable) would suggest that an ordinary linear regression may be appropriate. However, using an ordinary linear regression assumes independence among the data. Considering that students are not independent, they are grouped by class or instructor, I would be violating an important assumption if I used an ordinary linear regression. Additionally, hierarchical models yield more precise model estimates because there may be group level variation that hierarchical models can explain that ordinary linear regressions cannot. Thus, with more precise model estimates, I have more statistical power; I am less likely to make a type 2 error (finding no significant result when there actually is one in reality). For these reasons, I used hierarchical models even if the ICC for a model was near 0%.

After analyzing the null model and finding the ICC, I create a new model that incorporates level one predictors. This is phase one of the two-phase model specification procedure (Snijders & Bosker, 2012). For my data, the first level in a two-level HLM will be student-level data. Snijders and Bosker recommend selecting level one predictors that are the focus of the research and theoretically relevant (i.e. included on the basis of subject-matter knowledge). If there are any interactions between these predictors, that should be included in the model as well. Also, if I expect any of these predictors to have a group dependent effect, then those should also be included as a random effect. Next, I run the analysis on that model (the phase 1 model) and determine which fixed and random effects are significant. Any non-significant effects should be excluded from the next model (the phase 2 model). At this point, I may check to see if any unexpected predictors might have a group-dependent effect by creating such a model and checking for significant random effects. However, Snijders and Bosker warn against including too many random effects in the model as it may lead to a long iteration process from the statistical software or the model may fail to converge, thus resulting in no estimates of any parameters. These difficulties arise due to less data available to create the estimates for random effects.

Once a phase 1 model is created containing only significant level one fixed and random effects, I may move on to phase two of the model specification, where I enter the level two predictors into the model. The rules for selecting level two predictors are the same as selecting level one predictors: the predictors should be the focus of the research or theoretically relevant. Now, I run the phase 2 model with the previously significant level one predictors, and the selected level two predictors to find which effects are significant. Keeping only those final fixed and random effects that are significant, I run the analysis one final time, creating the final model, to obtain the parameter estimates and significance statistics. To determine the effect size of this final model, I calculate the proportional reduction in unexplained variance between the final model and the null model (see Equation 16). Calculating effect size using Equation 16 yields an estimate analogous to the R^2 value found using ordinary linear regression, where an R^2 of 0.04 is the "recommended minimum effect size representing a 'practically' significant effect for social science data" (Ferguson, 2009, p. 533) , 0.25 represents a moderate effect, and 0.64 represents a strong effect. For two level hierarchical models, there are two sources of unexplained variance: the unexplained group level variance (τ_{00}) and the unexplained individual level variance (σ^2), the same parameters used to estimate the intraclass correlation coefficient.

$$R^2 = 1 - \frac{\tau_{00(final)} + \sigma^2_{final}}{\tau_{00(null)} + \sigma^2_{null}} \quad (16)$$

Effect sizes for a hierarchical logistic model are calculated differently than for a hierarchical linear model. With logistic models, the outcome variable is assumed to follow a logistic distribution with a fixed variance of 3.29 (Snijders & Bosker, 2012). Thus, σ^2 is equal to 3.29. This alters both the calculation for the ICC (Equation 15) and the calculation for the model effect size. The effect size for a hierarchical cumulative logistic model or hierarchical logistic model is defined as the proportion of variance explained to the total variance. The variance explained, σ_F^2, is the variance of the predicted outcome values based on the model. The unexplained variance is the group level variance, τ_{00}, plus the fixed level one variance of 3.29. Thus, the effect size, which is analogous to an R^2 value, is as defined in Equation 17. It is important to note that effect size estimates calculated in this way are known to be lower than typical R^2 values obtained for linear regressions (Snijders & Bosker, 2012).

$$R^2 = \frac{\sigma_F^2}{\sigma_F^2 + \tau_{00} + 3.29} \qquad (17)$$

5.5.3 Change in Student Agreement

Using the student responses to the PUMBA from the beginning and end of the semester, I determined how student agreement changed over the course of the semester, addressing research question Q1e. At the beginning of the semester, the PUMBA measures student agreement with their expectations of the class format. By the end of the semester, the PUMBA measures student agreement with the class format based on their experience. I determined the change in student agreement over the semester by comparing this measure at the beginning and end of the semester. A positive change indicates that students began to agree more with the class format as the semester went on. A negative change indicates that students agree less with the class format at the end of the semester than they did at the start. Student agreement is measured on four overall scales:

agreement with 1) in-class activity importance; 2) out-of-class activity importance; 3) in-class activity time; and 4) out-of-class activity time. These measures are all continuous.

I used two methods to investigate the change in student agreement from the beginning to the end of the semester. For the change in student agreement within one class, I used a paired t-test with students who took the survey at the start of the semester as one group and the same students who took the survey at the end of the semester as the other group. If the usual assumptions of normality of the data for each group or equality of variances (homoscedasticity) between the two groups were violated, I used the non-parametric equivalent of a paired t-test, the Wilcoxon ranked-sum test. These tests tell me whether there is a statistically significant change in student agreement but do not provide information about the practical significance of the change (Thompson, 1988). An effect size estimate provides information about the magnitude of the relationship between variables (Ferguson, 2009) and is recommended to be reported in addition to statistical significance testing (Wilkinson, 1999). As these are comparisons of a continuous outcome variable between two groups, Cohen's d (Cohen, 1988) is the appropriate measure of effect size for this analysis. Using Cohen's d, an effect size between 0.2-0.5 is considered small, 0.5-0.8 is medium, and 0.8 or higher is a large effect. Cohen (1992) described a medium effect as something likely observable to the naked eye (i.e. a practical significance). Small effects were set to be less noticeable than a medium effect but non-trivial. Cohen set large effects to be the same distance from a medium effect as small is from medium.

Given the hierarchical nature of my data, a hierarchical linear growth model is an appropriate model of the change in student agreement over time for all students together. The t-test method would not be appropriate to look at all students together because it cannot account for the grouping of students based on class like a hierarchical model can. I created four, three-level HLMs, one for each overall student agreement measure. Each model has a structure like that shown in Equation 6, with parameters defined in Equations 7-10. For the growth model, X_1 is the time variable (when the survey was completed: beginning of the semester or end of the semester) and no other predictors are included. Effect sizes for these models are calculated by finding the proportional reduction in the sum of variances at each level of the final model compared to the null model (Equation 16).

I used two different types of models to investigate the change in student agreement to specific activities over time. I explored change in student agreement with the time spent on in-class activities using a hierarchical linear growth model like the one just described. However, instead of representing an overall agreement score for each student, Y_{ijk} represented the absolute difference in proportion of class time a student would spend on an activity in their HMEC compared to their actual class. A smaller value for this absolute difference at the end of the semester indicates an increase in agreement. I also measured the change in student agreement to the other aspects of the class format (e.g. importance) using a three-level hierarchical logistic regression model. A hierarchical logistic regression model is necessary because the outcome is dichotomous, either students agreed with the importance or time spent on an activity or they did not. The parameter estimates resulting from this model are used to calculate the probability that a student would agree with the activity in question and if that probability changes from the beginning to end of the semester. Also, from this model I can determine if this effect varies from class to class.

As I was looking at change in agreement from the beginning to the end of the semester, it was important that I only used the matched students, or those that took the survey at the beginning and end of the semester. Depending on the outcome variable used, there were a different number of students used in each analysis, due to the different responses of those students who made the time-importance mistake.

5.5.4 The Effect of Student Agreement on Student Outcomes

I hypothesized that students who buy in to the class format will perform better in the class because students who buy-in will: 1) make it easier for instructors to implement research-based instructional strategies: and 2) be more receptive of such strategies. In this analysis, I searched for a connection between student agreement with aspects of the class format and their performance in the class, addressing research question Q2. Student agreement was measured on the student level as their overall agreement to the four aspects of the class format (importance placed on and time spent on in-class and out-of-class activities). Student performance was quantified at the end of the semester in three ways: 1) their score on conceptual inventories like the FCI (Hestenes et al., 1992) or CSEM (Maloney et al., 2001); 2) their favorable attitudes as measured by the CLASS (Adams et al., 2006); and 3) their expected final grade. Each analysis included other controlling factors, such as student or instructor gender and experience in studio classes, as well as pre-scores on the FCI, CSEM, or CLASS.

Most of these analyses followed a two-level hierarchical linear model (HLM) with students nested within class. An HLM is appropriate because the outcome variables are continuous and the predictors are either continuous (e.g. student agreement level) or dichotomous (e.g. gender). When looking at the effect of student agreement on their expected final grade, the HLM applied a cumulative logit link to the categorical outcome variable. This is necessary when analyzing categorical outcomes with more than two levels (Ene, Leighton, Blue, & Bell, 2015) such as expected final grade where students expect to receive either an A, B, C, or D or lower.

I also wanted to determine if any specific activities have a significant effect on a student outcome. So, for all of these models, if I found any of the four overall agreement measures as a significant predictor of a student outcome, I further explored this effect to see if any specific activities were the cause. In a different model, I replaced the significant overall agreement measure with the agreement measures for the activities that comprised it. The agreement scores to the individual activities acted as a new set of predictors to determine if there were any specific activities that had an effect on student outcomes. Running the analysis on this new model allowed me to determine if a specific activity significantly effects a student outcome. I report both overall agreement measures and specific activities that had a significant effect on a student outcome.

5.5.5 The Effect of Instructor Discussion of Class Format on Student Agreement

So far, I have explored which aspects of the class students agree or disagree with, how that changes throughout the semester, how student agreement affects class performance, and the ways in which instructors try to gain student buy-in. The final piece of my investigation into student buy-in explores the effectiveness of the instructor's discussion on the class format in terms of gaining student buy-in. With this analysis, I attempt to answer research question Q4 which asks, "What effect do instructors' discussions of the class format have on student buy-in?" Answering this question should provide me with the strategies that influence student buy-in which I may then disseminate to other instructors. I may be able to tell instructors what students agree or disagree with and how that affects their performance in class. The natural question I should expect from every instructor is, "How can I get my students to agree with the class format?" Although there are materials out there to help instructors do this ("Framing the Interactive Engagement Classroom, Boulder: Science Education Initiative at the University of Colorado, Boulder," 2013), there are no prior research-proven methods.

Like the majority of the analyses used in my research, I investigated the effect that instructor's discussions of the class format had on student agreement with the time spent on in-class activities by using a hierarchical linear model with students nested within class. An HLM works for this data because the outcome variables are continuous and the predictors are either continuous (e.g. age) or dichotomous (e.g. the "comparison between reformed and traditional classes" method was or was not used). For each HLM, the outcome variable was some form of student agreement with the time spent on a single in-class activity. I used a different HLM for each activity because instructors used different methods to discuss the class format regarding different activities. The results of these analyses tell me which specific methods or formats that instructors use in regard to a specific activity significantly affected student agreement with that activity. At first, I ran an analysis on the difference between the proportion of class time a student would spend on an activity in their HMEC and the proportion of their actual class they expected or experienced was spent on that activity. That is to say, the outcome variable was the value defined by HMEC-Perceived_Actual for a specific activity. Note that this is not the absolute value of the difference like that used in the calculation for student agreement.

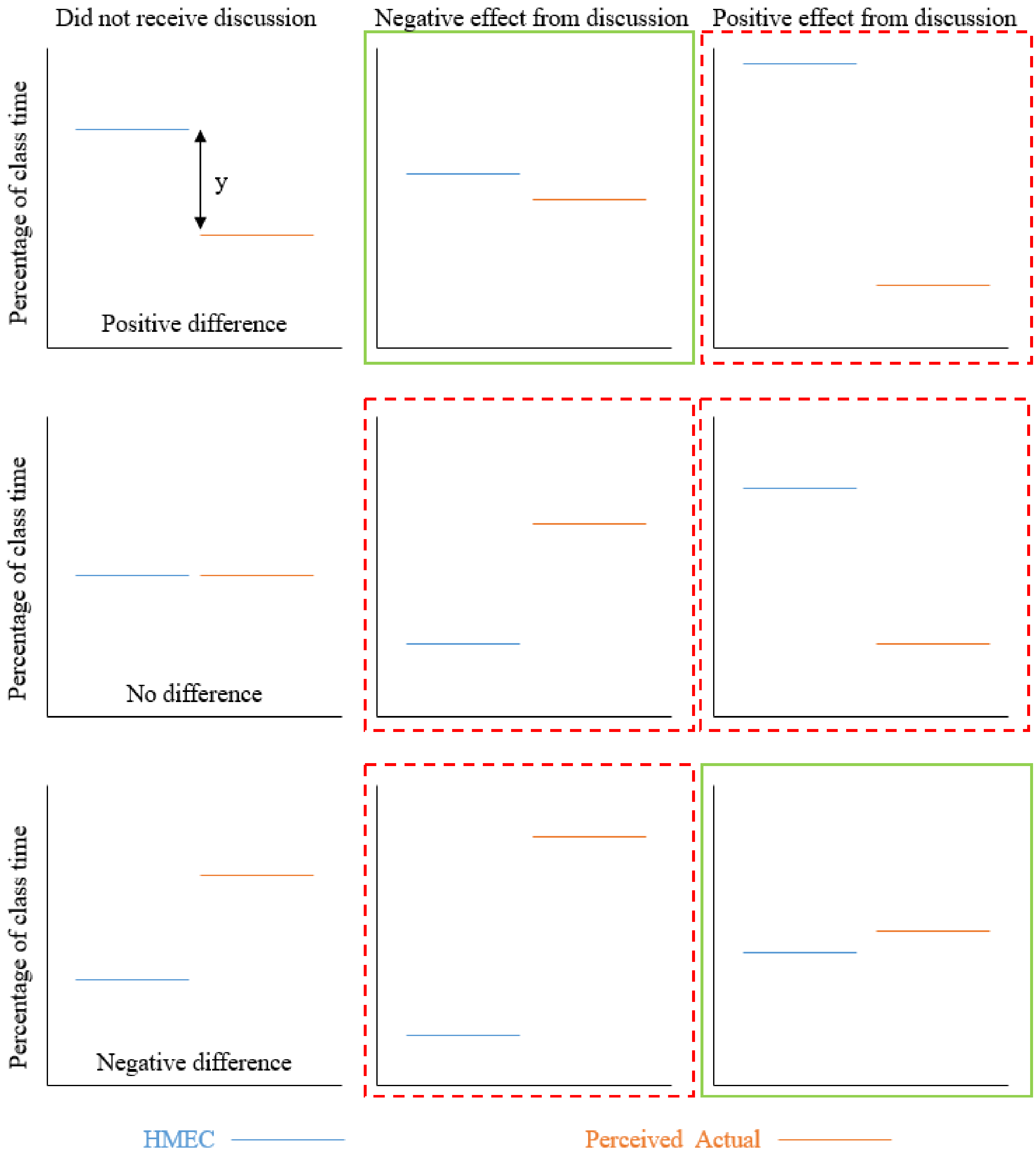


Figure 9. An illustration of the possible circumstances yielding a negative or positive effect using the difference between HMEC and Perceived_Actual as the outcome variable, “y”.

I explored the effects on this difference measure because I wanted to see if the instructors' discussions were affecting this difference in a particular direction to all students. For example, if the effect of a discussion was a negative, that would indicate that the difference between HMEC and Perceived_Actual is either a smaller positive value or a larger negative value for those that experienced the discussion compared to those that did not (see the center column of Figure 9). This may happen if either the HMEC proportion of class time spent on an activity decreased due to the discussion or if the perception of the proportion of class time expected or experienced in the actual class increased due to the discussion or both. For example, imagine a student whose difference measure is +10 (indicating they would prefer more of an activity) and another student whose difference measure is -10 (indicating they would prefer less of an activity). If a certain strategy is found to have a negative effect of 5 on the difference, then these students' agreement levels would decrease to +5 and -15. The effect acts in the same direction regardless of if a student prefers more or less of an activity. Figure 9 illustrates this example and the other possible results of this analysis if an effect were found. Note that in Figure 9, the difference "y" is the outcome variable in the analysis and is defined as HMEC-Perceived_Actual. The intercept value of this analysis determines which row to look in; a positive intercept (meaning the student prefers more) corresponds to the top row, a zero intercept to the middle row (meaning the student agrees), and a negative intercept (meaning the student prefers less) to the bottom row. The sign of the effect determines which column to look in. Results in Figure 4 that have a solid green outline are those that indicate an increase in student agreement with the time spent on that activity. Results with a dashed red outline are those that indicate a decrease in student agreement with the time spent on that activity. The examples in Figure 9 assume that no effect is so large that it actually flips the gap between the HMEC and the Perceived_Actual, though it is certainly possible.

However, an instructor strategy could have a different effect on different students depending on if they had preferred more or less of an activity. It may be that students that would have a positive intercept (preferring more of an activity) experienced a negative shift, and students with a negative intercept (preferring less of an activity) experienced a positive shift. In other words, the discussion of the class format may have had the effect of bringing a large variation of agreement levels to a smaller variation. For example, if we consider the same two students as before, one whose difference was initially +10 and another whose difference was -10, this effect could lead to these differences going to +5 and -5. In this example, the strategy had the effect of bringing students closer to agreement no matter where they started. I ran a second analysis to check if discussions of the class format reduce the variation in student agreement scores instead of shifting the scores one way or the other. To do this, I used the student agreement score for each activity, which is the absolute value of the difference between the HMEC and Perceived_Actual proportions. With the analysis based on the absolute value of this difference, a negative effect would always indicate an increase in student agreement compared to those students that did not experience the discussion. A negative effect means that the magnitude of the difference decreased, indicating that the amount of time a student preferred to spend on an activity is closer to the amount of time they expected or experienced in their actual class.

I also analyzed the effect of instructor strategies to gain student agreement with the importance of class activities and the time spent on out-of-class activities. At the student level, these measures are dichotomous, either the student agreed or they did not. Since the outcome is dichotomous and the predictors are either continuous or dichotomous, a hierarchical logistic model is appropriate. The hierarchical logistic model calculates the probability that a student will agree with the activity given the predictors.

5.6 Summary

In this chapter, I described the procedures I used for collecting and analyzing data; see Table 26 for a summary matching my data and analysis methods to the research questions. First, I discussed the timeline of my study, and described those who participated in the research. Next, I discussed the interviews I completed with students and instructors in which we mainly talked about the influences on student agreement and what strategies could be the most effective at changing student agreement. In the interview section of this chapter, I described how I analyzed the interviews by using a coding scheme and demonstrated reliable application of the codes by by achieving inter-rater reliability with a colleague. After discussing the interviews, I summarized the data that I collected from the SIMBA and PUMBA. I had discovered that students were making the "time-importance mistake" on the survey and so I discuss how I analyze my data knowing that some students may have answered the PUMBA with invalid responses. Finally, I describe the various methods that I used to analyze the data from the surveys. These methods relied heavily on hierarchical models.

Table 26. An outline of the analysis methods used to answer each research question.

Research Question	Data	Analysis Method
Q1a: How well do students understand how time will be spent in class?	Student and instructor perceptions of actual class time spent in class	Descriptive statistics of the magnitude difference between instructor and student responses.
Q1b: How well do students understand how frequently specific activities will occur in class?	Student and instructor perceptions of actual class time spent in class	Descriptive statistics of the difference between instructor and student responses for specific activities.
Q1c: How well do students agree with the class format?	Student overall agreement	Descriptive statistics
Q1d: What specific class activities do students agree or disagree with?	Student agreement to individual activities	Descriptive statistics
Q1e: Does student understanding and agreement change between the beginning and end of the semester?	Student agreement and student vs instructor perceptions of actual class time spent	Longitudinal hierarchical linear model
Q2a: Does students' agreement with the class format predict their final grade?	Student agreement and expected final grade	Hierarchical cumulative logistic model
Q2b: Does students' agreement with the class format predict their performance on concept tests?	Student agreement and concept test scores	Hierarchical linear model
Q2c: Does students' agreement with the class format predict their performance eon attitudinal surveys?	Student agreement and CLASS scores	Hierarchical linear model
Q3a: What class activities do instructors commonly talk about with their students to set their expectations and get their agreement with those expectations?	Instructor indicated discussed activities	Descriptive statistics
Q3b: What methods or topics do instructors commonly use to get student agreement?	Instructor indicated formats used	Descriptive statistics
Q3c: Through which formats do instructors commonly have these discussions?	Instructor indicated methods used	Descriptive statistics

Research Question	Data	Analysis Method
Q3d: Do these things change from the beginning to the end of the semester?	Instructor indicated discussed activities and strategies used	Descriptive statistics
Q4a: Is the use of certain methods or formats to get student agreement associated with higher levels of student agreement?	Instructor strategies used and student agreement	Hierarchical models and interview responses

CHAPTER SIX: QUANTITATIVE RESULTS

In this chapter, I report the results of the quantitative analyses described in Chapter Five in order of the research question they address. First, I describe the students' responses to their survey. I investigate how well they understand how the class time will be spent in their actual class and which activities they are properly or improperly understanding. Then, I explore how well students agree with different aspects and activities of the class. I reveal an interesting relationship between student perceptions of time spent on class activities and their desired amount of that activity. Second, I investigate how student agreement predicts student outcomes as measured by performance on concept tests, attitudinal surveys, and their expected final grade. Next, I describe the strategies instructors use to gain buy-in by exploring the activities they commonly talk about and the methods and formats they use when discussing those activities. Finally, I discuss how these instructor strategies might affect student agreement with individual activities.

6.1 Student Buy-in

6.1.1 Differing Perspectives Phenomenon: Students Who Want Less of an Activity Perceive More of It

Before beginning my investigation into the aspects of a studio class that students do or do not buy in to, I must first address the phenomenon of differing perspectives, which is when a student perceives the time spent on an activity in class differently than other students based on their agreement with that activity. While reporting survey results to instructors, I realized that students who would prefer less time spent on an in-class activity in their HMEC compared to their actual class perceived more of that activity in their actual class than those students who agreed with the time spent on the activity or those who would prefer more time spent on the activity. Figure 10 illustrates this "differing perspectives phenomenon". One might expect that after being oriented to the class format, all students in one class would have, on average, similar perspectives of how much time would be spent on each in-class activity. Similarly, one would expect similar perspectives by the end of the semester, when students have experienced the class format. With this assumption, students would differ primarily in the amount of time they would spend on an activity in their HMEC, not in how they perceive the class. However, for one group of students I my data collection, this was not the case.

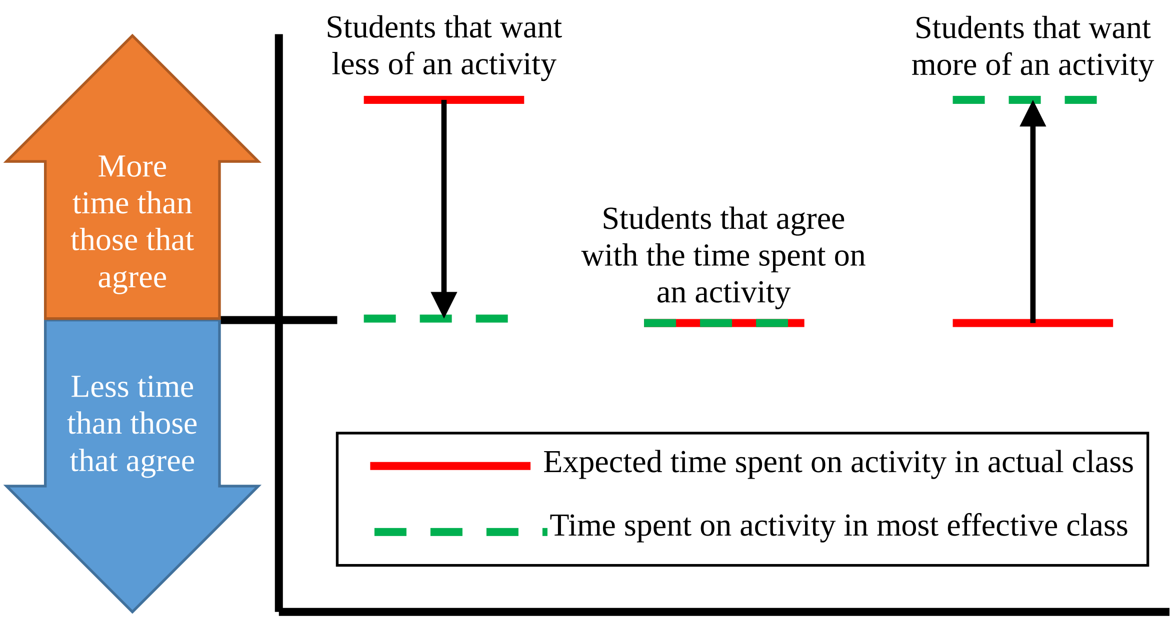


Figure 10. An illustrative model of the “differing perspectives phenomenon”.

For each in-class activity (Table 9), I separated students into three groups based on if they preferred less time, more time, or agreed with the time spent on that activity. Students could be in a different group for each activity. I used an ANOVA followed by a Tukey post-hoc test or a Kruskal-Wallis test followed by a post-hoc Dunn’s test on class-based z-scores for the expected time spent on an activity. If the ANOVA or Kruskal-Wallis test indicated that there were differences between the groups, I used the post-hoc test to determine which groups of students differed. For every distribution of the survey and for every activity, the students who would prefer less time spent on an activity in their HMEC compared to their actual class perceived, to a statistically significant degree, that the activity would or did occur more often in their actual class than at least one of the other two groups of students. The large majority of the post-hoc pairwise comparisons were significant to the $p < .001$ level. In most instances (33 out of 36 possible comparisons), students who wanted less of an activity perceived more of that activity compared to *both* of the other groups of students. The exceptions are described in Table 27.

Table 27. Differences or exceptions to the trend that students who prefer less of an activity perceive more of that activity than students who agree and prefer more.

Survey Administration	**Activity**	**Groups Who Differed**
Beginning of Fall Semester	Student Presentation	Students Who Preferred Less > Students Who Agreed
End of Fall Semester	Quiz	Students Who Agreed > Students Who Preferred More
End of Fall Semester	Step-by-Step Labs	Students Who Agreed > Students Who Preferred More
End of Fall Semester	Demonstrations	Students Who Agreed > Students Who Preferred More
Beginning of Spring Semester	Lecture	Students Who Agreed > Students Who Preferred More
End of Spring Semester	Lecture	Students Who Preferred Less > Students Who Preferred More Students Who Agreed > Students Who Preferred More
End of Spring Semester	Quiz	Students Who Agreed > Students Who Preferred More
End of Spring Semester	Step-by-Step Labs	Students Who Agreed > Students Who Preferred More
End of Spring Semester	Demonstrations	Student Who Preferred Less > Student Who Preferred More

There are two ways to interpret the meaning of this phenomenon which differ in the direction of causality between perception of the class and agreement with the class. With the “agreement first” interpretation, student perceptions of how the class works are biased by their agreement (in other words, agreement drives perception). For example, after the class orientation, a student may have decided that she does not agree with how much investigative lab work there will be. When asked how much time would be spent on investigative lab work in the PUMBA, she might put down what she would prefer for her HMEC and then knowing that she wants less than her actual class, put a larger number for her actual class. So, her perceptions are based on her agreement. This scenario could arise if the instructor was not clear as to how much time would the class would spend on investigative labs or if the student missed that part of the orientation. Possible evidence against this interpretation is the fact that it only occurs for students that want less of an activity and not for students that want more of an activity. However, this may be due to the response limitations for this section of the PUMBA, which I discuss in the “perception first” interpretation.

Research by Appleton-Knapp and Krentler (2006) has found that satisfaction with a course effects how students respond regarding course expectations (i.e. the “agreement first” interpretation). Specifically, when asked at the end of the semester, students satisfied with their class incorrectly recalled their expectations from the beginning of the semester. Students who were satisfied with their class would recall their initial expectations as consistent with what they experienced in the class, even though their actual initial expectations were not similar to their experiences. For example, if a student initially expected an activity to happen infrequently, experienced it to happen frequently and was satisfied with it, they would then recall their initial expectations as saying that the activity would happen frequently. Appleton-Knapp and Krentler identified this as a hindsight bias. However, I find the differing perspectives phenomenon at the beginning of the semester when students are not recalling information but rather indicating their current views. It is possible that hindsight bias may explain some of the differing perspectives phenomenon found at the end of a semester.

Using the “perception first” interpretation, students, for one reason or another, perceive that a class activity will or did occur often. If we assume that these students would prefer a class format similar to the majority of the other students in the class, then they would prefer less time on that activity because they perceived it as occurring more often. In this case, the perceptions cause the disagreement. A scenario such as this might arise if students are understanding the class format differently from the rest of the class. Perhaps, these students who ultimately prefer less of an activity came into the class with prior expectations that were not influenced by the instructor’s class orientation.

The StRIP survey (Nguyen et al., 2017) found evidence of perspectives of the class format affecting student response to the class (i.e., the “perception first” interpretation). Student expectations for the amount of time spent on passive lecture was positively related to how well the students valued the class. If students expected more passive lecture, then students valued the class greater. There was a negative relationship found between expectations of active learning lecture and how students evaluated the course. These results indicate that student expectations or perceptions of passive or active lecture predicted how they responded to (valued or evaluated) the class.

We might expect that students’ perspectives of the class time spent on an in-class activity would model a normal distribution centered around the true time spent on the activity (at the end of the semester) or centered around the true amount of time the instructor said would be spent on that activity (at the beginning of the semester). With this lens, there will be a set of students who perceived more than the class average. At some point, the amount of more time on an activity that these students perceive will surpass the amount of time they would like to spend on that activity, thus being categorized into the group of students that would prefer less time on an activity. These students might have perceived more time spent on the activity due to natural variation in perceptions. However, if I find some other commonality among the students who perceived more time than the rest of the class, then it may be that the commonality is the reason they perceived more time. For an example in a different context, people may run a mile in different times due to natural variation. However, if I find that the people who run faster miles also share the commonality of working out at a gym every day, I may consider that working out at a gym is connected to running faster miles. Likewise, for my data, student may perceive more time spent on an activity due to natural variation, but if I find that these students also all preferred less of that activity, I may consider that as connected to their increased perceptions. Indeed, that is what I find in my data; students who prefer less of an activity perceive that activity more often. One issue with this normal distribution model as it applies to my data is that it suggests I should also find students who perceived less time spent on an activity than the class average, and for 29 of the 36 comparisons, that was not the case. Although it might explain why there are a group of students who perceive more of an activity (who just so happen to also prefer less of the activity), it does not explain why there is not a group of students who perceive less of an activity.

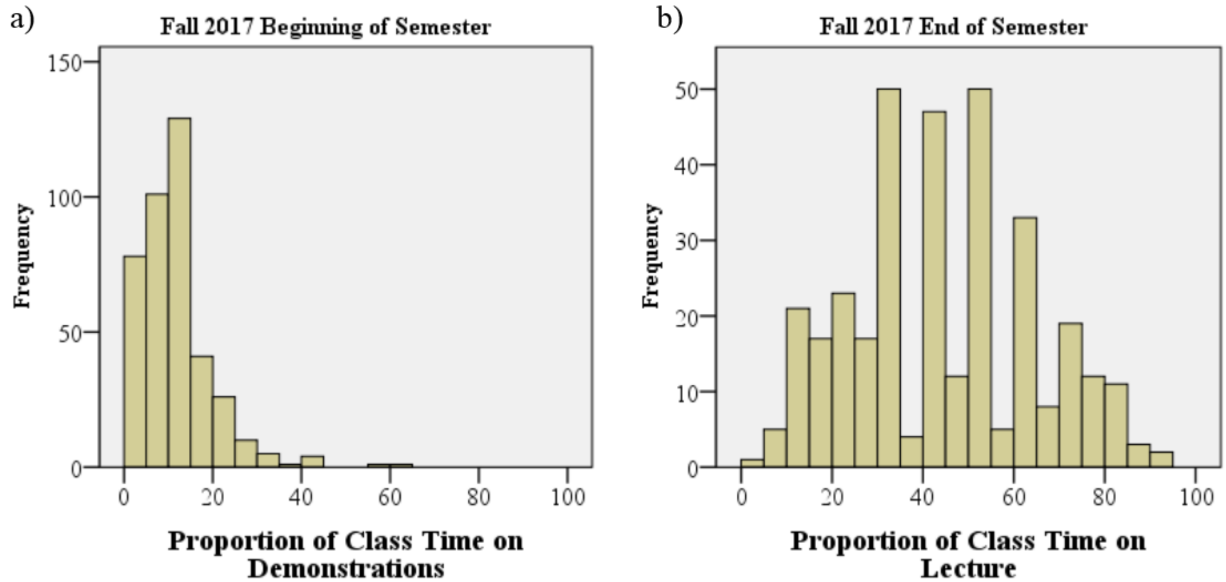


Figure 11. Histograms of student responses to expected proportion of class time on demonstrations (a) and perceived time spent on lecture (b).

Considering the limitations to the student responses (i.e. they can only enter a value between 0% and 100%), approximating student responses with a normal distribution may be invalid. Many activities are expected or perceived to occur infrequently, close to the 0% limitation. Thus, with these activities, there is less room to perceive the activity occurring less often and more room to perceive it as occurring more often. So, a histogram of student responses would look less like a normal distribution and more like a distribution skewed towards lower values (see Figure 11a) meaning only students who perceive more time spent on activity can differ from other students to a statistically significant degree. This is the case for many activities. While this approach could explain the phenomenon for many activities, it does not explain why there are still different perspectives for activities that do occur more often, thus having a normal distribution (see Figure 11b). Perceptions of lecture at the end of the Fall 2017 semester approximates a normal distribution with similar amounts of students perceiving more or less lecture than the average. Yet, I still find that student who prefer less lecture perceived it as occurring more often while students who preferred more lecture did not perceive it as occurring less often. This means that preferring more lecture was not a commonality of the students that perceived less lecture.

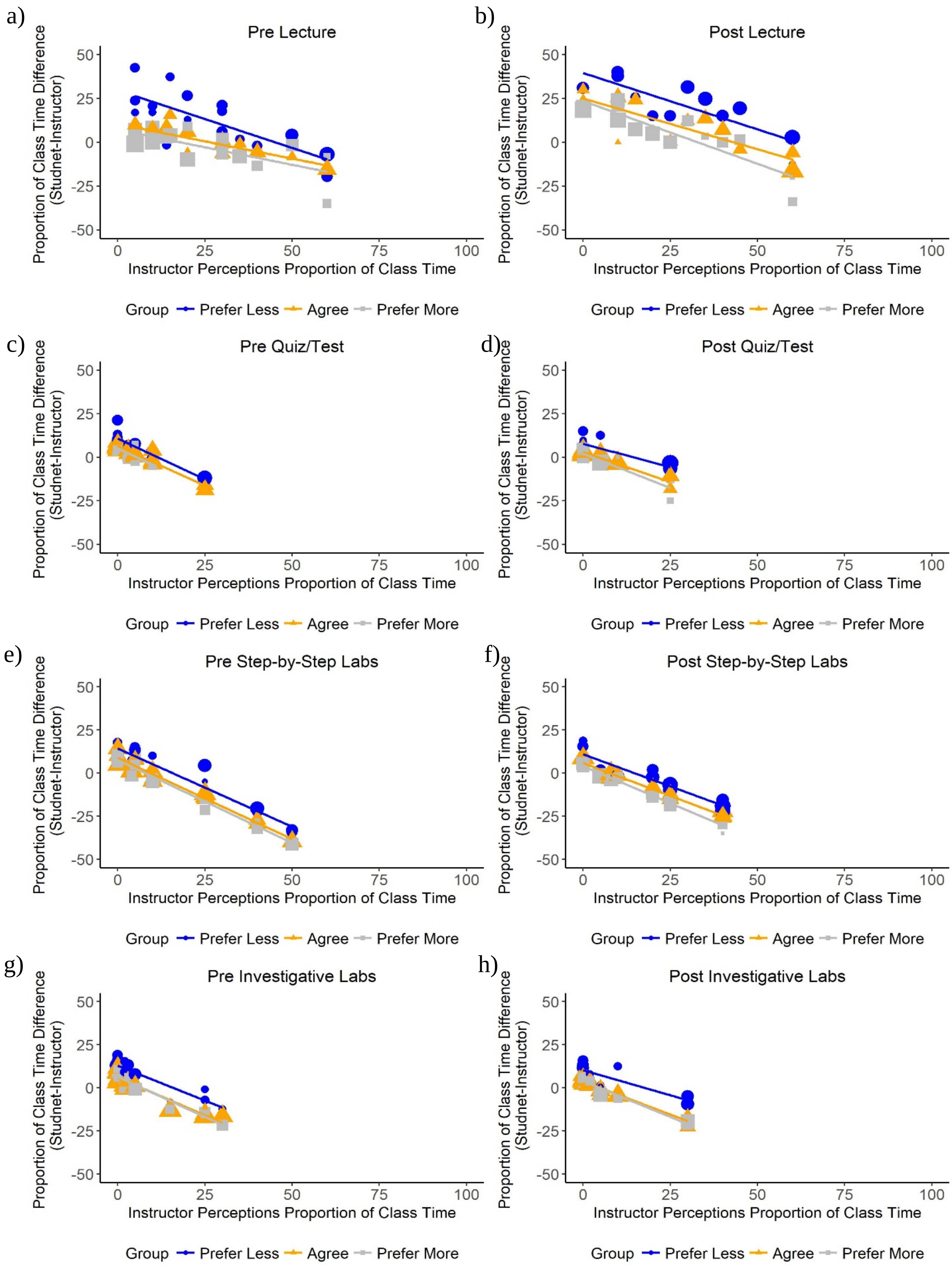
The important take-away from the differing perspectives phenomenon is that instructors should be aware that their students perceive the class time in different ways, particularly based on if they agree with it. Instructors should know that if a student prefers less of an activity, then the student may perceive that activity to occur more frequently than other students who agree with or want more time spent on the activity. This knowledge should help instructors change student agreement as they will have a more accurate idea of how the student perceives the class.

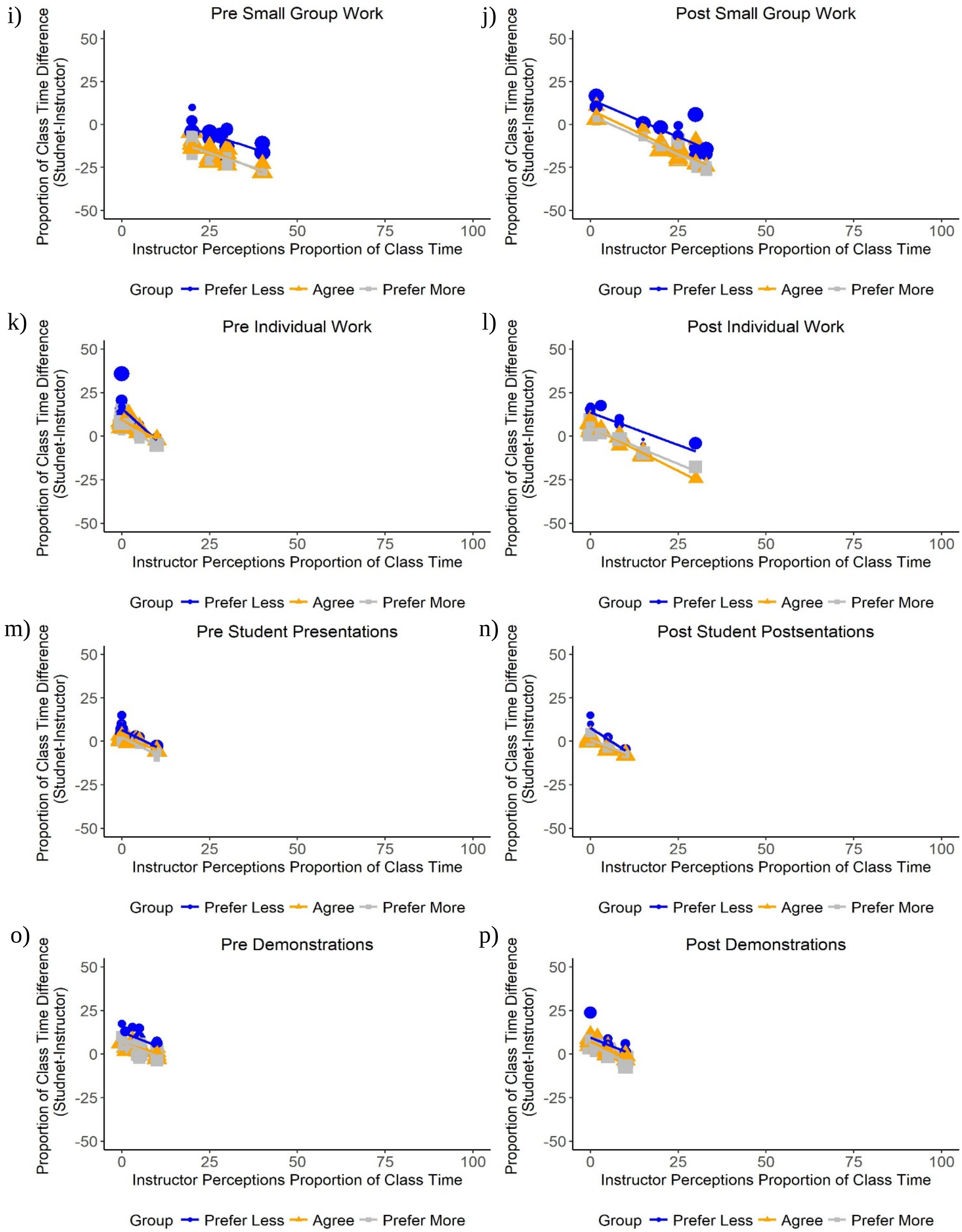
To provide a possible explanation of the differing perspectives phenomenon through the "perception first" approach, I explore student understanding of how the class time is spent across the three agreement groups. It may simply be that students are misunderstanding how the class actually functions. Fortunately, with the PUMBA, I can determine if students are understanding the class format, based on the instructor's definition of the class format.

6.1.2 Student Understanding

Whether or not one takes the "agreement first" or "perception first" interpretation of the differing perspectives phenomenon, the solution is for the instructor to be explicitly clear about the time spent on class. That way, there should be no opportunity for the students to perceive the class differently than the instructor expects them to. Using student responses from the PUMBA and instructor responses on the SIMBA, I can measure how well students understand the class format as defined by the instructor. Remember, though, that student and instructor perceptions of how class time is allocated to in-class activities is a somewhat subjective measure as they are different people with different ways of perceiving things (e.g. 50% to the instructor may be 40%

to the student). When examining groups of students, however, the differences in the way they perceive the class should cancel out to an average perceived proportion of class time spent on an activity that can be used in comparison to the instructor's perceptions.


a)
Pre Lecture
b)
Post Lecture
c)
Pre Quiz/Test
d)
Post Quiz/Test
e)
Pre Step-by-Step Labs
f)
Post Step-by-Step Labs
g)
Pre Investigative Labs
h)
Post Investigative Labs
Proportion of Class Time Difference (Studnet-Instructor)
Instructor Perceptions Proportion of Class Time
Group
Prefer Less
Agree
Prefer More
50
25
0
-25
-50
0
25
50
75
100


i)
Pre Small Group Work
j)
Post Small Group Work
k)
Pre Individual Work
l)
Post Individual Work
m)
Pre Student Presentations
n)
Post Student Postsentations
o)
Pre Demonstrations
p)
Post Demonstrations
Proportion of Class Time Difference (Studnet-Instructor)
Instructor Perceptions Proportion of Class Time
50
25
0
-25
-50
0
25
50
75
100
Group
Prefer Less
Agree
Prefer More

q) r)

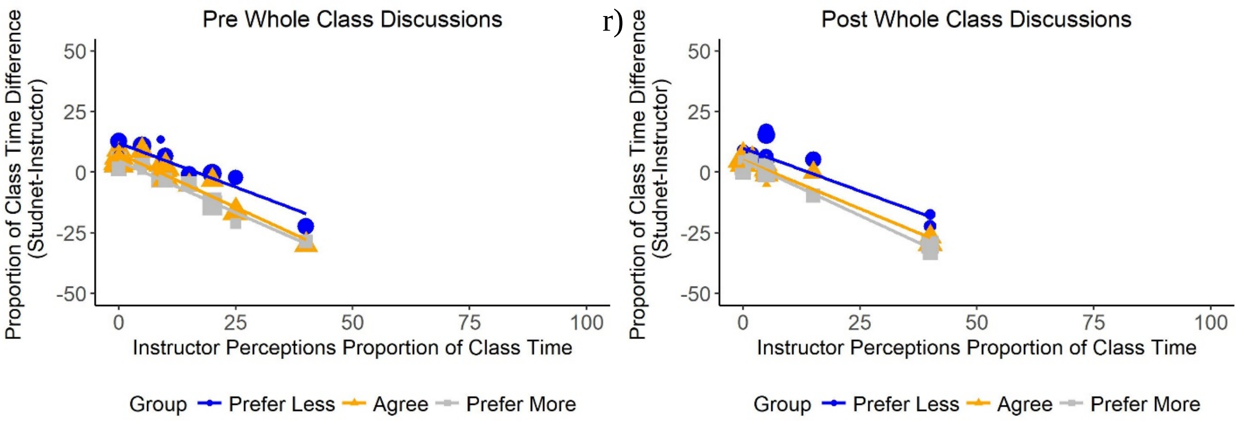


Figure 12. The difference in perceptions between students and instructors for each activity at the start and end of a semester. Marker size represents the proportion of the class in that group.

I examine student understanding between the groups of students who prefer less of an activity, more of an activity, and agree with the activity in order to explore a possible explanation for the differing perspectives phenomenon through the perspectives first approach. The plots in Figure 12 graph the difference between students' and instructors' perceptions of the time spent in class versus the instructors' perceptions of the time spent in class. Each data point is the average of the student-instructor difference for the three groups of students (those who prefer less/equal/more time spent on an activity) in a class. Each data point is sized by the proportion of students in their class that are categorized as each group, with larger proportions corresponding to a larger marker size. If a data point lies above the x-axis, the group of students are "overestimating" the amount of time spent on that activity. If a data point lies below the x-axis, the group of students are "underestimating" the amount of time spent on that activity. There are several points to note when reading these plots:

1) Because the x-axis is the instructor's perception of the time spent on the plotted activity, students from one class all have the same x-value. By observing the variation in the x-

values, one observes the variation among the instructors in how they spend time in their class on that activity.

2) The values of the y-axis represent a group of students' average over/underestimation of the plotted activity. A y-value of zero indicates that group of students correctly understood the time spent on that activity. Values further from zero indicate greater misunderstanding.

3) The trendlines represent the effect that varying instructor perceptions have on student perceptions of the class. For steeper slopes, student over/underestimation varies more with instructor perceptions than less steep slopes. The distance between the trendlines helps to visualize how student perceptions in general for one group compares to the others.

4) The size of the marker represents the proportion of the students in that class that are categorized into that group. Larger markers correspond to a larger proportion of students in that group.

In every activity, the students who prefer less of an activity perceive that activity as one that will or did happen more frequently than at least one of the other two groups of students. This is shown visually through Figure 12 as the "prefer less" group is always above the other groups. It may be that the students who prefer less of an activity are overestimating the amount of time the instructor says will be spent on the activity more so than the other students. An example of this is seen in Figure 12a and 12b with the lecture activity; the "prefer less" group responses are more positive than the other groups, meaning they are overestimating the amount of time spent on lecture more than the other groups. In such a case, the students who prefer less of an activity will have a greater misunderstanding compared to the other groups. It may also be that students who prefer less of an activity are underestimating the time spent on that activity to a lesser degree than the other students. An example of this is seen in Figure 12i and 12j with the small group work activity. The "prefer less" students, had a less negative understanding than the other groups, meaning they are underestimating the time spent on small group work less than the other students. In such a case, the students who prefer less will have a statistically significant better understanding of an activity. Another possible scenario is if the students who prefer less are overestimating the time spent on an activity while the other groups of students are underestimating. Whole class discussions at the start of a semester (Figure 12q) exhibits this scenario. Students who "prefer less" are largely overestimating while many of the other groups of students are underestimating. In this case, there may not be a statistically significant greater misunderstanding between the groups, because misunderstanding is defined as the magnitude of the difference between instructor and student, thus an overestimate and an underestimate could have the same magnitude.

Now, I explore the differences between the three student groups in terms of the magnitude of their over or underestimation of what the instructor says was the time spent on activities in class, or in other words, the magnitude of their misunderstanding. I find that in 35 out of 72 possible instances, students who preferred less of an activity had a statistically significant different level of understanding than the other two groups. I measure misunderstanding of an activity by taking the absolute difference between what the students perceive will or did occur for an activity and what the instructor perceived. In Figure 12, this is a measure of how far each data point is from the x-axis. Table 28 outlines the significant differences in student understanding between the group of students who prefer less of an activity and the other two groups of students. To obtain the results summarized in Table 28, I transformed each students' understanding score to a z-score based on their class. This accounts for any class level differences in their understanding so that I may run an analysis across all classes at once. I find that for many activities over each distribution of the survey, there is a greater misunderstanding from the students who prefer less of an activity than the other two groups of students. This is indicative of these students overestimating the amount of time that will be or was spent on that activity in their actual class. On the other hand, for the small group work activity, I find that students who prefer less of that activity actually have a statistically significantly better understanding than the other groups of students. This means that these students have the correct understanding or are underestimating the time spent on small group work to a lesser degree than the other groups of students are underestimating it.

In some cases, such as with student presentations and whole class discussion, there were distributions of the survey in which there was no difference in understanding between the three groups of students. How can there be no statistically significant difference in understanding when I know that there was a statistically significant difference in perceptions? Two scenarios could yield such a result. First, a “no difference” result could mean that students who prefer less of an activity were overestimating the amount of time spent to roughly the same degree that the other groups of students were underestimating the time spent on that activity. For example, if students who prefer less of an activity overestimate the time spent on that activity by five percentage points and the other groups of students underestimate by five percentage points, the misunderstanding measure for all three groups would be five percentage points, thus not different. It is also possible that a “no difference” result stems from the transformation of over/underestimation values to misunderstanding values. During this transformation (recall that the misunderstanding measure is the absolute value of the over/underestimation measure), the average for a group of students may change if the range of responses for that group contained zero. If the range of responses for a group of students in the over/underestimation measure contains zero, then taking the absolute value of these responses will increase the average value. For example, imagine four students in the "agree" group whose over/underestimation values are -1, 0, 1, and 2. The average for this group would be 0.5. When transforming this to the misunderstanding measure, their scores are now 1, 0, 1, and 2 which has an average of 1. Because this transformation increased the average, bringing this group's average closer to the "prefer less" group which I know has a higher average, a statistical analysis is less likely to find a significant difference between the two groups' averages (i.e. a "no difference" result).

Table 28. Statistically significant differences of student misunderstanding z-scores between students who prefer less of an activity and students who agree or prefer more of that activity.

Student understanding by agreement group			Beginning of Fall 2017		End of Fall 2017		Beginning of Spring 2018		End of Spring 2018	
			...students who agree.	...students who prefer more.	...students who agree.	...students who prefer more.	...students who agree.	...students who prefer more.	...students who agree.	...students who prefer more.
Students who prefer less…	Lecture	…compared to…	GM	GM	GM	GM	GM	-	-	GM
	Quiz		-	-	-	GM	GM	GM	-	-
	Step-by-step labs		-	LM	LM	LM	-	GM	-	-
	Investigative labs		-	-	GM	GM	GM	GM	-	-
	Small group work		LM	LM	LM	LM	LM	LM	LM	LM
	Individual work		GM	GM	-	-	GM	GM	GM	GM
	Student presentations		-	-	LM	-	-	-	-	-
	Demonstrations		GM	GM	-	GM	-	-	-	-
	Whole class discussion		-	-	-	-	-	-	-	-

Greater Misunderstanding (GM). Lesser Misunderstanding (LM). No Difference (-). All differences are statistically significant to the $p<.05$ level.

The major takeaway from Table 28 is that in many cases, it is true that the group of students who prefer less of an activity are misunderstanding how the class is supposed to or did operate. This occurred particularly often for the “lecture” and “individual work” activities, and slightly less often for “quiz”, “investigative labs”, and “demonstrations”. Only in some cases (e.g. “small group work”) could we say that the group of students who prefer less of an activity might have an accurate understanding of the class format regarding that activity. These results show that the different perspectives phenomenon occurs for several activities, most notably “lecture” and “small group work” which are two of the most distinctive activities in a studio class (the lack of lecture and the prevalence of group work). To aid in gaining student buy-in to these activities by establishing the first step in student buy-in (proper understandings of the class format), instructors must be extremely clear as to how much time they will spend on each activity so that there is no chance for misunderstanding among the students.

Through the exploration into the different perspectives phenomenon, I have addressed much of question 1a: “How well do students understand the class format?” I have learned that different groups of students understand the class format differently depending on how they agree with the class format. To fully address Q1a, I examine the magnitude and direction of the misunderstanding among the students for each activity, as shown in the plots of Figure 12. Doing so also answers Q1b: “What specific class activities do students properly or improperly understand?”

I find that there were several activities with large misunderstandings: lecture, step-by-step labs, small group work, and individual work.

Lecture was overestimated to a large degree by students who prefer less lecture in their HMEC. At the start of a semester, the groups of students who preferred less time spent on lecture usually perceived it as occurring about 20 percentage points more than the instructor said it would occur. This remained fairly consistent by the end of the semester; the groups of students who preferred less lecture by the end of the semester perceived lecture as occurring about 30 percentage points more than the instructor did. Even the groups of students who agreed and preferred more lecture were overestimating the time spent on lecture, except for students in the class where the instructor indicated 60% of the class was spent on lecture.

Step-by-step labs are often underestimated by students, especially if the instructor indicated that step-by-step labs would occur for 20% or more of the class time. Across all groups of students, if the instructor said 20% or more of the class would be devoted to step-by-step labs, students tended to underestimate this by up to 40 percentage points.

The amount of time that will be spent or was spent on *small group work* is almost always underestimated by all students. However, students who prefer less small group work have a better understanding of how much time will be spent on small group. These students are underestimating small group work to a lesser degree than other students.

Finally, the amount of *individual work* involved in their studio classes is overestimated by students. Many instructors indicated that individual work would not or did not occur for their class. This message is evidently not getting through to the students because almost all groups of students said individual work would or did occur for up to 20% of the class time.

Results such as these can help inform instructors in the future about which class activities require more clarity as to how much time will be spent on them. For example, it is clear that students have incorrect expectations for the amount of time that will be spent on lecture, step-by-step labs, small group work, and individual work. Instructors can learn from this and adjust their presentation about the class format regarding these activities. Instructors should also note that students that want less of an activity have a different perception of that activity than students who agree or want more of that activity and so these students may require extra effort from the instructor when establishing expectations for those activities. What is intriguing, and would require further research, is why by the end of the semester students continue to have different perceptions from the instructor of how the class time was spent. It appears that for some reason, how the instructor perceives past class time is different than how students perceive past class time, thus resulting in the measured misunderstanding for some activities at the end of a semester.

6.1.3 Student Agreement

The second aspect of student buy-in, after proper understandings of the class format, is student agreement. A student's agreement depends only on the student's own perceptions and not the accuracy of those perceptions because student agreement is based on two student reported values: what the student perceives for their actual class and what they would prefer for their most effective class. It is irrelevant if their perceptions are accurate because the important measure of student agreement is the difference between whatever they perceive and what they would prefer. Thus, there is no need to worry about the differing perspectives phenomenon when exploring student agreement. However, it is still important to separate students on the basis of having made the time-importance mistake once, more than once, or not at all. For the analyses on student agreement, only students who did not make the mistake and any groups of students that responded similarly to those who did not make the mistake for the measurement under analysis were included.

In this section, I answer questions 1c-e, which ask how well students agree with the class, which specific activities they agree or disagree with, and how these change throughout the semester.

6.1.3.1 Overall Agreement

Agreement with the class in general is measured through agreement with 1) importance of in-class activities, 2) importance of out-of-class activities, 3) time spent on in-class activities, and 4) time spent on out-of-class activities. In Tables 29 and 30, I report average student agreement (and standard deviations) with the class format through the Fall 2017 and Spring 2018 semesters for each class. Recall that for the importance measures and the out-of-class time agreement measure, the value reported is the average proportion of class activities that students agree with. The in-class time agreement measure does not have a simple, concrete interpretation like the other measures, but all measures, including in-class time agreement, are higher values when students agree more with the class format.

There are some important things to keep in mind when looking at Tables 29 and 30. First, due to the time-importance mistake, there are a different number of students for each agreement measure. Only students who responded similarly on the measure to students who did not make the mistake were included when reporting the results for that measure. In some cases, this severely limited the sample size within a class. For example, in class 4, no students were kept for the importance measures and out-of-class time agreement measures at the end of the semester. The student groups that were used for each measure were not always consistent between the beginning and end of the semester. For instance, at the beginning of the Fall 2017 semester, students who made the mistake once could be included with those that did not make the mistake for the out-of-class time agreement measure. However, at the end of the semester, only students that did not make the mistake were used. In these cases where the usable student groups differed across the semester, when looking at the change in agreement from beginning to end of the semester, the students used were those who were used at both the beginning and end of the semester. So, following from the previous example, only students who did not make the mistake or made the mistake once at the beginning of the semester *and* did not make the mistake at the end of the semester were used when exploring the change in student agreement. Also notice that instead of standard deviations, I report standard errors (SE) for the change in student agreement across all classes. This is due to the different means in which these values were obtained; standard deviations apply to descriptive statistics while standard errors apply to inferential statistics. For the majority of the table, the values are an average of a population. For this section, however, the values reported are an estimate resulting from a longitudinal hierarchical linear model, thus I report the standard error instead of the standard deviation. One final note is that classes 11 and 12 did not take the PUMBA at the end of the semester.

Table 29. Overall agreement with the course format throughout the Fall 2017 semester. Each value is reported as “Average (Standard Deviation or Standard Error (SE))”.

Class	Beginning of Semester				End of Semester				Change (End - Beginning)			
	In-Class Importance (N=333)	Out-of-Class Importance (N=213)	In-Class Time (N=277)	Out-of-Class Time (N=284)	In-Class Importance (N=242)	Out-of-Class Importance (N=138)	In-Class Time (N=340)	Out-of-Class Time (N=111)	In-Class Importance (N=104)	Out-of-Class Importance (N=49)	In-Class Time (N=139)	Out-of-Class Time (N=46)
1	57 (21)	60 (26)	82 (18)	58 (24)	72 (22)	66 (25)	79 (13)	60 (31)	**29 (22)***	34 (26)	2 (26)	38 (13)
2	57 (25)	57 (27)	79 (18)	59 (26)	53 (20)	73 (26)	72 (18)	57 (25)	4 (31)	13 (27)	1 (20)	-18 (27)
3	58 (20)	51 (26)	82 (19)	48 (23)	72 (21)	71 (26)	77 (20)	40 (38)	15 (32)	22 (36)	-5 (18)	-18 (42)
4	50 (25)	63 (31)	78 (19)	47 (24)	-	-	75 (13)	-	-	-	15 (49)	-
5	63 (22)	54 (29)	74 (13)	60 (17)	61 (23)	66 (24)	72 (21)	38 (10)	-8 (19)	-50 (0)	**-12 (7)***	-50 (0)
6	58 (15)	39 (20)	80 (27)	42 (18)	57 (24)	54 (20)	70 (26)	65 (27)	-13 (28)	4 (19)	-15 (45)	8 (26)
7	57 (21)	56 (18)	80 (16)	56 (19)	61 (23)	63 (25)	75 (16)	63 (24)	4 (27)	2 (26)	-3 (17)	-21 (27)
8	64 (21)	57 (20)	85 (13)	58 (21)	66 (19)	63 (20)	80 (13)	63 (20)	0 (20)	4 (25)	-2 (12)	4 (37)
9	60 (23)	52 (20)	77 (11)	54 (22)	61 (24)	62 (25)	79 (15)	58 (23)	-9 (27)	13 (25)	-2 (17)	7 (15)
10	54 (23)	50 (14)	85 (9)	50 (20)	63 (21)	57 (28)	82 (11)	55 (23)	15 (31)	25 (25)	-6 (8)	33 (20)
11	63 (22)	64 (21)	82 (14)	51 (18)	-	-	-	-	-	-	-	-
12	61 (22)	65 (19)	77 (12)	59 (23)	-	-	-	-	-	-	-	-
Average of all students	58 (22)	56 (23)	80 (16)	54 (23)	64 (22)	64 (25)	77 (16)	57 (27)	4 (SE=5)	**11 (SE=4)****	**-3 (SE=2)****	1 (SE=6)

* p-value of paired t-test or Wilcoxon ranked-sum test is less than 0.05.

** p-value of time effect in a hierarchical linear growth model is less than 0.05.

Table 30. Overall agreement with the course format throughout the Spring 2018 semester. Each value is reported as "Average (Standard Deviation or Standard Error (SE))".

Class	Beginning of Semester				End of Semester				Change (End - Beginning)			
	In-Class Importance (N=211)	Out-of-Class Importance (N=136)	In-Class Time (N=262)	Out-of-Class Time (N=239)	In-Class Importance (N=155)	Out-of-Class Importance (N=93)	In-Class Time (N=234)	Out-of-Class Time (N=141)	In-Class Importance (N=92)	Out-of-Class Importance (N=43)	In-Class Time (N=159)	Out-of-Class Time (N=88)
13	65 (23)	61 (20)	81 (16)	48 (24)	68 (26)	71 (26)	82 (13)	65 (30)	5 (29)	13 (21)	1 (19)	**17 (26)***
14	61 (23)	62 (26)	78 (17)	59 (26)	61 (25)	74 (24)	75 (15)	59 (25)	-1 (25)	-1 (28)	-2 (19)	0 (29)
15	65 (26)	54 (31)	78 (20)	47 (27)	63 (21)	45 (25)	70 (17)	44 (31)	-4 (17)	-23 (29)	**-13 (15)***	10 (44)
16	55 (16)	54 (20)	82 (13)	60 (24)	63 (23)	57 (27)	83 (12)	58 (22)	8 (26)	6 (31)	-1 (18)	1 (32)
17	63 (28)	54 (24)	81 (16)	53 (22)	70 (18)	73 (21)	85 (13)	66 (27)	3 (20)	-13 (0)	**7 (8)***	-2 (15)
18	66 (26)	67 (27)	82 (15)	43 (30)	59 (24)	65 (29)	89 (11)	63 (30)	-7 (15)	0 (0)	**16 (18)***	24 (40)
Average of all students	62 (23)	58 (24)	80 (16)	52 (26)	64 (24)	65 (27)	81 (15)	60 (27)	3 (SE=3)	0 (SE=5)	1 (SE=4)	**9 (SE=5)****

* p-value of paired t-test or Wilcoxon ranked-sum test is less than 0.05.

** p-value of time effect in a hierarchical linear growth model is less than 0.05.

At the beginning of the Fall 2017 semester, I find that for the importance agreement measures and the out-of-class time agreement measure, students on average agree with 50-60% of the in-class or out-of-class activities. As a reminder, these scores represent the average number of activities that students agree with out of the total number of activities for that measure. For example, figure 13 shows a histogram of the number activities agreed with for in-class importance for all students at the beginning of the Fall semester. The histogram models a normal distribution centered on five activities agreed with out of nine total (an agreement score of 55%). This figure is representative of the distributions for the number of activities agreed with for all importance measures and out-of-class time measures in each distribution of the PUMBA. I specifically find that students across all classes agreed with the importance placed on an average of 58% of in-class activities and 56% of out-of-class activities. Also, across all classes, students agreed with the time spent on an average of 54% of out-of-class activities. Clearly, student agreement with these aspects of a studio class could be improved. An instructor may take this information and better prepare their class orientation to discuss the importance of class activities and how they expect the students to spend their time on out-of-class activities. For the time spent on in-class activities, student agreement, across all classes, is relatively high with a value of 80 out of a maximum possible value of 100. This indicates decent agreement with the time spent on in-class activities but with room for improvement.

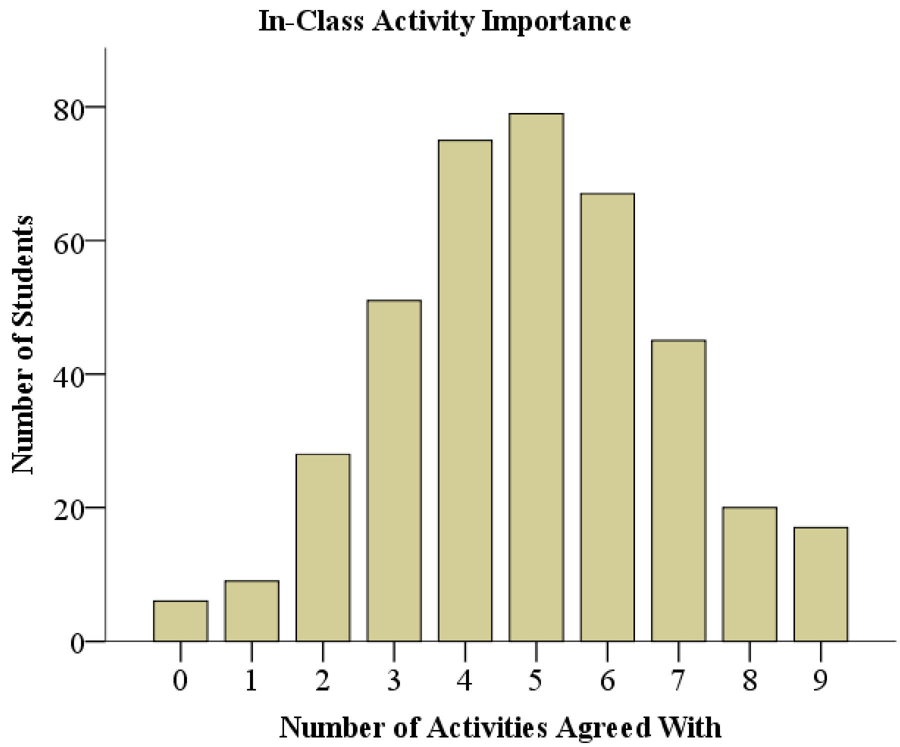


Figure 13. The distribution of the number of in-class importance activities students agreed with in the Fall 2017 semester.

By the end of the Fall 2017 semester, the averages across all classes increased for the importance agreement measures and the out-of-class time agreement measure. However, for the in-class time agreement measure, the average across all classes decreased from the beginning to the end of the semester. These changes in agreement may be due to differing populations at the beginning and end of the semester. A better indication of how student agreement changed over the semester examines the change in student agreement between two surveys matched to the same student. This analysis is reported on the right side of Tables 29 and 30 under the "Change (End – Beginning)" header. For this section, only students who took the survey at both points in the semester were considered. In addition to running descriptive statistics to determine the class average and standard deviation in the change of an agreement measure, I also ran inferential statistics to see if this change was statistically significant. Specifically, I ran a paired t-test or a Wilcoxon ranked-sum test (if normality or homoscedasticity assumptions were violated for the paired t-test) on each class. The changes which were statistically significant at the $\alpha=0.05$ level are bolded and marked with "*". There were few changes in the Fall semester that were significant within one class, possibly due to the low sample size in a single class. Class 1 has a significant increase of 29 percentage points for agreement to the importance of in-class activities for class 1. The effect size, as measured by Cohen's d, is 1.44 which is a very large effect (Rosenthal, 1996). The other within-class significant change in agreement occurred in class 5 regarding agreement to the time spent on in-class activities. Agreement dropped by 12 percentage points, an effect size of 0.73, which is a medium to large effect.

When looking across all classes, I use a hierarchical linear growth model. Statistically significant changes are found if the fixed effect of the time variable is a significant predictor in the model. I find a statistically significant change for student agreement with the importance placed on out-of-class activities; it increases by 11 percentage points, equivalent to one more activity that students agree with. This is a small effect size of 0.07. There is also a statistically significant decrease in agreement with the time spent on in-class activities by 4 percentage points. This is also a small effect size of 0.08. Changes in in-class importance agreement and out-of-class time agreement were not statistically significant.

In the Spring semester, the agreement levels at the beginning of the semester are similar to the Fall semester. Students agree with 50-60% of the activities in terms of their importance or the time spent on out-of-class activities. Again, the overall student agreement with the time spent on in-class activities was 80, which is relatively high but could still be improved. When looking at the agreement measured at the end of the semester, I find an improvement for each agreement measure. However, this does not compare the same students. Using students who took the survey at the beginning and end of the semester, I find that only the time agreement measures change significantly for some classes. Class 13, the instructor of which also taught class 1, had a statistically significant improvement in their agreement with the time spent on out-of-class activities. This change is considered a small to medium effect size (Cohen's d = 0.44). This instructor also had a large improvement in this measure in the Fall semester, but it was not statistically significant. In the Fall semester, this instructor did achieve an increased agreement with the importance of in-class activities but this is not found in the Spring semester. Class 15, whose instructor taught class 4, had a large decrease in agreement with the time spent on in-class activities (Small to medium effect size, Cohen's d = 0.44). Classes 17 and 18, taught by the same instructor who taught class 10, had a statistically significant increase in their agreement with the time spent on in-class activities. The effect size for the change in agreement for class 17 is small (Cohen's d = 0.28) and medium (Cohen's d = 0.53) for class 18. No instructor had consistent statistically significant changes in agreement for any of the four agreement measures. Likewise, there is no consistency in statistically significant changes across all classes each semester. In the Fall semester, out-of-class importance agreement increased and in-class time agreement decreased to a statistically significant degree. In the Spring semester, both of these agreement measures hardly changed across all classes. The only significant change across all classes in the Spring semester occurred with the time spent on out-of-class activities, increasing by nine percentage points, but with a very small effect size less than 0.04. These inconsistencies between classes taught by the same instructors might be evidence that instructors have little effect on student agreement throughout the semester. There were similarities in student agreement at the beginning of the semester suggesting that instructors might have a consistent effect on students at that point in time. However, as the semester continues and student agreement changes, it changed in different ways for different students with the same instructor, suggesting that the instructor might not have a noticeable effect on student agreement throughout the semester. This will be explored further when I investigate instructor methods and their effect on student buy-in.

By looking down the columns of Tables 29 and 30, I can examine the distribution of class averages for each overall agreement measure. I find that there is little variation in average agreement score across classes. For the importance measures and the out-of-class time agreement measure, the class average typically falls between 45-65 at the start of a semester. For the in-class time agreement measure, the class averages at the start of a semester range from 74-85. By the end of the semester, although classes overall may have increased or decreased in an agreement measure, there is still some consistency between classes. I also find that the variation of student agreement within a class is large, with standard deviations typically in the 20s which is about one third to one half of the magnitude of the average. The standard deviations for in-class time agreement is, unsurprisingly, smaller than the rest. This metric is more precise than the others as students can indicate exactly how much more or less of an activity they would prefer. Although these variations within class are smaller than the other agreement measures, they are still on the same order of magnitude as the average, meaning they are somewhat large. From class to class, student agreement is on average the same. Within a class, however, there is a large range of student agreement levels; some students strongly disagree with the class format and some students strongly agree.

6.1.3.2 Agreement with Specific Activities

Naturally, the next question to ask is what activities exactly are students agreeing or disagreeing with (Q1d)? Within each class, I report to instructors the proportion of students in their class who agree with each activity regarding the importance and time spent on it. I also report the proportion of students that would prefer more or less importance or time spent on that activity so that the instructors have an idea of how they should tailor their message to the students to increase agreement. Classes do not differ from each other in the specific activities their students agree or disagree with. There were few instances where a class had a statistically higher or lower agreement for an activity than the rest of the classes (see Table 31). I ran a chi-square goodness of fit test on the number of students who agreed with an activity across each class, comparing that number to the expected number of students if each class had the same proportion of students who agreed. The expected proportion of students who agree was defined as the average of the proportions of students who agree in each class. Statistically higher or lower agreement for a class was determined by the standardized residuals of this test with a magnitude larger than 2 (Agresti, 1996). Few differences were found between the classes. Some examples of a class having a statistically significant different agreement level with an activity include class 3 having higher agreement with lecture time and demonstrations time at the beginning of the semester, and class 14 having higher agreement with the importance of extra reading at the end of the semester. There were only 14 instances of a class having a statistically significant different agreement level than other classes across all activities for importance and time measures and across all distributions of the survey. I will examine trends across all classes when describing which activities students typically agree or disagree with. Student agreement across all classes is found by averaging the proportion of students who agree with an activity for each class. I do this instead of using the proportion of all students who agree because that value would be biased towards the agreement level of larger classes.

Table 31. Activities for which a class has a different proportion of agreement compared to the other classes. Cells shaded in blue have a lower proportion of agreement and cells shaded in orange have a higher proportion of agreement.

	Beginning of Semester						End of Semester			
Class	Online Resources Importance	Lecture Time	Small Group Work Time	Demonstrations Time	Whole Class Discussion Time	Assigned Reading Time	Student Presentations Importance	Small Group Work Time	Individual Work Time	
1	62%	26%	62%	47%	66%	45%	30%	44%	63%	
2	65%	27%	39%	41%	34%	58%	38%	41%	31%	
3	40%	45%	55%	62%	49%	61%	21%	61%	52%	
4	43%	27%	36%	52%	55%	52%	-	25%	25%	
5	69%	13%	44%	13%	7%	42%	-	33%	56%	
6	25%	38%	63%	75%	38%	14%	25%	50%	36%	
7	50%	42%	33%	38%	45%	65%	25%	40%	33%	
8	55%	33%	50%	40%	39%	77%	35%	33%	41%	
9	42%	17%	19%	33%	42%	32%	56%	41%	45%	
10	18%	27%	32%	50%	50%	67%	-	-	-	
11	56%	32%	52%	48%	47%	44%	-	-	-	
12	49%	36%	36%	28%	29%	59%	-	-	-	
	Beginning of Semester						End of Semester			
Class							Extra Reading Importance	Lecture Time	Investigative Labs Time	Whole Class Discussion Time
13	N/A						51%	41%	52%	62%
14							66%	27%	20%	34%
15							45%	10%	62%	41%
16							36%	32%	61%	48%
17							30%	30%	70%	63%
18							33%	57%	60%	87%

In Tables 32 and 33, I report the top three most agreed with activities and top three most disagreed with activities for each distribution of the PUMBA. Alongside the activities is the proportion of students on average who agree with that activity. Also included with the most disagreed with activities is the proportion of students who would prefer less and more importance or time spent on that activity in their most effective class compared to their actual class and the amount of time more or less that students would prefer of the in-class activities if they disagreed with it. Note that for out-of-class time agreement, the proportions may not sum to 100% because students are given an option to say that they have no feeling as to how much time they are intended to spend on an activity outside of class. I obtained these values by treating each class as having one value and averaging those across the classes. For example, 50% of class A agrees with activity X and 60% of class B agrees with activity X, so the reported value would be 55%. In doing so, one class cannot adversely affect the results as it receives equal weighting as all other classes.

Tables 32 and 33 are similar to how I report the PUMBA results to the instructors. I first highlight aspects in which they are doing well in achieving student agreement. Then I point out the aspects and activities for which they could work to improve student agreement. To aid in the development of their message to the students, I provide instructors with the information on what their students would prefer so that instructors know if they need to emphasize why an activity is less/more important or occurring less/more often in the actual class. I emphasize in the report that this information is not to tell the instructors how they should be teaching their class to appease their students, but instead how to tailor their message to get students to buy-in to how the class is intended to be taught.

In-class Importance: Looking first at the in-class importance agreement measure, I find that lecture and quiz were often among the most agreed with activities. These two activities were in the top three most agreed with activities for each survey distribution aside from the end of the Fall semester. In both semesters, demonstrations and step-by-step labs appeared as two of the least commonly agreed with activities. For both activities, the majority of students that disagreed with these activities would place greater importance on them in their most effective class than they felt their actual class was going to. By the end of the semesters, step-by-step labs were no longer one of the most commonly disagreed with activities, but in the Fall semester, despite an overall increase in the proportion of students who agreed with it, demonstrations remained the most disagreed with activity in terms of its importance. Students continued to prefer a greater importance placed on demonstrations.

Table 32. The top three most agreed with and most disagreed with activities for Fall 2017 and for each agreement measure.

		Beginning of the Semester						End of the Semester					
		Top 3 Most Agreed		Top 3 Most Disagreed				Top 3 Most Agreed		Top 3 Most Disagreed			
		Activity	Average %	Activity	Average %	Average % Preferring Less (Average percentage points less preferred for In-class Time)	Average % Preferring More (Average percentage points more preferred for In-class Time)	Activity	Average %	Activity	Average %	Average % Preferring Less (Average percentage points less preferred for In-class Time)	Average % Preferring More (Average percentage points more preferred for In-class Time)
Fall 2017	In-class Imp.	Lecture	71	Demo.	42	5	53	Step-by-step labs	69	Demo.	59	5	36
		Quiz	68	Step-by-step labs	55	10	35	Small group work	68	Lecture	60	11	29
		Whole class discussion	62	Student Present.	57	34	9	Inv. labs	66	Whole class discussion	60	5	35
	Out-of-class Imp.	Lab Reports	61	Extra Reading	47	38	15	Homework	70	Extra reading	56	28	15
		Reading	60	Online Resources	48	18	35	Group projects	70	Extra homework	57	24	18
		Homework	58	Help	49	7	44	Lab reports	69	Online resources	59	16	25
	In-class Time	Quiz	60	Lecture	30	25 (-15)	45 (+17)	Student present.	68	Lecture	28	36 (-16)	37 (+17)
		Student present.	58	Whole class discussion	42	30 (-8)	28 (+7)	Quiz	58	Step-by-step labs	39	38 (-11)	23 (+10)
		Inv. labs	48	Small group work	43	39 (-12)	18 (+7)	Inv. labs	52	Small group work	42	38 (-11)	20 (+8)
	Out-of-class Time	Homework	60	Extra Reading	41	36	16	Homework	69	Reading	42	32	20
		Lab Reports	59	Online Resources	43	16	34	Help	54	Online resources	45	15	33
		Help	55	Extra Homework	44	23	28	Extra homework	53	Extra reading	46	31	11

Table 33. The top three most agreed with and most disagreed with activities for Spring 2018 and for each agreement measure.

		Beginning of the Semester						End of the Semester					
		Top 3 Most Agreed		Top 3 Most Disagreed				Top 3 Most Agreed		Top 3 Most Disagreed			
		Activity	Average %	Activity	Average %	Average % Preferring Less (Average percentage points less preferred for In-class Time)	Average % Preferring More (Average percentage points more preferred for In-class Time)	Activity	Average %	Activity	Average %	Average % Preferring Less (Average percentage points less preferred for In-class Time)	Average % Preferring More (Average percentage points more preferred for In-class Time)
Spring 2018	In-class Imp.	Quiz	77	Demo.	56	3	41	Lecture	64	Student present.	35	58	6
		Lecture	71	Step-by-step labs	57	13	30	Individual work	64	Whole class discussion	51	23	25
		Individual Work	62	Small group work	58	21	21	Quiz	60	Inv. labs	52	27	21
	Out-of-class Imp.	Homework	60	Reading	56	32	12	Help	68	Lab reports	41	49	10
		Lab Reports	60	Online Resources	56	21	23	Homework	64	Extra reading	43	44	12
		Group Projects	60	Extra Reading	57	32	11	Reading	64	Group projects	48	35	17
	In-class Time	Student present.	62	Lecture	35	30 (-16)	35 (+15)	Student present.	85	Lecture	33	36 (-15)	31 (+14)
		Quiz	61	Small group work	41	42 (-12)	17 (+8)	Quiz	61	Small group work	40	39 (-10)	21 (+6)
		Whole class discussion	58	Demo.	44	18 (-7)	38 (+8)	Whole class discussion	56	Step-by-step labs	42	29 (-9)	30 (+9)
	Out-of-class Time	Homework	61	Extra Reading	33	43	12	Homework	59	Online resources	45	17	31
		Reading	55	Online Resources	41	20	31	Lab reports	59	Extra reading	46	30	12
		Group Projects	53	Lab Reports	46	29	12	Group projects	55	Help	47	11	36

Out-of-class Importance: Next, looking at student agreement with the importance placed on out-of-class activities, I find that at the start of a semester, students commonly agree with the importance placed on assigned conceptual/problem solving homework ("homework") and lab reports. In both semesters, roughly 60% of students agreed with the importance placed on these activities. The importance of assigned homework and lab reports remained among the most commonly agreed with out-of-class activities in the Fall semester. In fact, agreement with the importance of these activities increased over the semester. The same can be said for the importance of assigned homework in the Spring semester. Interestingly, lab reports, dropped from the most commonly agreed with activity at the beginning of the Spring semester to the least commonly agreed with activity at the end of the semester. Nearly half of all students would have placed less importance on lab reports in their most effective class. Classes 14, 15, and 18 are the main reasons for this dramatic change. At the beginning of the semester, 62%, 61%, and 89% of these classes respectively agreed with the importance placed on lab reports. By the end of the semester, these proportions decreased to 47%, 27%, and 42%. Other out-of-class activities with which students commonly disagreed with the importance placed on them were reading beyond what is assigned ("extra reading") and using other online resources ("online resources"). Both of these activities remained as commonly disagreed with activities from the beginning to the end of the semester, except for the "online resources" activity at the end of the Spring semester. Students that disagreed with the importance placed on "extra reading" typically preferred less importance on that activity. Students favored more importance placed on "online resources".

In-class Time: The most agreed with activities for the time spent in-class remained the same from the beginning to the end of a semester. In both semesters, students commonly agreed with the time spent on quizzes and student presentations. At the start of the semester, roughly 60%

of the students agreed with the time spent on these activities. By the end of the semester, the same proportion of students agreed with the time spent on quizzes and a greater proportion of students agreed with the time spent on student presentations. Students expected or experienced both of these activities to occur infrequently, between 3-8% of the class for quizzes and 1-3% of the class for student presentations. So, students are agreeing that these activities should occur very rarely. In the Fall semester the “investigative labs” activity remained the third most agreed with activity from the beginning to the end of the semester and in the Spring semester the “whole class discussion” activity remained as the third most agreed with activity. Students were also consistent about the activities they disagreed with regarding the time spent on them in class. In every distribution of the survey, “lecture” was the least agreed with activity and “small group work” was the second or third least agreed with. Twenty-eight to thirty-five percent of the students surveyed agreed with the time spent on lecture. Although students often agreed with the importance placed on lecture, they did not agree with how much time they spent on it. One might expect students to always want more lecture than what they are getting, but somewhat surprisingly it is more often the case that a similar number of students prefer less lecture as the students who prefer more. Typically, at the start of the semester there are slightly more students who want more lecture but by the end of the semester students are uniformly distributed between the three agreement groups. With the PUMBA, I can determine how much more or less lecture these students want. For the “lecture” activity, the degree of disagreement with the activity is on average relatively large. Students would prefer an average ranging between 14 and 17 percentage points more or less lecture than what they perceive lecture will or did occur in their class. At the end of both semesters, the time spent on step-by-step labs became one of the least agreed with activities as well.

"Small group work" was also one of the least commonly agreed with activities, appearing in the bottom three for each distribution of the survey. On average only 40-43% of the students who took the survey agreed with the time spent in small group work. Unlike the "lecture" activity, there is a consensus on what the students would prefer in their most effective class: less time spent on small group work. The average proportion of students who preferred less small group work ranged from 39-42% while the average proportion of students who preferred more ranged between 17-21% across the four distributions of the survey. The students who preferred less small group work would have had 10-12 percentage points less time spent on that activity while the students who preferred more would have spent 7-9 percentage points more time.

"Step-by-step labs" appeared in the least agreed with activities at the end of both semesters with 39% and 42% of the students surveyed agreeing with that activity. In the Fall semester, most students who disagreed preferred less time spent on the labs (11 percentage points less time then what they perceived occurred in their actual class). In the Spring semester, there was no consensus on what the students preferred. Twenty-nine percent of the students in the Spring semester would prefer less time spent on step-by-step labs ,and 30% of the students would prefer more time. On average both sets of students would spend nine percentage points less or more time on step-by-step labs in their most effective class compared to their actual class.

Out-of-class Time: Lastly, looking at the agreement with the time spent on out-of-class activities, I find that, like the importance placed on it, students are agreeing with the time spent on completing assigned conceptual/problem solving homework ("homework"). This activity is consistently the most agreed with activity in this category. In the Fall semester, the time spent getting help outside of class via office hours, tutors, help center, etc. ("help") was a commonly agreed with activity at the beginning and end of the semester. In the Spring semester, the time spent working on group projects was commonly agreed with at the beginning and end of the semester. For the least agreed with activities in this category, completing reading beyond what is assigned (“extra reading”) and using online resources consistently appeared in the bottom three. “Extra reading”, which was also commonly disagreed with in terms of its importance, was an activity that students would prefer to do less often for their most effective class. Students would have preferred to spend more time using other online resources than the time they felt they were intended to spend on that activity for their actual class.

Tables 32 and 33 certainly present a vast amount of information that an instructor could learn much from. However, there are two key takeaways that I would like to emphasize regarding the “lecture” and “extra reading” activities. The time spent on lecture was the least agreed with activity out of all of the types of agreement. Only approximately one third of the students who took the surveys agreed with the time spent on lecture. This may not be too surprising given that the reduction of lecture in comparison to a traditionally taught physics class is a feature of the studio style. What is surprising is the fact that there are similar proportions of students who want less time spent on lecture as students who want more, especially by the end of the semester. This means that when an instructor is trying to get buy-in to lecture, they cannot only talk about why they are doing less lecture because there is a significant proportion of the student population who thinks there should be even less. An instructor will have to discuss why their class lectures to the amount that they do considering that some students will want more lecture and some will want less. Looking at the amount more or less lecture that students typically want (approximately 15 percentage points more or less lecture) shows that instructors must put in a large amount of effort to convince students why their class shouldn’t have that much more or less lecture. It is clear that students have a strong disagreement with the time spent on lecture, one way or the other, so if instructors want to achieve buy-in, discussing lecture should be a top priority.

The second major take-away from tables 32 and 33 is that students have some idea of the importance of "extra reading" needed for the class and how much time they feel they need to spend on it. Students may get this idea if the instructor recommends text to supplement their learning, or from an abstract idea of the amount of effort they would need to make outside of class to learn the material. In every distribution of the survey, the "extra reading" activity appeared as one of the least commonly agreed with out-of-class activities. Most students disagree with the importance placed on or time spent on the "extra reading" activity, typically preferring less importance or time. By working to get buy-in to the "extra reading" activity, instructors can eliminate the most common disagreement. First, instructors should be explicit about how much time they feel their students will need to do extra reading. Doing so achieves the first step in student buy-in: proper understandings. Then, instructors need to discuss why students should need to do that amount of extra reading. Students may see extra reading as having to teach themselves the material outside of class and disagree with that concept. For example, one student explains,

> *"So, when it came to physics by the end of the time, I realized it was a lot, it was very heavily reliant on what I did out of class, rather than what I did in class… So that's why I valued it equally like, equal time spent on actually reading the text out of class and beyond what's assigned."*

Another student found it necessary to do "extra reading" to understand the material, saying, "Well, for [reading] beyond what was assigned, I think in the beginning I was just, I didn't want to. And then afterwards, I figured that I had to, uh, in order to understand." Given that students may see the "reading beyond what is assigned" activity as equivalent to the concept of learning outside of class, instructors may want to discuss the benefits of learning outside of class, thus allowing for productive time spent in class, when trying to get buy-in to the “extra reading” activity.

6.1.3.3 Change in Agreement with Specific Activities

In tables 32 and 33 I examined some activities that were consistently top or bottom ranked in terms of their agreement from the beginning to end of the semester. I was looking for activities in which the agreement level did not change over the course of the semester. My next analysis focuses on activities for which student agreement did significantly change over the semester. To further investigate the change in student agreement to specific activities, and answer research question 1e, I use a hierarchical linear growth model for the time spent on in-class activities and a hierarchical logistic growth model for the importance placed on activities and the time spent on out-of-class activities. If the effect of the time variable is significant in the model, then the outcome variable, or agreement with the activity, changed between the beginning and end of the semester. A summary of the significant findings is found in Table 34.

Table 34. Summary of the significant changes in agreement with individual activities. Positive effects indicate greater agreement unless noted with "*".

Activity Agreement	Logistic "†" or Linear "*" Effect of Time on Agreement	p-value	R^2 Effect Size
Fall 2017			
Reading/annotating assigned text *importance*	0.15†	<.0001	0.16
Reading beyond what is assigned *importance*	0.87†	0.0313	0.06
Step-by-step labs *time*	2.15*	0.0014	0.34
Spring 2018			
Quiz/test *importance*	-0.93†	0.0087	0.14
Whole class discussion *time*	1.72*	0.0046	0.18
Reading beyond what is assigned *time*	0.78†	0.0034	0.06
* A positive effect for this activity indicates greater disagreement. This is due to how I calculate agreement to the time spent on individual activities; it is the magnitude of the difference between HMEC and actual class.			

Few activities changed to a statistically significant degree between the beginning and end of the semester. The activities with significant changes are discussed below.

Step-by-step Labs. In the Fall semester, the time spent on step-by-step labs changed significantly. For the time spent on in-class activities, the outcome variable used in the model is the average absolute difference in the proportion of class time between the student's most effective class and their actual class. For this measure, values closer to zero indicate greater agreement. The average absolute difference with step-by-step labs at the start of the semester was 4.65. However, there was a large variation in students' initial agreement level as measured by the variance of the random intercept in the hierarchical linear growth model (intercept variance = 9.3, p=.0045). By the end of the semester the average absolute difference increased by 2.15 percentage points, indicating less agreement with step-by-step labs as time went on. This was a statistically significant change (p=.0014) with a medium effect size (measured by R^2; Equation 16) of 0.34 but this effect also varied by student. The standard deviation of the effect of time on this absolute difference was 5.8 percentage points per unit time (p<.0001). Overall, this hierarchical linear growth model is showing that initially, there was a large variation in student agreement with the time spent on step-by-step labs. In general, as the semester progressed, agreement with the time spent on step-by-step labs decreased but this effect varied when looking at specific students. This finding is consistent with tables 32 and 33. At the beginning of the semester, agreement with the time spent on step-by-step labs was neither in the top or bottom three most agreed with activities. By the end of the semester, however, step-by-step labs did appear in the bottom three activities meaning that agreement with this activity worsened.

Reading Outside of Class. Agreement with the importance of reading outside of class, assigned or not, increased as the Fall semester progressed. By the end of the Fall semester, a student was 1.15 time more likely to agree with "assigned reading" than they were at the start of the semester. This represents a small effect size of 0.16. There was a greater increase in agreement with "extra reading". Students were 2.4 times more likely to agree with this activity at the end of the semester than they were at the start. The effect size for this change was 0.06, which is smaller than the effect size for "assigned reading" despite "extra reading" being a larger change. This indicates that there is more variation in the change of agreement with "extra reading" than there is for "assigned reading".

Quizzes. Agreement with three activities changed significantly during the Spring semester. Agreement with the importance placed on quizzes decreased over the Spring semester. The odds that students were to agree with the importance placed on quizzes at the end of the semester were reduced by a factor 2.54 compared to the odds that students were to agree at the start of the semester. In other words, students were 0.39 (the inverse of 2.54) times as likely to agree with the importance of quizzes at the end of the semester as they were at the beginning of the semester. This effect was statistically significant at p=0.0087 with a small effect size of 0.14. Referring back to table 33, quizzes remained as one of the top most agreed with activities for the importance placed on in-class activities, but the proportion of the class to agree with it decreased. As determined by the growth model, at the start of the semester the probability that a student would agree with quizzes was 83%. An odds ratio of 0.39 brings the probability of agreement at the end of the semester to 67% which is still relatively large compared to the other activities, despite the decrease in agreement.

Whole Class Discussion. The agreement with the time spent on whole class discussion also changed in the Spring semester. At the beginning of the semester, the average absolute difference between what students wanted in their most effective class and what they expected in their actual class was 2.7, though this value was not significant in the growth model and did have a significant amount of variation (5.14 p=0.016). In general, as the semester went on, the absolute difference increased by 1.7 (p=0.0046) with a small effect size of 0.18. Like step-by-step labs in the Fall semester, this effect varied from student to student. The measured variation in the growth model was 10.8 (p=0.0116). Overall, agreement with the time spent in whole class discussion worsened as the semester progressed, but this effect varied for individual students.

Reading Beyond What is Assigned. Lastly, the agreement with the time spent on “extra reading” increased over the Spring semester. At the beginning of the semester, the estimated probability of a student agreeing with the time spent on “extra reading” was 30% (p=0.0087) which is consistent with the actual proportion of students that agreed at the start of the semester. The probability of agreement increased by the end of the semester to 48%, a statistically significant increase ($p=0.0034$) with a small effect size of 0.06. The odds ratio for this effect is 2.18, meaning that the likelihood of agreeing with the time spent on extra reading was 2.18 time larger at the end of the semester than it was at the beginning. No other activities for any other agreement measures across either semester showed a statistically significant change in agreement using the hierarchical growth models.

Implications for Instructors. With information about how agreement changes, or doesn't change, during the semester instructors can learn which activities they may need to focus on throughout the semester for maintaining student buy-in. Activities such as step-by-step labs, quizzes, and whole class discussion lost agreement as the semester progressed so instructors may need to make an extra effort to maintain buy-in to these activities. Instructors may even use the information that agreement with "extra reading" increased over the semester. They might show these results to their students at the beginning of the semester to illustrate that the students' conceptions of the time spent on "extra reading" is not as bad as they may initially think. An instructor could say that by the end of the semester, more students agree with the time spent on extra reading than they did at the start of the semester so whatever the students are thinking now may be inaccurate. Considering that "extra reading" is one of the activities students most often disagree with, this information could be a valuable tool for increasing student buy-in to this activity.

6.1.4 Summary

Using the PUMBA, I aimed to determine how well students buy-in to their studio class. I did so investigating five questions. First, I wanted to determine how well students understand the class format and which activities they understand or misunderstand (Q1a-b); here, I operationalize "understand" as perceiving the class format similar to how the instructor perceives it. I found that student understanding was often dependent on if the students wanted less of an activity, agreed with it, or wanted more of an activity. Students who wanted less of an activity perceived that more time would be spent or was spent on that activity in their actual class. Thus, for most activities their understanding of the time spent on that activity was worse than the other students. The exception is for the "small group work" activity in which all students underestimated the time spent on it. Students that wanted less small group work underestimated to a lesser extent than the other students, thus having a better understanding of small group work. Overall, students had a poor understanding (measured as the magnitude of the difference in percentage points spent on an activity between what students and instructors perceive of the class) of the time spent on lecture, step-by-step labs, small group work, and in some cases, individual work. Students overestimated the amount of time spent on lecture and individual work, while they underestimated the amount of time spent on step-by-step labs and small group work. For all other activities, and in some cases with these four activities depending on the survey distribution or the student group, students had a decent understanding of the time spent on the activities.

Next, I explored how well students agreed with the overall class format (Q1c). Students typically agreed with the importance placed on 55-65% of class activities. Likewise, students agreed with the time spent on 50-60% of out of class activities. Agreement with the time spent on in-class activities generally ranged between 70 to 80 overlapping percentage points. Overall average agreement was consistent between classes but varied greatly among the students within individual classes. Agreement may change between the beginning and end of the semester, however there was no consistency in how it changed between the two semesters. In the Fall, agreement with the importance placed on out-of-class activities increased by 11 percentage points, and agreement with the time spent on in-class activities decreased by 3 percentage points. In the Spring semester, there were no significant changes in agreement for those two measures, but there was a statistically significant increase in the agreement with out-of-class activities by 9 percentage points.

Finally, I investigated how students agreed with specific activities and how their agreement changed from the beginning to the end of the semester (Q1d-e). One of the major findings is that students greatly disagree with the time spent on lecture. Though one might expect most students to prefer more lecture in their most effective class, this is not necessarily the case. There were similar proportions of students wanting less lecture as there were students wanting more lecture, especially so at the end of the semester. Students would prefer on average approximately 15 percentage points more or less lecture in their most effective class than their actual class. This means that instructors should put in a large effort to gain agreement with lecture and appeal to students who want more and less lecture, not just students who want more. The second major finding is the disagreement with the importance of and time spent on “extra reading”. This activity was one of the least agreed with activities in every survey. Instructors could eliminate a sizeable portion of disagreement just by working on increasing agreement with “extra reading”.

When looking at how agreement with specific activities changes over the semester, I found that few changed significantly. In general, agreement with the time spent on step-by-step labs decreased over the Fall semester but this varied from student to student. Agreement with the importance of quizzes decreased in the Spring semester, though it still remained one of the most agreed with activities in that category. Agreement with the time spent on whole class discussion decreased, in general, in the Spring semester, though this effect did vary from student to student. Lastly, agreement with the time spent on “extra reading” did increase in the Spring semester, though it remained one of the least commonly agreed with activities.

6.2 Student Agreement and Student Outcomes

In the previous sections, I described how well students agree or disagree with the importance placed on and time spent on in-class and out-of-class activities, both as overall measures and for each specific activity. My motivation for this research was to reduce student resistance to reformed instructional strategies by increasing student agreement. The idea was that if student resistance is reduced, instructors might have less difficulty implementing the reformed strategies, which are purported to increase student learning. To indirectly test this hypothesis, I used hierarchical linear and logistic modeling to see how student outcomes are predicted by student agreement. This does not test whether or not increased student agreement makes implementation of the research-based instructional strategies easier; rather, it simply tests if greater student agreement is associated with better student outcomes. These analyses answer research questions 2a-c, which separate student outcomes into student performance on concept inventories, attitudes, and expected final grade.

Student performance was measured through the conceptual tests the FCI (Hestenes et al., 1992) and the CSEM (Maloney et al., 2001). I also measured student performance based on their expected final grade as indicated in the version of the PUMBA completed at the end of the semester. I also investigate if student agreement predicts responses to the attitudinal survey the CLASS (Adams et al., 2006). Students often take the CLASS with conceptual tests at the beginning and end of the semester. It is a common measure to see how a class affects the level of expert-like views of and attitudes towards physics that students have. As it is an attitudinal survey, the CLASS is another survey of student affect, like the PUMBA. Thus, I might expect a connection between student agreement and the level of expert-like responses to the CLASS.

I used the performance scores as the outcome variable in my hierarchical models. For the FCI, CSEM, and CLASS, the outcome is continuous. The FCI and CSEM scores are the proportion of correct responses to the items on the tests. For the CLASS, the score is the proportion of expert-like responses to the survey, which is defined as a response similar to how a physics expert would respond. For these outcomes, I used a hierarchical linear model. Through the survey, I collected expected final grades as either “A”, “B”, “C”, or “D or lower”. This outcome is of the ordinal data type with four levels and so I used a hierarchical cumulative logistic model. For all models, students are grouped by their class thus necessitating the hierarchical model structure.

Research has shown that aspects such as student gender, instructor gender, and pre-score on conceptual tests are significant predictors of post-scores (Feldman, 1993; Hestenes & Halloun, 1995; Potvin et al., 2009). To correctly analyze the effect that student agreement has on student outcomes, I must control for such features of the students and their class. For all analyses, I include each type of demographic information collected through the surveys, including student and instructor studio experience, student and instructor age, the type of course, and the student’s standing (first year in college, second year, etc.).

I selected the four overall agreement scores as measured at the end of the survey as the main predictors. I am interested in the student outcomes measured at the end of the semester to see how well they have learned in the class and if that was influenced by their agreement throughout the semester. Thus, I use the agreement measures from the end of the semester instead of those from the beginning. As a result, classes 11 and 12 are excluded from these analyses since students did not take the PUMBA at the end of the semester. I do not include agreement with specific activities because the overall agreement scores are mathematically defined by student agreement with individual activities. Including them would result in a multicollinearity issue that negatively effects the interpretations of the model estimates. I built the models following the procedure outlined in section 5.5.2.2. Level 1 predictors were introduced to the model first. These include all student level demographics as well as the four student agreement measures. Any predictors that were significant predictors in this model are included in the next model which introduces the level 2 predictors, the instructor demographics and course information. Keeping only predictors that were significant, I created a final model to determine the effect estimates.

Not all classes provided me with conceptual test and/or attitudinal survey scores. Classes 10, 17, and 18 did not distribute a concept test nor the CLASS. Classes 5 and 6 did provide concept test scores but did not distribute the CLASS to their students. Table 35 summarizes the results of each analysis of the effect of student agreement on a student outcome.

6.2.1 Student Agreement and Conceptual Test Scores

Classes that participated in this research distributed a different conceptual test depending on the course. Students took the FCI in mechanics classes and the CSEM in electricity and magnetism classes. Because these are different measures, I ran a separate analysis for each test.

Table 35. A summary of the student outcomes analyses results. Italicized predictor variables indicate a significant agreement measure.

Student Outcome	Significant Predictor	Effect Estimate	p-value	Model Effect Size
Fall 2017				
FCI Post-Score	Student Gender	0.06	0.0214	0.41
	Pre-Score	0.56	<.0001	
CSEM Post-Score	Student Gender	0.10	0.0074	0.18
CLASS Post-Score (Overall Model)	Student Gender	0.08	<.0001	0.6
	Pre-Score	0.75	<.0001	
	In-Class Time Agreement	0.17	0.0021	
CLASS Post-Score (Model with Each Activity)	Student Gender	0.09	<.0001	0.6
	Pre-Score	0.77	<.0001	
	Student Presentation Agreement	-0.01*	0.0135	
Expected Final Grade (Overall Model)	Intercept for "A" Grade	-2.25	.0003	0.72
	Student Gender	0.70	0.0098	
	Student Age	-0.16	0.0027	
	Out-of-Class Time Agreement	1.17	0.0126	
Expected Final Grade (Model with Each Activity)	Intercept for "A" Grade	-2.02	.0003	0.69
	Student Gender	0.56	0.018	
	Student Age	-0.12	0.0113	
	"Extra Reading" Time Agreement	0.85	0.0008	
Spring 2018				
FCI Post-Score	Pre-Score	0.73	<.0001	0.68
CSEM Post-Score	Student Standing	-0.16	0.0115	0.11
CLASS Post-Score	Pre-Score	0.73	<.0001	0.49
Expected Final Grade (Overall Model)	Intercept for "A" Grade	-1.48	.0056	0.51
	Out-of-Class Time Agreement	1.05	0.0223	
Expected Final Grade (Model with Each Activity)	Intercept for "A" Grade	-1.39	.0052	0.54
	"Reading" Time Agreement	0.68	0.0163	

* A negative effect for this predictor mean greater agreement leads to greater post-scores.

6.2.1.1 Force Concept Inventory

Seven classes (classes 2-8) in the Fall semester and four classes (classes 13-16) in the Spring semester were mechanics classes and so took the FCI. As shown in Table 35, in the Fall semester, no overall agreement measure was a significant predictor of FCI score. Only the student's gender and their pre-score predicted their Post-Score. Men received a post-score six percentage points higher than women. A one-unit increase in pre-score corresponded to a 0.56-unit increase in post-score. In the Spring, only pre-score was a significant predictor of post-score on the FCI. A one-unit increase in pre-score corresponded to a 0.73-unit increase in post-score. There is no evidence that a student's agreement to any of the four aspects of the course format effects their performance on the FCI.

6.2.1.2 Conceptual Survey of Electricity and Magnetism

Few classes in my study taught electricity and magnetism. In the Fall semester, only three of the twelve classes taught E&M, and only two of those provided me with concept test scores. In the Spring semester, only one class taught E&M so any results from that class may be conflated with class or instructor level variables. As shown in Table 35, men, in the Fall semester, received a score nearly ten percentage points higher than women. This was the only significant predictor of post-score in the Fall semester. In the Spring semester, students in their third year or more of college received a score 16 percentage points lower than students in their first two years of college. However, remember that these are results from one class where the majority of the class are in their third year or more. Only 11 of the 61 students in that class were in their first two years of college. Again, though, I find that student agreement has no effect on concept test scores.

6.2.2 Student Agreement and Attitudinal Survey Scores

In the Fall semester, I received CLASS scores from seven classes. These were the classes at Universities A and C. In the Spring, I received the CLASS scores from four classes, again from Universities A and C. The results of the analyses on student agreement and attitudinal survey scores are outlined in Table 35. In the Fall semester, I find three significant predictors to a student's post overall favorable score. As a reminder, the overall favorable score is the percentage of items on the survey for which the student responded in a similar manner to an expert. It is indicative of their favorable attitudes towards physics. A one-unit increase in a student's pre-score corresponded to a 0.75-unit increase in their Post-Score. Also, men received an overall favorable score eight percentage points higher than women. I also found that student agreement with the time spent on in-class activities was a significant predictor of their overall favorable post-score. A one-unit increase in agreement with the time spent on in-class activities corresponded to a 0.17-unit increase in their post-score. As might have been expected, a measure of student agreement with the class has a positive relationship with a measure of student attitudes towards physics. This suggests that efforts to increase a student's agreement with the activities that occur in class should also help students have more expert-like attitudes towards physics. However, the contribution of in-class time agreement to explaining the variation in the CLASS post-score (i.e. the effect size of in-class time agreement specifically) is very small (Ferguson, 2009), with an R^2 of 0.02. I determined if agreement with any specific activities were contributing to students attitudinal post-scores by rerunning the final model and replacing the overall in-class time agreement measure with the agreement with the time spent on specific activities. So, I ran a model with pre-scores, student gender, and the nine in-class time activity agreement scores. In this model, student gender and pre-score had similar effect estimates as before. Agreement with only one specific activity, student presentations, was found as a significant predictor of attitudinal post-score, though with a very small effect size of 0.01. As the difference between the time spent on student presentations in their actual class and their most effective class increases (indicating stronger disagreement), students' attitudinal post-scores decrease. However, in the Spring semester, only a student's pre-score predicted their post-score so the results are inconsistent between semesters.

6.2.3 Student Agreement and Expected Final Grade

On the end of the semester PUMBA, students were asked to indicate what grade they expected to receive in their class. So, for this outcome measure, every student that completed a post survey in every class has a data point. The results of the analysis on student agreement and its effect on expected final grade is summarized in Table 35. In the Fall semester, the student's gender, age, and agreement with the time spent on out-of-class activities were significant predictors of the probability for receiving a specific grade. I describe these effects relative to the probability of a student expecting to receive an "A" grade; I calculated the results following the examples shown in Equations 11 and 13 (section 5.5.2.1). The effect estimates that are listed in Table 35 are added to the exponent of each exponential, as demonstrated in Equation 13, to determine the probability of a student expecting to receive an "A" grade. For example, the probability of a woman student, of average age, who completely disagrees with the time spent on out-of-class activities, expecting to receive an "A" in the class is about 10%. The probability of a man student, of average age, who completely disagrees with the time spent on out-of-class activities, expecting to receive an "A" in the class increases to 17%. Adding the effect of agreement with the time spent on out-of-class activities, which has a specific effect size of 0.04, the probability of a woman student, of average age, who completely agrees with the time spent on out-of-class activities, expecting to receive an "A" is 25%. For a man the probability would increase to 41%. The negative sign on the effect estimate for the student's age indicates that older students have a lesser probability of expecting to receive an "A" compared to younger students.

In the Spring semester, agreement with the time spent on out-of-class activities was the only significant predictor of grade expectation. A student who completely disagrees with the time spent on out-of-class activities has a probability of expecting to receive an "A" grade of about 19%. If a student completely agrees with the time spent on out-of-class activities then the probability increases to 39%. The specific effect size for agreement with the time spent on out-of-class activities is very small, less than 0.01.

In both semesters, agreement with the time spent on out-of-class activities had a positive relationship with student's expectation for receiving a higher grade. However, no other agreement measures had an effect. When replacing the overall out-of-class time agreement measure with student agreement with each specific out-of-class activity, I find that doing reading beyond what is assigned in the Fall semester and doing assigned reading in the Spring semester were the only specific activities to significantly predict expected final grade. Both activities had a positive relationship with the probability of expecting to receive a higher grade but both had a specific effect size of 0.01. That is to say that if a student agreed with "extra reading" in the Fall, or assigned reading in the Spring, they were more likely to expect to receive a higher grade. This may be due to students correlating reading outside of class with the concept that students may have to teach themselves outside of class in the studio style. Learning outside of class time is the "flipped classroom" or "upside-down pedagogy" strategy that is a feature of some studio implementations, particularly SCALE-UP where "upside-down pedagogy" is the meaning of the "UP" in the name. If students agree with this feature of the studio style then they are likely doing the learning outside of class, coming prepared to the classroom, and getting the most out of the class time, thus likely receiving a higher grade.

6.2.4 Summary

I hypothesized that increased student agreement reduces student resistance to reformed instructional strategies. With reduced resistance, instructors would have one less barrier to the implementation of such strategies, thus making it easier to use those strategies. As these strategies are based in education research, if they are successfully implemented they should lead to increased student learning. Once I had measured student agreement using the PUMBA, I investigated whether student agreement resulted in increased student outcomes measured by their performance on conceptual tests, their scores on attitudinal surveys, and their expected final grade. I found that student agreement at the end of the semester did not have an effect on performance on conceptual tests at the end of the semester. In one semester, I did find that agreement with the time spent on in-class activities, particularly student presentations, had an effect on students attitudinal scores; if students had greater agreement their attitudinal scores increased. However, this had a very small effect size. This may seem odd considering that on average students perceived student presentations to occur very infrequently (1-3% of the class time). However, students who preferred less time spent on student presentations, perceived it to occur at 5-7% of the class time and they would typically prefer up to 2% of the class time on this activity. Student who would prefer more time on student presentation perceived it to occur for about 1-2% of the class time but would have preferred about 5 percentage points more. Although these disagreements were small, it was evidentially related to their attitudes about physics. More research should look further into this relationship. In both semesters I found that student agreement with the time spent on activities outside of the classroom, particularly reading, affected their expected final grade. These effect sizes were also small or very small. Students who agreed more with the time spent on out-of-class activities were more likely to expect to receive a higher final grade. These data show that increasing student agreement does have some positive results on student outcomes, though they may have a small effect. Getting students to agree with the time spent in-class may increase their attitudes towards the class. Also, getting student agreement with the time spent outside of class, particularly on reading, may increase their expected grade in the class.

6.3 Instructor Strategies to Gain Student Buy-in

In addition to exploring the aspects of the class students agree or disagree with and how that agreement influences student performance in the class, I explored how instructors can influence or gain student buy-in. In this section, I take a descriptive approach to describe the methods and formats instructors use when discussing the class format. First, I explore the activities of the class that instructors typically discuss with their students (Q3a). Then, I look at the methods and formats instructors used most often and if any methods or formats were used more or less often for specific activities (Q3b-c). Throughout these sections, I discuss how the use of these strategies changed from the start to the end of the semester (Q3d). The answers to these questions are found by inspecting Tables 36 through 39, which outline how often instructors used a method or format for a class activity at the beginning and end of the Fall and Spring semesters. In these tables, the value in the cell is the proportion of instructors who used the method or the format (column) out of the number of instructors that discussed that activity (row). I report the proportion of instructors instead of the counts because different total numbers of instructors discussed each activity and I would like to make comparisons across activities. The column with the header “Discussed” tells you how many instructors talked about that activity with their students. To help visualize the data, cells are shaded darker if the value of the cell is larger.

Table 36. Methods and formats used by instructors in the beginning of the Fall semester.

		Methods						Formats					
Beginning of the semester	Discussed (9 max)	Advantages of engagement	Comp. b/w reformed and trad.	Disc. about students	Evidence-based disc.	Extended example	Try and see	Lecture	Class disc./ activity	Reading/ syllabus	A/V	Remind	Reward
Lecture	9	78%	78%	33%	22%	22%	33%	89%	22%	89%	0%	22%	22%
Quiz	7	57%	29%	29%	14%	14%	14%	86%	14%	71%	0%	43%	29%
Labs	8	63%	63%	75%	13%	25%	13%	88%	38%	88%	13%	13%	25%
Small group work	9	89%	89%	67%	11%	22%	22%	89%	56%	67%	22%	33%	11%
Individual work	8	63%	38%	38%	13%	13%	25%	75%	13%	88%	0%	13%	38%
Student present.	1	100%	100%	100%	100%	100%	100%	0%	100%	100%	0%	100%	100%
Demonstrations	2	50%	100%	50%	50%	50%	50%	50%	50%	50%	0%	50%	50%
Whole class disc.	3	67%	67%	33%	33%	67%	33%	33%	100%	33%	33%	67%	33%
Out of class work	9	67%	56%	78%	11%	22%	22%	89%	22%	78%	0%	11%	33%

Table 37. Methods and formats used by instructors by the end of the Fall semester (excluding the first week of class).

		Methods						Formats					
End of the semester	Discussed (6 max)	Advantages of engagement	Comp. b/w reformed and trad.	Disc. about students	Evidence-based disc.	Extended example	Try and see	Lecture	Class disc./ activity	Reading/ syllabus	A/V	Remind	Reward
Lecture	5	100%	40%	40%	20%	0%	0%	100%	20%	40%	0%	40%	0%
Quiz	4	50%	25%	50%	25%	50%	0%	50%	50%	50%	0%	0%	25%
Labs	6	33%	50%	33%	33%	17%	17%	67%	50%	33%	0%	17%	33%
Small group work	5	80%	80%	40%	20%	20%	0%	80%	80%	40%	0%	40%	20%
Individual work	3	67%	33%	33%	0%	33%	0%	100%	33%	100%	0%	67%	33%
Student present.	1	0%	0%	0%	0%	0%	100%	0%	100%	0%	0%	0%	100%
Demonstrations	0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Whole class disc.	1	100%	100%	100%	100%	0%	0%	100%	100%	100%	100 %	100%	100%
Out of class work	5	80%	40%	40%	0%	0%	20%	80%	0%	60%	0%	60%	40%

Table 38. Methods and formats used by instructors in the beginning of the Spring semester.

		Methods						Formats					
Beginning of the semester	Discussed (5 max)	Advantages of engagement	Comp. b/w reformed and trad.	Disc. about students	Evidence-based disc.	Extended example	Try and see	Lecture	Class disc./ activity	Reading/ syllabus	A/V	Remind	Reward
Lecture	5	60%	80%	60%	60%	20%	20%	100%	20%	80%	20%	40%	20%
Quiz	4	75%	50%	25%	25%	0%	50%	100%	50%	75%	0%	25%	50%
Labs	4	100%	50%	100%	25%	50%	25%	100%	50%	75%	25%	50%	25%
Small group work	4	75%	75%	50%	25%	0%	50%	100%	100%	75%	0%	50%	25%
Individual work	2	100%	50%	100%	50%	0%	50%	100%	50%	100%	0%	100%	50%
Student present.	1	100%	0%	100%	0%	0%	0%	100%	0%	0%	0%	0%	0%
Demonstrations	0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Whole class disc.	3	33%	67%	100%	33%	0%	33%	100%	100%	33%	0%	33%	67%
Out of class work	5	60%	40%	40%	20%	0%	40%	100%	20%	100%	0%	40%	20%

Table 39. Methods and formats used by instructors by the end of the Spring semester (excluding the first week of class).

		Methods						Formats					
End of the semester	Discussed (5 max)	Advantages of engagement	Comp. b/w reformed and trad.	Disc. about students	Evidence-based disc.	Extended example	Try and see	Lecture	Class disc./ activity	Reading/ syllabus	A/V	Remind	Reward
Lecture	3	100%	100%	67%	33%	0%	33%	100%	33%	33%	0%	33%	33%
Quiz	3	67%	33%	67%	67%	0%	67%	100%	67%	33%	0%	33%	33%
Labs	3	100%	67%	100%	0%	0%	33%	100%	33%	67%	0%	67%	67%
Small group work	3	100%	67%	33%	0%	0%	33%	100%	100%	67%	0%	67%	33%
Individual work	2	50%	100%	50%	0%	0%	50%	100%	50%	50%	0%	50%	50%
Student present.	0	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Demonstrations	3	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Whole class disc.	3	67%	33%	100%	33%	33%	33%	67%	67%	33%	0%	67%	67%
Out of class work	3	67%	33%	67%	33%	33%	33%	33%	33%	67%	0%	100%	33%

6.3.1 Commonly Discussed Activities

First, addressing research question 3a, I look down the "Discussed" column in Tables 36-39 to determine which activities instructors commonly talk about with their students to set expectations and/or to get students to agree with it. In the Fall semester, at the beginning and throughout the semester, at least seven of the nine instructors talked about lecture, quizzes, labs, small group work, individual work, and out-of-class work. One, two, and three instructors talked about student presentations, demonstrations, or whole class discussions, respectively. Instructors planned to spend less than 10% of the class time on student presentations and demonstrations. . At the start of the Fall semester, only three of the nine instructors indicated that they planned to spend any time on student presentations (of these instructors 8% of their class time would be spent on student presentations on average). Although all but one instructor stated that demonstrations would occur, the average intended time spent in the class on demonstrations was 5.3% of the class. For whole class discussions, five instructors indicated that they planned on incorporating that activity into their class, despite only three instructors discussing it with their students. Of those that did plan on doing whole class discussions, the average proportion of class time they intended to spend on that activity was 14.8%. This, combined with the fact that very little class time was intended to be spent on individual work (1.3% of the class on average) despite all but one instructor discussing it, shows that not planning for an activity in class does not necessarily mean it is not discussed. A possible reason for this is that instructors may discuss with their students that an activity the students might be expecting will not happen in the class.

By the end of the semester, the activities discussed are more reflective of how time was spent in class. Three instructors discussed individual work, and they are the only three to report spending some amount of class time on that activity (12.8% average). One instructor discussed student presentations, and that instructor was one of two who reported doing student presentations. Demonstrations did occur in most instructors' classes, but they did not occur often (3.67% average), perhaps explaining why no instructor talked about demonstrations or why students should agree with them. One instructor discussed whole class discussion throughout the semester and this instructor reported that it was a major part of their class (40%). The other instructors, who did not discuss expectations or agreement with whole class discussion, indicated that they spent an average of 3% of the class time on that activity.

In the beginning of the Spring semester, at least half of the instructors discussed lecture, quizzes, labs, small group work, whole class discussion, and out-of-class work. Two or fewer instructors discussed individual work, student presentations, and demonstrations. Only one instructor indicated that individual work would occur in their class, and that instructor is one of the two that discussed it with their class. Instructors may not have discussed student presentations because, again, it was not intended to occur frequently (1.4% of the class time). Despite every instructor stating that demonstrations would occur in their class, and at an average proportion of class time at 7.2%, no instructors discussed these expectations with their students.

Each of the five instructors in the Spring semester were also involved in the Fall semester, so I can track which activities they did or did not discuss between the two semesters (see Table 40). One more instructor (instructor 1) discussed quizzes/tests in the Spring semester than in the Fall semester. Instructor 2 began discussing student presentations and whole class discussion in the Spring semester when he did not in the Fall. More instructors stopped discussing activities than started, particularly for individual work and demonstrations.

Table 40. The change in discussed activities across semesters.

Activity	Number of instructors (out of 5) who started/stopped discussing an activity at the beginning of a semester between the Fall and Spring semesters.		Number of instructors (out of 4) who started/stopped discussing an activity at the end of a semester between the Fall and Spring semesters.	
	Started	Stopped	Started	Stopped
Lecture	0	0	0	1
Quiz	1	0	1	1
Labs	0	0	0	2
Small Group Work	0	1	0	1
Individual Work	0	2	1	1
Student Presentations	1	1	0	1
Demonstrations	0	2	2	0
Whole Class Discussion	1	0	2	0
Out-of-Class Work	0	0	0	1

By the end of the Spring semester, the same activities most instructors discussed at the start continued to be discussed throughout the semester. Three instructors throughout the semester later discussed demonstrations, which they did not discuss at all at the beginning of the semester. Individual work and student presentations remain as activities instructors do not often talk about with their students. One instructor indicated that individual work occurred for 30% of their class, and that instructor did talk about expectations for or agreement with individual work with their students throughout the semester. No instructor discussed student presentations, likely because it did not occur in class. Only one instructor said that it did but only for 5% of the class time.

Between the two semesters, the activities discussed by the four instructors (one instructor did not take the SIMBA at the end of the Fall semester so there are only four to compare) who were present in both semesters varied quite a bit. There were six instances in which an instructor talked about an activity in the Spring semester that they did not discuss in the Fall and there were eight instances where an instructor didn't discuss an activity in the Spring but did in the Fall (see Table 40).

In general, instructors often discuss expectations for and/or why students should agree with lecture, quiz/tests, labs, small group work, and out-of-class work, at the beginning and throughout the semester. Individual work, student presentations, demonstrations, and whole class discussion are discussed less often. In most cases, these activities are not discussed because they are not a part of the class. Demonstrations is an exception to this because though it was not discussed by many instructors, it did occur in the class (though not for much time).

Implications for Instructors My findings from this analysis can inform areas in which an instructor can improve in gaining student buy-in. For instance, few instructors discussed student presentations and demonstrations with their students when describing the class format. Previously, I found that the importance of student presentations and the importance of and time spent on demonstrations were often one of the least agreed with activities among students (see Tables 32 and 33). Students typically want less student presentations and more demonstrations than they perceived. If instructors begin to discuss these activities, they might be able to 1) clarify to students how much student presentations and demonstrations there will be (this is particularly needed for demonstrations at the start of a semester; Figure 12o), and 2) encourage agreement with these activities. By discussing these activities with their students, instructors should be able to increase agreement with two of the more commonly disagreed with activities.

6.3.2 Commonly Used Methods

By looking down the columns of Tables 36-39, I can explore which methods these instructors commonly used to explain a class activity or to get student agreement with an activity. As a reminder, “methods” are the topics of conversation an instructor might use when describing the class format for a certain activity. For commonly used methods, I am searching for methods where most cells have large percentages, or are shaded darker. At the beginning and end of both semesters, the instructors in my study often used the “advantages of engagement”, “comparison between reformed and traditional courses”, and the “discussion about students” methods. In the Fall semester, the use of these methods decreased by the end of the semester, but they were still the most prevalent methods used throughout the semester. In the Spring semester, the proportion of instructors using those three methods out of the instructors that discussed an activity remained relatively similar. Also in the Spring semester, there appears to be greater use of the “try and see” method both at the start of the semester and throughout, however this is more due to instructors who didn’t use the “try and see method” in the Fall semester not participating in the Spring semester. Only one instructor who didn’t use the “try and see” method in the Fall used it in the Spring.

“Evidence-based discussion” and “extended example” methods are less commonly used in each semester. No more than two instructors ever indicated that they used the “extended example” method for an activity. The same can be said for “evidence-based discussion” aside from three instructors using that method regarding lecture at the start of the Spring semester. It is possible that few instructors used the “evidence-based discussion” method because few knew of evidence they could use to get students to agree with the studio format. For instance, when asked if she didn’t use the “evidence-based discussion” method because she didn’t have that information, instructor 1 said,

> *“Yeah, I mean, I'm generally familiar with the fact that a lot of this research has shown things but I haven't had personally, recently especially, the time to like go through and say like alright this is this, uh, research that shows this one thing. Right? And I wish I did and I would like to improve on future semesters. It would ... But yeah, if I had more hard evidence about, especially about, specific things, I think it would, I would be happy to use that. Yeah.”*

This is one possible explanation for the lack of “evidence-based research” being used among instructors. It is unknown why certain other methods were used more commonly over others. This would be an area to pursue when continuing this research.

Inspection of Tables 36-39 can provide insight into which methods are more or less commonly used for certain activities among this group of instructors. This type of question could be aided by a chi-square goodness of fit inferential test which would tell me if more or less instructors use a method for certain activities than was expected. However, the data violates multiple assumptions of the goodness of fit test: 1) the sample size is small, so the expected count in multiple cells is less than 5 and sometimes less than 1, and 2) the same instructor may be represented in multiple cells, violating the assumption of independence. Because of these violations, it is inappropriate to run inferential statistics on the data. However, I can examine the data and find trends among the sampled population that could guide future research.

At the beginning of the Fall semester, many instructors talked about the advantages of being engaged in reference to any class activity. Slightly more often than the other activities, instructors used this method for lecture and small group work. Making a comparison between traditional and reformed classes also occurred slightly more often for lecture and small group work. This may be due to the fact that these are two activities where traditional and reformed courses differ the most. Reformed studio classes typically exchange lecture for more group work. When explaining the studio format, it is natural to compare these activities to the traditional course as they are the main difference. Instructors might justify this difference by highlighting the studio goal of student engagement. Thus, in one conversation, instructors may have discussed lecture and small group work by comparing it to a traditional class and reasoning that it keeps students engaged, which is a good thing. Indeed, one instructor made this connection between the “comparison” and “advantages” methods, saying “Most of [the “comparison” method] actually comes in telling them what the advantages of the engagement are, for instance.”

I also find that the "discussion about students" method is commonly used with labs, small group work, and out-of-class work. These three activities are largely student-centered activities; these are the three activities where students are the most engaged and interactive. Thus, it should be no surprise that the instructors talk about the students when getting their agreement for these activities more often than they do for other activities. This same reasoning might explain why the “class discussion/activity” format is used more commonly for small group work. The remaining methods, “evidence-based discussion”, “extended example”, and “try and see” are used uniformly across the activities at the beginning of the Fall semester.

By the end of the Fall semester, the “advantages of engagement” method was still being used more commonly for lecture and small group work activities. It was also used commonly with out-of-class work. As the semester continued, perhaps instructors needed to motivate students to do homework or come prepared by discussing the benefits of engaging in such activities. Making a comparison between reformed and traditional classes was used commonly for small group work, similar to the beginning of the semester, but unlike the beginning, it was not used more commonly for lecture. Perhaps as the semester goes on, there is less of a need to describe the differences in lecture between the two class styles because students will have experienced it by then. At the beginning of the semester the “discussion about students” method was used more commonly for student-centered activities, but throughout the semester it was used generally equally among the activities.

At the beginning of the Spring semester, all four instructors who discussed labs used the “advantages of engagement” method. Though, due to the low number of instructors discussing any activity, there may not be much meaning to this fact; this is only one more instructor using this method than the number of instructors using this method for “lecture”, “quiz”, and “small group work” activities. It does seem that the “advantages of engagement” method was used less commonly for whole class discussion, where only one of three possible instructors used that method. The “comparison” method is still used more commonly for lecture and small group work, perhaps for the same reason described earlier. A discussion about students is used less commonly for quizzes and more commonly for labs and whole class discussions. This seemingly refutes my earlier theory for the use of the “discussion about students” method because it is not used more commonly with the “small group work” or “out-of-class work” activities. However, due to the low sample sizes, these findings should be interpreted with some caution. Lastly, the “evidence-based discussion” method seems to be used more often with the “lecture” activity. By the end of the semester, the use of these methods appears to be more spread out for most activities. Considering that at most, three instructors discussed one activity over the course of the semester, it would not be wise to make comparisons on the use of certain methods between different activities.

In both semesters, three methods are by far the most commonly used: “advantages of engagement”, “comparison between reformed and traditional classes”, and “discussion about students”. These activities do seem to be used more for certain activities than for other, particularly in the Fall semester. The “advantages” and “comparison” methods may be used together when describing how the studio class has reduced lecture and increased group work. The “discussion about students” method was used more often with student-centered activities. The other three methods, “evidence-based discussion”, “extended example”, and “try and see” were used less commonly (though “try and see” did see some use in the Spring).

Implications for Instructors. Information like this can help instructors learn about new ways of getting student agreement. Perhaps they had not thought to use those final three methods. It may be that an instructor hadn’t thought to use a discussion about students when talking about student-centered activities. If an instructor feels that their methods are not working, this information might encourage the instructor to try and use a less frequently employed method for a certain activity to see if it works. These results also suggest that instructors may not know of evidence that could be used when describing the class, since that method was used infrequently.

6.3.3 Commonly Used Formats

Like I did with the methods shown in Tables 36-39, I explore commonly used formats by inspecting the formats columns which have higher proportions of instructors using them and are shaded darker. As a reminder, “formats” are the physical means of having a conversation about class expectations and/or why students should agree with them. At the beginning and end of both semesters, the “lecture” conversation format, or simply telling the students something without their input, was used very commonly. This is not surprising given how quick and easy it is for an instructor to simply explain an aspect of the class format to their students. At the beginning of both semesters, the “reading/syllabus” format was also used frequently. When asked to specify what text they provided to their students that they considered as the “reading/syllabus” format, all instructors stated they were referring to placing information in the syllabus. Thus, it is not surprising that this format was used more often in the beginning of a semester than it was at the end, nor is it surprising that it was used frequently because the first day of class usually involves reviewing the syllabus. The “class discussion/activity”, “reminders throughout the semester”, and “rewarding appropriate behavior” formats were used but to a lesser extent than “lecture” and “reading/syllabus”. The use of these three activities remained relatively consistent throughout a semester. The only format to not be used by many instructors was the “audio/video” format. Across each distribution of the instructor survey, only one or two instructors used “audio/video” for only one or two activities. For most activities, this format was not used at all.

Also through inspection of Tables 36-39, I can search for formats that may be used more or less commonly for certain activities. However, due to the same assumption violations described in the previous section, I cannot make statistical determinations. I can examine the data to search for possible trends that could guide future research. These trends may only be true for this population of instructors and so should not be generalized to all studio instructors. That would be an avenue for continued research on instructor buy-in strategies.

The “lecture” format was used fairly consistently across all activities (excluding those activities that very few instructors discuss). Throughout the Fall semester, the “lecture” format was used slightly less often regarding expectations or agreement with quizzes/tests. Throughout the Spring semester it was used less commonly for out-of-class work. There does appear to be a consistent use of the "class discussion/activity" format regarding expectations or agreement with small group work as evidenced in each distribution of the SIMBA. This may be because both small group work and a class discussion or activity are both student-centered. So, instructors may be using a student-centered format to describe the student-centered class activity. The remaining activities are used, or not used, relatively uniformly across each activity at both the beginning and end of a semester.

6.3.4 Summary

Instructors often discussed lectures, quizzes/tests, labs, small group work, and out-of-class work with their students when setting their expectations and trying to get their agreement with the class format. Fewer instructors discussed individual work (aside from the beginning of the Fall semester), student presentations, demonstrations, and whole class discussions. Combined with the investigation into the specific activities students disagree with, it is clear that more instructors should discuss student presentations and demonstrations as these two activities are often one of the least agreed with activities.

The most common methods for getting agreement were to discuss the advantages of being engaged, make a comparison between reformed and traditional courses, and have a discussion about the students. The first two of these seem to be used more often when describing lecture and small group work, perhaps being used in tandem. Having a discussion about students seems to be used more commonly with student-centered activities. It is important to note that conclusions about specific methods being used more or less frequently for specific activities should not be generalized to anyone outside of these instructors. The remaining methods, "evidence-based discussion", "extended example", and "try and see" were used less often. Instructors may not have used an “evidence-based discussion” because they did not have such evidence readily available to them. To help an instructor achieve buy-in for their reformed course a department may want to provide such materials for their instructors.

Regarding the formats instructors use to discuss expectations and why students should agree with the class format, “lecture” and “reading/syllabus” were the two most common formats. Instructors typically referred to the syllabus when indicating that they use the “Reading/syllabus” format. “Lecture” remained a common format throughout the semester whereas “readying/syllabus” was used less often throughout the semester. “Class discussion/activity”, “reminders throughout the semester”, and “rewarding appropriate behavior” were used for many activities but to a lesser extent than “lecture” and “reading/syllabus”. The “Audio/video” format was rarely used.

This information simply describes what instructors might do to try and achieve buy-in. With this data I can identify possible strategies these instructors may have used, strategies which may be further explored in future research. For example, it appears that making a comparison between reformed and traditional courses may be used in combination with discussing advantages of engagement to describe lecture and small group work in the studio class. Instructors may also be using student-centered methods and formats, like “discussion about students” and “class discussion/activity” to describe student-centered activities like the “small group work” activity. The natural follow-up question to this information is which methods and formats are effective at changing student agreement?

6.4. The Effect of Instructor Strategies on Student Agreement

Up to this point, I have described the aspects of the studio class format with which students agree or disagree, the specific activities they agree or disagree with, what that means for how they perform in the course or their attitudes towards physics, and the ways instructors in this study have tried to explain the course format and why students should agree with it. The final piece in my investigation into student agreement is to determine which practices or strategies are effective at gaining student agreement so that I may make practical recommendations to instructors teaching studio classes. In doing so, I address my final research question, is the use of certain methods or formats to get student agreement associated with higher levels of student agreement (Q4a)?

I first attempt to answer this question through a quantitative analysis, combining the results of both the PUMBA and the SIMBA to see if the methods and formats instructors used in their class actually resulted in a change in student agreement. In the next chapter, I explore this question qualitatively through instructor and student interviews to determine what they feel are effective and ineffective strategies to gain student agreement.

6.4.1 The Analysis Procedure

As I have framed it in the instructor survey, the SIMBA, the methods and formats instructors might use to gain student agreement are always used regarding a specific activity. Thus, when testing to see if these methods and formats are effective, I must use student agreement with that specific activity as the outcome variable. The four overall student agreement measures will not be used as part of these analyses. There are nine activities that an instructor might refer to when discussing the class format: lecture, quiz/tests, labs, small group work, individual work, student presentations, demonstrations, whole class discussions, and out-of-class work. Many of these activities overlap with a single activity the students might agree to as measured in the PUMBA. For the students, “labs” are separated into “step-by-step labs” and “investigative labs”. The strategies that instructors use to discuss labs will be applied to both “step-by-step labs” and “investigative labs”. The strategies that instructors use to discuss out-of-class work will be applied to each of the out-of-class activities (see Table 10). The reasoning for these differences between the activities on the SIMBA and PUMBA are described in section 4.2.2.

There are nine in-class activities and eight out-of-class activities that students can agree with. Additionally, their agreement is based on both the importance placed on these activities and the time spent on these activities. Thus, for each activity there are two agreement measures. In total, there are 34 different agreement measures for each distribution of the PUMBA.

For the importance agreement measures and the out-of-class time agreement measures, students either agree with the activity or prefer more or less. Thus, this agreement measure is dichotomous: they agree or they do not. As this variable is a categorical outcome variable, hierarchical logistic regressions were used to estimate the effect of instructors’ strategies on the outcome. Agreement with the time spent on in-class activities could be measured as the mathematical difference between the proportion of class time a student would prefer to be spent on an activity versus the time it actually occurred (referred to as “difference”) or as the magnitude of this difference (referred to as “agreement”). Both types of measurements are on a continuous scale, so a hierarchical linear regression was used. The hierarchical models and how they were specified in the analysis are described in section 5.5.2.

I tested these agreement measures against the strategies that instructors used to try to gain agreement. I included demographic information about the student, instructor, and course in the model to account for differences in agreement based on factors outside of the instructor strategies. I also included two other variables that I refer to as “perceived discussion” and “perceived agreement discussion”. These variables come from the first section of the PUMBA, where the students indicate if they perceived that an instructor used any strategies to talk about an activity (“perceived discussion”) and talked about why they should agree with an activity (“perceived agreement discussion”).

When running these analyses, I came across an issue of multicollinearity, which is when the contribution of one predictor could be explained by the contribution of other predictors. For a simple example (and this is not necessarily the case for my data), it could have been that all instructors who taught an electricity and magnetism class were women. If these two predictors remained in the model, I would not be able to trust the effect estimates as generalizable. Multicollinearity increases the standard errors of effect estimates meaning that the magnitude of the effect is more variable between samples. The effect estimate obtained from my sample may not be found in a different sample. I also would not be able to determine which of the two predictors is actually responsible for the outcome (Field, 2009). Does a student agree better because they were in an E&M class or because their instructor was a woman? As my goal with these analyses is to determine the effect of specific methods or formats, I must resolve the multicollinearity issue.

I could conduct a principal component analysis or a factor analysis to combine predictors that explain similar portions of the data (Field, 2009). Doing so would reduce the number of predictors I use by clustering similar predictors into one. However, this prevents me from achieving my goal of pinpointing specific strategies to use to gain student agreement. Instead, I inspected the relationships between the predictors using the variance inflation factor (VIF) and tolerance statistics, which are the inverse of each other. These statistics indicate if a predictor has a linear relationship with other predictors. If the VIF is greater than 10, there is an issue of multicollinearity in my data (Bowerman & O'connell, 1990; Myers, 1990). I checked the VIF of each predictor used together in a model by running them through a linear regression in SPSS and having the program output multicollinearity diagnostics. The outcome variable of this model is not important as VIF or tolerance is only dependent on relationships between the predictors. SPSS automatically eliminates from the model predictors which have a tolerance below 0.001, which are the predictors that are almost completely defined by a combination of other predictors. The predictors that remained usually had a low VIF with the others. Sometimes, however, there were still predictors with a VIF above the 10 threshold. In these cases, I removed the predictors with the largest VIFs one at a time, rerunning the model each time, until all VIFs were below 10. Because of this, I may not be able to say definitively if one method or format leads to better student agreement. If one of these methods or formats that remains in the model is found to be significant, I must first check to see if a combination of other predictors might explain that predictor. I did this by running a regression where the significant predictor was now the outcome and all other variables remain a predictor. If I find that other predictors explained the significant predictor, I can only say that either that method or format affects student agreement or the combination of some other set of predictors might affect student agreement. I assume that any observable effect (i.e. a statistically significant effect) is the result of only one variable or the combination of two variables. It is possible that the effect found from predictor A is actually the effect of the combination of predictors B, C, and D, which predictor A may be significantly collinear with, and not from predictor A itself. However, I am assuming that if an effect appears to come from one or a combination of two predictors, it is more likely from those rather than a combination of more predictors working together. For example, I could find that using a student focused discussion significantly increases student agreement, but that method may be collinear with the "reading/syllabus" format, the "try and see" method, and the "comparison between reformed and traditional classes" method. I am assuming that the student's agreement is responding to the one, more salient, method, rather than the less salient combination of numerous formats and methods. So, I will report if a predictor, found to be significant in the analysis, is significantly collinear with only one or two other predictors. If a significant predictor is significantly collinear with three or more other predictors, I will assume the effect is actually from the one predictor.

6.4.2 Agreement with In-Class Activity Importance

The effect of instructor methods and formats on students' agreement with the importance of in-class activities was measured using a hierarchical logistic regression. Table 41 summarizes the in-class activity importance agreement in which the analysis found a statistically significant effect from an instructor method or format. Recall that the effect estimates in a hierarchical logistic regression must be interpreted using the example shown in Equation 14. The estimates are used to calculate the probability that a student would agree with the importance of an activity. One may still use the estimates in their raw form to determine the relationship the predictor has with the outcome and the relative magnitude of the effect.

Table 41. The instructor strategies that significantly affect student agreement with the importance placed on in-class activities.

Survey Distribution	Outcome	Significant Predictors	Effect Estimate	p-value	Effect Size	Possible related predictors	Adjusted R^2 of predictor regression
Beginning Fall 2017	Step-by-step labs importance	Lecture Format	0.756	0.0066	0.02	Labs Discussion, Reminders Format	1
End Spring 2018	Lecture Importance	Perceived Discussion	1.801	0.003	0.07	Perceived Agreement Discussion	0.155
End Spring 2018	Demonstrations Importance	Perceived Discussion	-1.040	0.0231	0.04	Student Age, Perceived Agreement Discussion	0.482
		Perceived Agreement Discussion	0.950	0.0164		Student Age, Perceived Discussion	0.475
End Spring 2018	Whole Class Discussion Importance	Perceived Discussion	1.019	0.0002	0.07		

In the beginning of the Fall semester, I only found evidence for one method or format influencing agreement with importance placed on an in-class activity. Either using the "lecture" format for discussing step-by-step labs expectations and agreement or a combination of having a discussion about labs in general and reminding students throughout the semester had a positive relationship with the probability that a student would agree with the importance of step-by-step labs. This is a very small effect, with an effect size of 0.02. As evidenced by the adjusted R^2 value of 1, the "lecture" format predictor was completely confounded with the general labs discussion and the "reminders throughout the semester" format, meaning the observed effect could be due to either sets of predictors. Only two instructors did not use the "lecture" format, so it might be that this effect is actually dependent on some aspect common to these two classes that is not included in the analysis.

At the end of the Fall semester and beginning of the Spring semester, I found no evidence that an instructor strategy could affect student agreement with the importance placed on in-class activities. However, throughout the Spring semester, there is evidence that if the student perceives there were discussions about an activity and/or discussions about why they should agree with an activity, their agreement with the importance placed on that activity might change. These are small effects, with effect sizes ranging from 0.04-0.07. Students were more likely to agree with the importance placed on lecture throughout the Spring semester if they perceived that their instructor discussed the lecture format with them. This predictor was found to be collinear with other predictors (but it was only significantly collinear with "perceived agreement discussion"); however, these other predictors only explained 15.5% of the "perceived discussion" predictor. Thus, it is likely that this effect is coming from the "perceived discussion" predictor. Students were also more likely to agree with the importance placed on whole class discussions if they perceived that their instructor discussed that aspect of the class with them throughout the semester. This predictor was significantly collinear with other predictors, but the collinearity was with more than three other predictors. Thus, based on my assumption, I am claiming that the effect found is most likely only from the "perceived discussion" predictor. The "perceived discussion" and "perceived agreement discussion" predictors both had an effect on student agreement with the importance placed on demonstrations. Both predictors were significantly collinear with each other as was the student's age. Approximately 48% of the variability in both of these predictors can be explained through other predictors, as evidenced by the adjusted R^2 values, meaning that their effect might actually be from other predictors. However, the effects found originally make theoretical sense, specifically that students who perceive a discussion about why they should agree with the importance of demonstrations are indeed more likely to agree. A possible reason that students who perceived a discussion about demonstrations (but not an agreement discussion) might be less likely to agree with demonstrations is that the discussion might reveal that demonstrations may not be as important as they would prefer. Had the discussion not taken place, students would remain ignorant about the role of demonstrations in the class and perhaps assume they are as important as they would hope.

6.4.3 Agreement with Out-of-Class Importance

I analyzed and interpreted student agreement with this aspect of the class in an identical manner as agreement with the importance placed on in-class activities. The instructor methods and formats that I found to have a statistically significant effect on student agreement are outlined in Table 42.

Table 42. The instructor strategies that significantly affect student agreement with the importance placed on out-of-class activities.

Survey Distribution	Outcome	Significant Predictors	Effect Estimate	p-value	Effect Size	Possible related predictors	Adjusted R^2 of predictor regression
Beginning Fall 2017	Lab Reports Importance	Rewarding Appropriate Behavior	-0.966	0.008	0.07		
End Fall 2017	Extra Homework Importance	Perceived Agreement Discussion	0.700	0.0101	0.04	Perceived Discussion	0.376
End Fall 2017	Lab Reports Importance	Perceived Discussion	-0.886	0.0494	0.05	Perceived Agreement Discussion, Lecture Format	0.391
End Fall 2017	Group Projects Importance	Comparison between reformed and traditional classes	1.246	0.0055	0.10	Algebra/Calculus-Based Class, Lecture Format	1
Beginning Spring 2018	Homework Importance	Class/Discussion Activity	-1.158	0.0068	0.05	Rewarding Appropriate Behavior	1
End Spring 2018	Extra Reading Importance	Reminders throughout the semesters	0.815	0.0003	0.05	Out-of-Class Activity Discussion	1
End Spring 2018	Lab Reports Importance	Perceived Discussion	-1.193	0.0241	0.09	Perceived Agreement Discussion	0.352
		Perceived Agreement Discussion	1.498	0.0023		Perceived Discussion	0.363

At the beginning of either semester, agreement with the importance placed on only one out-of-class activity was affected by instructor strategies to get student agreement. At the start of the Fall semester, students were less likely to agree with the importance placed on lab reports if the instructor discussed rewarding appropriate behavior for lab reports. This was a small effect with an effect size of 0.07. Only two instructors used this format when discussing labs so the effect found may be due to a common aspect between these two classes that was not included in the model. The “rewarding appropriate behavior” format may have also had a negative effect on the agreement with the importance placed on the “homework” activity at the beginning of the Spring semester. However, this effect might have come from the “class discussion/activity” format as it was perfectly collinear with the “rewarding” format. Only one instructor used these formats at the start of the Spring semester so it is possible that the found effect is actually due to an aspect of that class that is not included in the model. Additionally, this was a small effect with an effect size of 0.05.

Assuming that the effects are due to the instructor strategies, it may be that at the beginning of a semester, these are two out-of-class activities that students would place little importance on. By using the “rewarding” format or the “class discussion/activity” format, the instructor is bringing attention to the actual importance of these activities and so students cannot remain ignorant. Of course, it is important for students to not remain ignorant about these activities. What this information tells me, if this is indeed the reasoning behind the results, is that instructors need to put an extra effort in to get agreement with these activities since once you describe your expectations for those activities, students may naturally disagree.

Like agreement with the importance placed on in-class activities, agreement with the importance placed on out-of-class activities by the end of the semester is sometimes influenced by students’ perceptions of whether or not a discussion about that activity and/or a discussion about why they should agree with the activity occurred. These effects were somewhat small, ranging from 0.04-0.09 effect sizes. At the end of the Fall semester, students were more likely to agree with the importance placed on “extra homework” if they perceived their instructor discussed why they should agree with it. It is possible that this could be the effect of students’ perceptions of if there was a discussion about “extra homework” in general. However, only 38% of the “perceived agreement discussion” predictor is explained at all by other predictors, of which only “perceived discussion” is significant, so it is likely that the effect is due to the agreement discussion. Likewise, at the end of the Spring semester, the probability of agreeing with the importance of lab reports increases if the students perceived an agreement discussion throughout the semester. This could perhaps be the effect of the perception of a general agreement discussion with out-of-class work but less likely due to only 37% of the “perceived agreement discussion” predictor being explained by other predictors. A discussion about out-of-class work expectations (which every instructor did at the beginning and end of a semester, except for instructors 7 and 8 at the end of the Spring semester) tends to reduce the probability that a student will agree with the importance of lab reports, as found in both semesters. This predictor is significantly collinear with at least one other predictor but in both semesters, the other predictors explain a somewhat small amount of the “perceived discussion” predictor thus, the results are likely from the “perceived discussion” predictor.

A comparison between reformed and traditional classes, a method used by two instructors, might have a positive effect, with effect size 0.10, on the likelihood that a student would agree with the importance of group projects. Though, this predictor is completely collinear with a combination of the instructor using the “lecture” format and the student being in an algebra or calculus-based class. Using the “reminders throughout the semester” format or talking about out-of-class activities in general (3 instructors used these strategies) are strategies an instructor might use that could improve agreement with the importance placed on reading beyond what is assigned, as evidenced by the results at the end of the Spring semester.

6.4.4 Agreement with In-Class Time

Unlike all other agreement measures, the outcome variable for a student's agreement with the time spent on in-class activities is continuous, so I used a hierarchical linear regression to explore the relationship between instructor strategies and student agreement. I explore the effect of these strategies on the difference in the proportion of class time spent on an activity between a student's actual class and their most effective class. I refer to this difference as an activity's "difference" score. I also examine how instructor strategies affect the magnitude of this difference (called "agreement" score) to see if the difference gets smaller. I differentiate between the "difference" score and the "agreement" score because a significant effect found in the "difference" score means that overall all students, regardless of if they preferred more or less of an activity, shifted their agreement in the same direction. From example, the effect of a method might be to change a difference of +10 to +5 and a difference of -5 to -10. The shift in the difference is the same regardless of whether the student wanted more or less of that activity. On the other hand, if there is a significant effect from a method or format when exploring "agreement" scores, the differences would both move towards or away from zero. For example, if a method had a negative effect (meaning it reduces the "agreement" score) it would change a difference of +10 to +5 and -10 to -5. Both scores moved closer to zero. Table 43 outlines the activities and score type with which a method or format had a significant effect. Remember that for the individual activity "agreement" score, a negative effect indicates stronger agreement because it means a smaller magnitude of the difference between what students prefer in their most effective class and what they perceive in their actual class. For the “difference” scores, the direction of the effect in conjunction with the intercept determines if the effect represents increasing agreement in general (see Figure 9), thus, for these scores, I also report the intercept.

Table 43. The instructor strategies that significantly affect student agreement with the time spent on in-class activities.

Survey Distribution	Outcome	Significant Predictors	Effect Estimate	p-value	Effect Size	Possible related predictors	Adjusted R^2 of predictor regression
Beginning Fall 2017	Lecture Agreement	Perceived Agreement Discussion	-5.1455	0.001	0.05	Perceived Discussion	0.148
		Extended Example	5.3112	0.005			
Beginning Fall 2017	Step-by-step Labs Agreement	Perceived Agreement Discussion	-3.7828	0.014	0.02	Perceived Discussion	0.534
		Student Standing	1.7398	0.012			
Beginning Fall 2017	Demonstrations Difference	*Intercept*	2.7273	<.001	0.02		
		Perceived Discussion	-2.0089	0.006		Perceived Agreement Discussion	0.587
End Fall 2017	Individual Work Difference	*Intercept*	2.7488	0.002	0.01		
		Perceived Agreement Discussion	-2.1608	0.035		Student Gender, Perceived Discussion	0.447
Beginning Spring 2018	Lecture Agreement	Perceived Agreement Discussion	-4.3828	0.017	0.02	Perceived Discussion	0.248
End Spring 2018	Lecture Difference	*Intercept*	10.7953	0.0345	0.06		
		Perceived Discussion	-16.1191	<.001		Perceived Agreement Discussion	0.155
		Perceived Agreement Discussion	4.638	0.043		Perceived Discussion	0.184
End Spring 2018	Lecture Agreement	Perceived Discussion	-13.8105	<.001	0.13	Perceived Agreement Discussion	0.155
		Student Gender	-3.5303	0.015			0.134
End Spring 2018	Quiz Agreement	Rewarding Appropriate Behavior	2.4956	0.005	0.05	Reminders throughout the semester	1
End Spring 2018	Investigative Labs Difference	*Intercept*	0.1064	0.5756	0.02		
		Class Discussion/ Activity	3.8665	0.028		"Try and see"	1
		Rewarding Appropriate Behavior	-2.6419	0.035		Instructor Gender, Lecture Format	1
End Spring 2018	Demonstrations Agreement	Perceived Agreement Discussion	-1.791	0.015	0.05	Student Age, Perceived Discussion	0.475

The majority of the effective strategies for gaining student agreement are the general "perceived discussion" and "perceived agreement discussion" predictors. Both of these predictors were always significantly collinear with the other, but usually a small proportion of one predictor was explained by all other predictors (see the adjusted R^2 values). There are a few exceptions, namely for step-by-step labs agreement, demonstrations difference, and individual work difference, where the effect of one predictor could be the effect of another due to multicollinearity. In every case, the "perceived discussion" or "perceived agreement discussion" predictors worked to increase agreement, implying that if an instructor discusses the class format and why their students should agree with it, they may achieve greater agreement with the activities listed in Table 43.

There were only a few specific methods or formats that had a significant effect on student agreement, all of which adversely affected agreement. At the beginning of the Fall semester, using an extended example effectively cancelled out any benefit received from a general discussion about why students should agree with lecture. However, only two instructors used that method at the beginning of the semester, so it may be that those instructors had worse agreement with the time spent on lecture for other reasons not included in the model. At the end of the Spring semester, using the "reminders throughout the semester" or "rewarding appropriate behavior" formats (they were completely collinear) seemed to reduce student agreement with the time spent on quizzes. However, only one instructor used these methods so this effect may be due other qualities of that one instructor or class. This instructor was also one of the two instructors to have used the "extended example" method at the start of the Fall semester. Lastly, "rewarding appropriate behavior" or a combination of the instructor gender and using the "lecture" format, had a negative effect on agreement with investigative labs. Again, the one instructor who used the other significant methods used these formats. So, it may be that these effects are significant because of one instructor who uses many strategies to gain student agreement but to little success. It is also worth noting that the three instructors that used the "lecture" format throughout the semester regarding labs all taught at University A, so it may be that there is some university effect on student agreement with investigative labs.

I determined effect sizes from the final model, meaning that if the final model had more than one predictor, the effect size represents the effect of all of those predictors together. So, in some cases, like step-by-step labs agreement and lecture agreement at the end of the Spring semester, the effect size is calculated based on the effect of multiple instructor strategies or a strategy with a student demographic predictor. However, the effect sizes were small for all final models, implying that the effects of the specific instructor strategies must also be small.

6.4.5 Agreement with Out-of-Class Time

Like the two importance measures, agreement with the time spent on out-of-class activities is measured dichotomously, necessitating a hierarchical logistic model. Table 44 summarizes the out-of-class activities with which an instructor strategy had a significant effect on student agreement. In the beginning of the Fall semester, students who perceived a discussion about expectations for out-of-class activities were more likely to agree with the time spent getting help from the instructor, GTAs, or tutors. Though, other predictors explain about 44% of the "perceived discussion predictor" including the "perceived agreement discussion" predictor, which is significantly collinear, so the effect might be from a perceived discussion about why students should agree with the out-of-class activities. Throughout the Fall semester, students in classes where the instructor used the "rewarding appropriate behavior" format for out-of-class activities were less likely to agree with the time spent doing assigned conceptual/problem solving work ("homework"). Two of the six instructors who took the SIMBA at the end of the Fall semesters indicated that they used this format. Agreement with the time spent doing conceptual/problem solving work beyond what is assigned ("extra homework") was positively affected by instructors discussing the advantages of engagement for out-of-class work. Four instructors used this method. All of the effects in the Fall semester were fairly small, with effect sizes of 0.04 or 0.06.

Table 44. The instructor strategies that significantly affect student agreement with the time spent on in-class activities.

Survey Distribution	Outcome	Significant Predictors	Effect Estimate	p-value	Effect Size	Possible related predictors	Adjusted R^2 of predictor regression
Beginning Fall 2017	Help Time	Perceived Agreement Discussion	0.64451	0.014	0.06	Perceived Agreement Discussion	0.436
		Student Standing	-0.5215	0.012			
		Student Studio Experience	-0.5671	0.032			
End Fall 2017	Homework Time	Rewarding Appropriate Behavior	-0.9583	0.039	0.04		
End Fall 2017	Extra Homework Time	Advantages of Engagement	0.841	0.007	0.04		
Beginning Spring 2018	Extra Reading Time	Student Studio Experience	-1.4236	0.005	0.13	Advantages of Engagement	0.066
		Reminders throughout the semesters	-0.7266	0.046			
		Rewarding Appropriate Behavior	0.9285	0.031			
End Spring 2018	Homework Time	Perceived Agreement Discussion	0.5973	0.029	.04	Perceived Discussion	0.363
		Rewarding Appropriate Behavior	-0.8303	0.046		Class Discussion/Activity OR Evidence-based Discussion	1
End Spring 2018	Extra Homework Time	Rewarding Appropriate Behavior	-1.0671	0.026	.04	Class Discussion/Activity OR Evidence-based Discussion	1

At the beginning of the Spring semester, the “advantages of engagement” method was significantly collinear with the “student studio experience” predictor, which significantly predicted agreement with the time spent on “extra reading”. However, all other predictors, including the “advantages of engagement” method only explained about 7% of the “student studio experience” predictor, so it is unlikely that the effect found is actually from the “advantages of engagement” method. Two formats were found to significantly predict student agreement with the time spent on “extra reading”. Students of the two instructors who used the “reminders throughout the semester” discussion format were less likely to agree with the time spent on “extra reading”. Considering that instructors might use the “reminders” format in response to student resistance to “extra reading”, it may not be surprising that this format was associated with lower agreement. “Rewarding appropriate behavior” had a positive effect on the likelihood that a student will agree with this activity. However, only one instructor used this method at the start of the semester so this result may actually be due to some aspect of that instructor or the class that is not included in the model. The effect of “student studio experience”, “reminders”, and “rewarding” predictors was somewhat sizeable with an effect size of 0.13.

By the end of the Spring semester, student who perceived that their instructor had discussions about why they should agree with out-of-class activities were more likely to agree with the time spent doing “homework”. 36% of the “perceived agreement discussion” predictor can be explained by other predictors, including the “perceived discussion” predictor that it was significantly collinear with. So, it is possible the effect is due to a “perceived discussion”. The “evidence-based discussion” method, “class discussion/activity” format, or the “rewarding appropriate behavior” format (all of which are perfectly collinear), had a negative effect on student agreement with the time spent on the “homework” and “extra homework” activities. However, only one instructor used these methods and formats, so the effect may actually be from an aspect of the instructor or class that is not included in the model. The effects of these discussion were small, with effect sizes of 0.04.

6.4.6 Summary of Quantitative Findings

Using the PUMBA and SIMBA, I quantified student agreement with the activities that occur for a studio physics class and the strategies instructors use to discuss those activities and/or get student agreement with those activities. Taking the results of these surveys and analyzing them together using hierarchical models, I was able to determine what affects student agreement to each activity in terms of the importance placed on them and the time spent on them. Most of the findings are that a discussion about the activity and/or a discussion about why students should agree with the activity does increase student agreement for various activities. There are some cases in which a general discussion about the activity (but not why the students should agree with it) leads to decreased agreement. This may be due to bringing students out of ignorance by discussing what they should expect of an activity they may prefer more or less of. For example, a discussion about demonstrations may cause the students to realize that the instructor does not place as much importance on them as they would prefer. If the instructor did not discuss demonstrations, it is possible that the students agreed with demonstrations because they *assumed* it would be just as important in their actual class as they would like in their most effective class.

There were few specific methods or formats that had a significant effect on student agreement. Often, even if a significant effect was found, only one or two instructors used that method or format, so it is uncertain if the result is due to that method or format or if it is due to some commonality between the classes that used the method or format that is not represented in the analysis model. Additionally, nearly every effect found with this analysis had a small effect size. On average, the effect size of a model was approximately 0.05, meaning that even though there is a statistically significant effect, it has little impact on the outcome. Future studies like this research would be improved by a larger, more diverse set of classes with instructors that use various methods and formats.

With these findings, I conclude that a discussion about the class format and why students should agree with it will usually lead to slightly greater agreement. When such a discussion does not lead to greater agreement, it may be due to students discovering that an aspect of the class will not meet their previous expectations that they might agree with (e.g. the importance of demonstrations). When this occurs, an instructor must be sure to discuss why students should agree with that aspect of the class as this always increases student agreement.

I have found that making an effort to gain student agreement is at times successful. However, my goal was to quantitatively determine specific strategies that instructors can use when having these conversations. Unfortunately, the quantitative data does not reveal many effective strategies. This does not mean that the methods and formats are ineffective. Finding no effective strategies with the collected data means that I do not have sufficient evidence to prove that these strategies are effective. Perhaps with more participating classes and more varied instructor strategies, I may find more effective methods and formats. Although I was unable to provide quantitative evidence of effective strategies for gaining buy-in, I may be able to qualitatively determine effective methods and formats.

CHAPTER SEVEN:
QUALITATIVE FINDINGS OF EFFECTIVE BUY-IN STRATEGIES

I interviewed five instructors and nine students with a primary goal of learning what they believed were effective and ineffective strategies for gaining student buy-in. In the instructor interviews (see Appendix E), I asked instructors what strategies they used to gain student buy-in and if they felt any of those strategies worked. Along with these questions, I presented the instructors with the list of methods and formats they could possibly use (see Tables 15 and 16). From this list, I asked instructors to indicate the methods and formats they felt would be effective or ineffective for gaining student buy-in. In the student interviews (Appendix F), I asked students what their instructor could have done to get them to agree with the class format. I provided the same list of methods and strategies to students so that they could list the specific strategies that they felt would be effective or ineffective. In this chapter, I discuss the responses to these questions. I also discuss what students stated were the reasons for their agreement at the start of the semester and the reasons for their change in agreement throughout the semester.

7.1 Student Responses

7.1.1 Actual Influences on Student Agreement

Before delving into the potential effective and ineffective methods and formats from the student responses, I wanted to investigate what influenced student agreement at the start and throughout the semester from the students' perspectives. For the student interviews, I coded student responses (see Table 45) searching for the methods and formats that could influence student agreement and other influencers such as prior knowledge/experience/opinions and experience in the class. In Table 45, methods, formats, and other influencers are represented in their abbreviated form. Methods begin with "M-", Formats being with "F-" and the other influencers are “PK” for “Prior Knowledge”, “CE” for “Current Class Experience”, “NIE” for “No Instructor Effect” and “TRUST” for “Trust/Faith in the Instructor”. See section 5.3.3 for a discussion on how these codes were developed, their definitions, and an example quote for each new code.

Table 45. An outline of the influences on student agreement at the start and throughout the semester.

Instructor	Student	Beginning In-Class Importance	Beginning Out-of-Class Importance	Beginning In-Class Time	Beginning Out-of-Class Time	Change In-Class Importance	Change Out-of-Class Importance	Change In-Class Time	Change Out-of-Class Time
1	1	PK, TRUST	M-BENEFIT, F-LECTURE	PK, F-LECTURE	PK, F-LECTURE, TRUST, M-BENEFIT	CE	CE	CE	CE
1	2	PK	PK	-	-	CE, NIE	CE	CE	CE, F-REMIND, M-ENGAGE, M-BENEFIT
2	3	PK, NIE	PK, NIE	PK, F-LECTURE	PK	CE, NIE	CE, NIE	CE	-
2	4	PK	PK, F-AV	-	-	CE	CE	CE	-
4	5	PK, F-LECTURE	PK, NIE, F-LECTURE	PK, NIE, F-LECTURE	M-BENEFIT, F-LECTURE	CE, F-LECTURE	CE, F-LECTURE	CE, M-ENGAGE	CE, M-BENEFIT, F-LECTURE
4	6	PK, F-LECTURE	PK, NIE	NIE	PK	CE, F-REMIND, M-ENGAGE	CE	CE, M-ENGAGE, F-REMIND	CE
7	7	PK, M-EVID	PK	PK, F-LECTURE, NIE	PK	CE, NIE	CE	CE	CE
7	8	PK, NIE	PK, NIE	PK, F-LECTURE, M-BENEFIT, NIE	PK, F-LECTURE, M-STUDENT	CE	CE	CE	CE, NIE
8	9	PK, F-LECTURE, M-BENEFIT	PK, F-LECTURE, M-BENEFIT	PK, NIE	PK, TRUST	CE	CE	CE, M-BENEFIT, TRUST	TRUST

“PK”: Prior Knowledge; “CE”: Current Class Experience; “TRUST”: Trust/Faith in Instructor; “NIE”: No Instructor Effect; “M-”: Method; “F-”: Format

Based on student responses, few strategies that an instructor might use appeared to influence student agreement. Nearly all the students mentioned prior knowledge/experience/opinions as the thing that influenced their agreement to each aspect of the class format at the start of the semester. If there was something done by the instructor, it was most often the instructor describing how the class will work (F-LECT) and the student recognizing that the class format is or is not aligned with their prior ideas of their most effective class. For example, student 7 was asked why he disagreed with the time spent on lecture and said in response,

> *"Um, just based on how he explained the class it sounded like we weren't gonna get a whole lot of lecture. And I would, was thinking that I would like more, a little bit more, because the way that he presented it to us it was only gonna be like 20 minutes of class. Um, which, and I was kind of hoping for a little bit more lecture because physics is hard and I need instruction with it."*

For this response, I coded "F-LECT" because the student said that the instructor "explained the class" and "PK" because he knows of himself that he needs more instruction with physics. Similarly, for student 9, when asked if anything the instructor did affected her agreement with the importance placed on out-of-class activities, she said,

> *"Um, more about his framing of the importance of things. So he very much framed that reading the text would be important. Um, doing the homework would be very important., and getting help from the help sessions or TAs or a tutor would be very important."*

When asked why her instructor saying something was important led to her to agree with it, student 9 said, "Um, most likely because I had taken [this course] before, and I was looking for something to be different so I could do better, hopefully." These responses were coded with "PK" because she was determining her agreement by comparing what she was told to the class she had taken before. So, by the instructor explaining how the class will work, students realized whether or not the class was formatted similarly to their hypothetical most effective class; a most effective class which they had unchanged previous conceptions of.

The other strategy that was effective at influencing agreement for multiple students at the start of the semester was the instructor simply stating that an aspect of the class will be beneficial. For example, student 8 agreed with less time spent lecturing and more time spent on his own because, "[The instructor] told us, like I said before, we needed to do that, that way we could actually learn, but we needed to have, you know, put our feet in the water a little bit first through lecture…" Student 5 was also positively affected by his instructor stating the benefits of a class activity. He stated that he agreed with the time spent on completing assigned problem-solving activities because "[The instructor] did put, like, a very heavy emphasis on that. That, you know, this is what's gonna help you out the most in the class."

I asked students what affected their agreement at the start of the semester and got the responses described above. After this, I asked students why their agreement changed throughout their current semester. Like the start of the semester, experience is the primary influence on agreement throughout the semester instead of any strategies used by the instructor. As the semester goes on, agreement is influenced by experience with their current class instead of experiences from before the semester began. Nearly every student, for every class aspect, mentioned that their agreement changed over the semester due to their experiences of their current class. Student 6 describes why her agreement with the importance of in-class activities increased over the semester saying, "Basically because I just thought about, like, what we did in class, um, and how that tied to doing well in the class, so, I mean, that, that's where I came from." Student 9 said, when describing her agreement with the time spent on student presentations, "So I just assumed that since we didn't really do it, and I'm doing fine that the amount of time that we're spending doing it probably is fine.”

At the beginning and end of the semester, there were few instructor methods that students cited as an influence on their agreement. Most of the time, students either implied there was no instructor effect (by not mentioning any instructor strategies) or explicitly said there was no instructor effect after being questioned about it (coded as “NIE”). It may be possible that there were instructor effects, but they were not salient to the student. It may also be the case that students incorrectly recalled the influences on their agreement, perhaps exhibiting a hindsight bias similar to that found in Appleton-Knapp and Krentler’s study (Appleton-Knapp & Krentler, 2006).

7.1.2 Potential Influences on Student Agreement

After discussing with students what influenced their agreement at the start of the semester and throughout the semester, and finding that instructor strategies had little effect, I asked students what they felt would be effective strategies to influence their agreement. For this section of the interview, I had students refer to the list of instructor methods and formats. Their responses, as well as instructor responses (see section for 7.2 for discussion), are summarized in Table 46. One of the nine students did not reach this section of the interview due to time constraints, so Table 46 summarizes responses from eight students. To see how each student responded, see Appendix G.

Table 46. Number of responses from students and instructors for potentially effective strategies for gaining buy-in.

	Students (8 max)		Instructors (5 max)	
Strategies	Effective	Ineffective	Effective	Ineffective
Methods				
Advantages of Engagement	4	-	-	2
Comparison Between Reformed and Traditional Classes	5	1	-	1
Student focused Discussion	4	1	1	-
Evidence-based Discussion	6	-	4	-
Extended Example	2	1	-	-
Try and See	2	4	-	1
Formats				
Lecture	4	-	2	1
Class Discussion/Activity	6	2	1	-
Reading/Syllabus	4	2		2
Audio/Video	-	5	1	-
Reminders Throughout the Semester	4	-	1	1
Rewarding Appropriate Behavior	5	-	1	1

7.1.2.1 Potentially Effective Methods

Students were fairly consistent with their responses for potentially effective strategies to gain their buy-in. Most methods and formats were mentioned by at least four students. I will discuss the two top methods and two top formats in terms of the number of students mentioning their potential positive influence on their agreement.

Six students stated that an evidence-based discussion would be an effective way to gain their agreement. Recall that few instructors used this method for any one activity. One student did specifically mention that an evidence-based discussion would only be effective in the context of a student focused discussion as well. Student 5 said,

> *"It's that, you know, um, logic is super important. Logic is super important. But also, you know, um, being able to just, like, talk to the students as well, like, relating to their experiences, relating to the emotions, you know, regarding to the class. It also has to like, take part. It's not just, like, you know, the evidence. Because sure, evidence, evidence is evidence. But the students are the ones that are experiencing the class."*

The other students did feel that evidence-based discussions without the student context would still be effective. Student 9 said, "I would have been more swayed if there was evidence based discussions that actually had clear evidence of how this would be beneficial towards all of us learning." Student 1 specifically stated an evidence-based discussion about how people learn would affect student agreement. When asked what would be the best way to convince her to agree with her studio class she said, "Probably discussion of how people learn, in all honestly. Just because, if you're trying to tell, if you're telling how other people learn, that would convince you more on how studio lab is run."

Five students said that a comparison between reformed and traditional classes would help sway their view of an effective physics class, sometimes alongside the "evidence-based discussion" method. Student 9 discussed using the "comparison" method in combination with an evidence-based discussion, stating,

> *"Um, and with the evidence based, if there was to be a comparison between the reformed and traditional classes, I think it might have been better for [the instructor] to talk about how much it'll be different and how much this will actually emphasize our self-learning."*

Student 7 also said something similar when describing why she said the "comparison" method would have been effective. She said, "Uh, I would've liked to hear more about advantages maybe that have been proven of the [class] ... I think that definitely would've swayed me more.” When asked to explain what she meant by “proven”, she responds, “Well I mean if they're like, I guess it would also go under like presenting studies and stuff but just really discussing like what the advantages are.”

Both of these methods could be classified as the logos type of argument using Aristotle’s rhetoric theory. An evidence-based discussion is clearly an argument rooted in logic, thus being classified as a logos argument. A comparison between reformed and traditional classes could also be a logical argument. For instance, student 8 described how his instructor (instructor 7) made a comparison between lecture and his reformed class. Student 8 said that the instructor told him that,

> *“there are like certain things that you can only really learn from firsthand, um, you can learn the material from lecture, but sometimes it's just ... it's easier to learn or it clicks better if you kind of do it yourself or you see it in action, so you have something to compare it to.”*

His instructor is appealing to reason or logic. In essence, the instructor is saying that student should agree with this class because they will learn better and in a way that cannot be done in a lecture class.

This is not to say that students do not think pathos, an appeal to their emotions, would not be an effective argument for gaining their agreement. Four students cited the “student focused discussion” method, which is primarily an appeal to students’ emotions, as a potential effective method to get their agreement with the class format. Student 4 appreciates the comforting students aspect of a student focused discussion. He says,

> *“Student focused discussion. Yes, that's just ... That's just automatic yes. Um, discussing, uh ... Encouraging, comforting students, letting them know that it's okay to be wrong, or making mistakes. Enjoyable? Um, I'm not a*

robot. I can't be perfect all the time, so knowing that it's okay if I mess up, always being told ... It always comforts me a little bit inside."

Student 6 agrees with the comforting aspect of a student focused discussion. When asked what she felt would be an effective method for gaining her agreement, she said,

> *"I think the student focused discussion, I know that throughout the semester, like, after tests, you know, he would let us know that it was okay, like, you know, if we did bad we'd have time to improve, um, you know, 'Stay engaged and you'll be fine,' like, that encouraging and comforting aspect allowed me to agree with him."*

7.1.2.2 Potentially Effective Formats

The most commonly cited potentially effective format was the "class discussion/activity format" indicated by six students. Student 4 explains his choice saying, "I'm actually persuaded by class discussion and activity, because this gives us a chance to kind of question why it kind of works the way that it does." Student 6 compares the effectiveness of the "discussion" format to that of the "lecture" format, saying,

> *"I mean, I would just say that, like, over lecture. Like, when I'm thinking about lecture I mean, me personally, like, I can, you know, hear what you're saying, but that won't necessarily, you know, allow me to agree. However, talking with other students, like, in a class and how they feel about it and everything, you know, I can compare, you know, how I feel to how they feel about it, um, you know, see if everyone is, like, on the same page, so I don't know. I think that I would benefit more rather than the lecture."*

Every instructor used the "lecture" format and the "class discussion/activity" at some point, but they used "lecture" far more often than they did a class discussion or activity. Instructors could improve their buy-in discussion by holding a class discussion or a class activity.

The next most commonly cited potentially effective format is the "rewarding appropriate behavior" format. Students recognize that incentives are a powerful way to get someone to do something. Students 6 and 7 put it bluntly, saying, "I don't know, maybe rewarding appropriate behavior. I mean, everyone likes rewards," and "Um, I mean, incentivizing, I'm a student, so I'll always do something for an incentive."

There were other formats that numerous students said would be effective, such as using the "lecture" format, placing information in the syllabus, and giving students reminders throughout the semester. Student 1 describes the potential effectiveness of the putting information about the class format in the syllabus by saying that it gives students an opportunity to learn how the studio will operate and decide then if they want to "partake in that kind of learning style." Student 5 describes the benefits of the "reminders" format saying that it helps the students focus on why "certain things about the class" are happening.

7.1.2.3 Potentially Ineffective Strategies

One method and one format were clear as likely ineffective strategies for gaining a student's agreement with the class format. Four of the six students who identified an ineffective method included the "try and see" method. Student 6 says this method would be ineffective because she could try it but not necessarily agree with it, so it would be ineffective at gaining her agreement. Student 7 dislikes the lack of substance and evidence of the "try and see" method. Student 9 also said that she would prefer evidence over the "try and see" method, despite being a trusting person. It appears, that students perceive the "evidence-based discussion" and "try and see" methods as opposite ends of a spectrum where they prefer the evidence.

For an ineffective format, five of six students said the "audio/video" format would not influence their agreement with the class format. Student 9 believes that using audio or video to persuade someone would be impersonal, saying, "Um, for the audio or video, why it would be ineffective is that, um, it would seem more of, like, an ingenuine way to persuade someone to think about it, whereas a lecture is more personal.” The “audio/video” format does not seem to connect well with student 6 either. She said, “And then for the ineffective, yeah, I don't think audio or video would really be helpful. Again, it goes along with reading, like, it's just something that you see and you’re just like, ‘Oh, okay.’”

From the student responses, it seems that they would be best influenced by an argument based in facts and to have a discussion about the class format in a more personal, connected format. This is evidenced by the majority of students indicating an “evidence-based discussion” method as an effective method and the “try and see” method as ineffective. Also, students would prefer to have a class discussion or activity when talking about the class format instead of listening or viewing something in the “audio/video” format, which they seem to find less personal.

7.1.2.4 Trust/Faith in the Instructor

Many of the instructor strategies to gain student agreement could be classified as either a pathos argument, logos argument, or a combination of both. The third aspect of rhetoric is ethos, which is the support of an argument based on the competency, caring, and character of the person making the argument (J. C. McCroskey & Teven, 1999). If an instructor has an established ethos, this could be identified by the students’ trust or faith in the instructor and this is evidently an influencer of students’ agreement with the class format.

Three students mentioned their trust in their professor as a reason for their agreement at the start or throughout the semester. When asked if the strategies her instructor used helped her agree with the class, student 1 said, “I believe so yeah. She told us a lot about how like, studio lab would work at the beginning. So I trusted her judgment, like I went with it.” Student 6 mentions trust in the instructor when describing why she feels the “advantages of engagement” method would be beneficial. She said, “Um, I guess, I guess because the teacher knows best, so it's like, they're, they're giving you the tools to, you know, pass the class…” Student 9 not only mentions her own trust in the instructor but also mentions that she was swayed by his confidence. When asked how her instructor’s strategies affected her agreement, she said, “Um, with his confidence in it being different, that did, uh, sway me towards being more receptive towards it. And I just trusted that they would- he would know what was best for learning and teaching physics.”

To help gain student agreement, instructors should strive to portray trustworthiness which helps establish their ethos. For these three students who cited trust in their professor, it took no other arguments to gain their agreement. Once the students decided that they trusted their instructor, they agreed with whatever the instructor told them. If the instructor’s credibility and trustworthiness (i.e. ethos), are established they may not need to put much effort into other agreement gaining strategies.

7.2 Instructor Responses

7.2.1 Instructors’ Perceptions and Experiences with Student Resistance

To explore instructors perceptions and experiences with student resistance, and also lay the groundwork for future research in this area, I began instructor interviews with a discussion about student resistance. Instructors experienced or expected student resistance to the general style of the studio class, particularly interaction, and the work done outside of class. One instructor said,

> *“It's much more, it requires more interaction of students and it's different than just like being able to sort of passively sit and maybe skip lecture or not pay much attention or things like that. Um, so I expected a little bit of resistance, yeah.”*

Another instructor expected resistance to many aspects of the class, both during the class time and outside of class.

> *"Oh, I knew exactly what they were going to resist. They were going to resist doing things on the whiteboard instead of their own notebooks. They were going to resist being put into groups and then any subsequent change in groups. And they were going to heavily resist any assigned reading. And then they also tend to resist just homework as an idea."*

Like this instructor, another instructor experienced student resistance to the work the students do outside of class. He said, "Uh, well, I definitely started recognizing they did not like, the ... what we were using for pre-lecture stuff." He continued on to say that he felt as though students were watching the pre-lecture videos and finishing the associated quizzes as quickly as they could, a form of resistance. As a result from speeding through the pre-lecture videos, the instructor said that the students did not come prepared to class. He said, "students really should have a small grasp on the content before they come in, but whenever I would teach, that didn't seem evident…"

In this preliminary qualitative investigation into instructor perceptions of student resistance I find some of the instructors in my study expected and experienced student resistance to out-of-class activities and the level of interaction required in a studio class. Future work on instructor perceptions of student resistance could look into the source of these expectations, and how instructors respond to experienced resistance (do they alter their instruction to appease the students?).

7.2.2 Potentially Effective Strategies

When asked what methods and formats they felt would be effective at gaining student buy-in, instructors had the list of strategies in front of them to refer to. These qualitative findings are based on the specific strategies in this list that the instructors responded with. In Table 46, I outline the methods and formats that instructors indicated could be effective or ineffective. Instructor 4, who is the instructor that used many of the strategies at once, is not represented in this table because he said, “Um, I just have try everything. I don't think any particular one will work, because students are different,” when asked which specific strategies he felt were effective.

There was little consistency between the four instructors in which formats they felt were effective or ineffective. Two instructors said the lecture format would be effective while one said it would be ineffective. Similarly, one instructor said the “reminders throughout the semester” and “rewarding appropriate behavior” formats would be effective while another says it would be ineffective. So there appears to be varying perspectives among instructors about how they should talk about the class format and why students should agree with it.

There is agreement among the instructors about which method they believe would be the most effective at gaining student buy-in: using an evidence-based discussion, or one based upon factual statements. Instructor 1, when discussing the potential of using an evidence-based discussion says,

> *“I think it would be interesting for like specific types of activities, like maybe not for the class as a whole but like if a specific topic, there's been something shown, this one activity works really well to help students learn it. So, evidence that that works. I think that would be an interesting thing. Plus, you could tie it into like how people learn science and what graphs are and it's a nice science way to tie it into.”*

Instructor 7 noted that when he used an evidence-based discussion it seemed to increase his class’ test results, thus making him believe it was an effective method at gaining buy-in. When asked if there was anything that he could do that he thinks would help get students to agree with the way the class works, instructor 8 said, “One thing that I would like to do is give a better evidence-based discussion, but we just really don't have a lot of evidence built up from our own implementation yet."

So, these instructors all feel that an evidence-based discussion could work or has worked. However, few instructors throughout the study actually used an evidence-based discussion. Like instructor 8 stated, this could be due to not having the evidence-based information readily available to them. Instructor 1 indicated something similar, saying, "I don't know that evidence-based things, I don't always…I didn't usually have an exact citation…" When asked if she would use that information in her class if she had it, she said that she might.

Few instructors mentioned possible ineffective methods, though two did state that the "advantages of engagement" method might be ineffective. Instructor 2 cited students coming to him at the end of the semester saying that they "don't read the book, because they don't get anything out of it". He went on to say, "that implies to me that I'm not conveying well enough, like, what are the advantages of actually, like, buying- buying into the course format." He is stating that even though he does discuss the advantages of engagement, it must not be effective because students are not engaging with the course.

7.3 Summary

In the quantitative analysis on the effect of instructor strategies on student agreement, I found few methods and formats that led to a statistically significant increase or decrease in agreement. Some of the more common effects were with the "rewarding", "class discussion/activity format", "lecture", and "reminders" formats. Few methods were found to be significant predictors of student agreement. The "evidence-based discussion" method was the most common significant method found in the quantitative analyses, appearing in two models. These strategies were often collinear with other strategies, and often only up to two instructors varied from the rest in terms of whether or not they used a strategy. To provide support for the quantitative analysis and attempt to answer research question 4 through a different means, I conducted interviews with instructors and students. Much of these interviews focused on what were or could be effective strategies to gain student buy-in.

I found that student agreement was largely based on experiences either before the class began or during the semester. This finding is consistent with research in expectancy value theory applied to active learning, which also found that students' positive experience with active learning primarily determined if they valued active learning (Cooper et al., 2017). Student agreement at the beginning of the semester was dependent on whether or not the class format, as described by the instructor, aligned with the hypothetical most effective class students had imagined before being oriented to their current studio class. That is to say, little of what the instructor might have done affected what the students thought of as the most effective physics class. If a student agreed, it was more likely due to the studio class coincidentally being of the same format as their previously imagined hypothetical most effective class. However, to get this agreement, the instructor has to describe the class format, which explains why the "lecture" format was somewhat commonly cited as an influencer on their agreement. This also explains why "perceived discussion" was a very common significant predictor of student agreement in the quantitative analysis. Throughout the semester, instructors still had little influence on a student's agreement. As the semester goes on, if the student liked aspects of the class, or felt they were doing well in the class, they were more likely to agree with the format.

Next, I explored what students and instructors considered as potentially effective and ineffective agreement gaining strategies. Instructors did not agree on what they felt were effective or ineffective formats, aside for two instructors agreeing that the "lecture" format was effective. There was better agreement among the students. They felt that a class discussion or rewarding appropriate behavior would be effective ways of talking about the class format and/or why they should agree with it. These two formats do appear as a significant predictor in the quantitative analyses and they are often collinear with each other. However, whereas the students say these would be effective at gaining their agreement with the class format, these formats were more often than not a negative effect on student agreement. The discrepancy could perhaps be due to the low number of instructors to use these methods, thus skewing the quantitative effect towards the overall agreement level of those few instructors. What students felt would be ineffective at gaining their agreement was using the "audio/video" format.

Moving on to potentially effective and ineffective methods, there was agreement within and between instructors and students. The majority of the instructors and students who I interviewed believed that an evidence-based discussion would be an effective method for gaining buy-in. Although most everyone agreed on its potential effectiveness, few instructors used the "evidence-based discussion" method. This is likely due to instructors not having such materials at the ready. Henderson and Dancy have found that instructors question the results of educational research (Henderson & Dancy, 2008), believing that there is too much focus on conceptual inventories, the comparisons made between typical instructors and teachers highly interested in education research is inappropriate, and that studies are too short term. So there is reason to question if instructors will successfully use an evidence-based discussion since they might not trust the research themselves. Finally, only one instructor agreed with the students that the “try and see” methods would be ineffective at gaining student agreement.

CHAPTER EIGHT: CONCLUSIONS

8.1 Summary

My prior research into the influences on Graduate Teaching Assistants in a reformed physics discussion/lab hybrid section motivated this dissertation. In my previous study, I found that even though GTAs bought in to the reformed teaching style, their in-class actions were more similar to what they thought their students wanted, which was not in the reformed style (Wilcox et al., 2016). It seemed that the GTAs' perceptions of their students' views had a larger influence on the instructors' actions than the instructors' own beliefs. I hypothesized that rather than asking instructors to ignore what they think their students want, it may be more productive for the instructor to get students to want to learn in the reformed manner that the class is taught. This concept (of the students understanding and agreeing with the class format) is often called student buy-in.

When students do not buy-in to the class that they are taking, they may exhibit resistance to the teaching methods implemented in the class. Instructors and GTAs across STEM fields have cited student resistance as a barrier to the implementation of research-based instructional strategies. Numerous studies have shown instructors experience student resistance, including studies on physics faculty (Dancy & Henderson, 2005, 2010; Henderson & Dancy, 2007; Turpen et al., 2016; Yerushalmi et al., 2007), GTAs and LAs in physics SCALE-UP classrooms (DeBeck et al., 2010), GTAs in biology and engineering (Bautista et al., 2014; Pinder-Grover, 2013), and faculty in engineering (Froyd et al., 2013). In my study, instructors experienced or expected student resistance to the general style of the studio class, particularly interaction, and the work done outside of class.

In an effort to decrease student resistance, I researched student agreement with reformed instruction, particularly the studio style of teaching introductory physics courses. The studio style was designed from physics education research and so employs many research-based instructional strategies that students may resist. First, I wanted to determine what aspects of the studio style class format students agreed or disagreed with. To do this, I developed a student survey called the "PUMBA" (Perceptions of Undergraduates Measure of Buy-in Achievement) to measure student agreement with the class format through their agreement with class activities. I defined the class format through four concepts: 1) the importance of in-class activities; 2) the importance of out-of-class activities; 3) the time spent on in-class activities; and 4) the time spent on out-of-class activities. In addition to seeing how students agreed with those four aspects, I further explored student agreement with the individual activities that construct each of the four overall measures. I hypothesized that increased student agreement would make it easier for instructors to use research-based instructional strategies which would then result in increased performance. To test this hypothesis, I examined the effect of student agreement on student performance through measures including conceptual test scores, attitudinal test scores, and expected final grade.

Following my investigation into the details of student agreement, exploring the aspects of the class format, the specific activities that students agree or disagree with, and how student agreement affects student performance, I wanted to discover the practical strategies that instructors could use to achieve student buy-in. To begin, I surveyed instructors using a survey I developed called the "SIMBA" (Survey of Instructor Methods for Buy-in Achievement), which was used to record the strategies instructors used to gain student buy-in. Instructor strategies were categorized as either a method, for discussing why students should agree with a class activity, or a format, the means of having the conversation. Once I described the commonly discussed activities, and the commonly used methods and formats, I investigated how these strategies affected student agreement in the instructors' classes. I approached this investigation through quantitative means by analyzing the responses to the SIMBA and PUMBA together. I also took a qualitative approach by interviewing instructors and students to learn what they felt were effective strategies for gaining student buy-in.

Below, I list my research questions and summarize the main findings while connecting the findings to my theoretical and conceptual frameworks (Chapter 2).

8.1.1 How Well Do Students Buy-In to Reformed Instruction?

8.1.1.1 How Well Do Students Understand the Class Format?

Student understanding of the class format (which is defined as the magnitude of the difference in perceptions of the time spent in class between the students and the instructor) depended on whether they preferred less, equal, or more time spent on an in-class activity. For every in-class activity, students who preferred less time spent on an activity would perceive that activity as one that would or did occur more often than the frequency that the other students expected or perceived. For most activities, students who would prefer more of an activity perceived that activity as one that would or did happen with the same frequency that students who agreed with the activity perceived it occurring. As student understanding is calculated based on how frequently a student perceives an activity will or did occur in their actual class, the answer to this question is different for these groups of students.

For roughly half of the in-class activities across the four distributions of the student survey, student understandings of a class activity were statistically similar between the three groups of students (see Table 28). When they were not similar, it was almost always the case that students who preferred less of an activity had a greater misunderstanding than the other students. Only with the "small group work" activity did students who preferred less of that activity have a statistically better understanding. It is somewhat surprising to find that student perceptions and understandings are not consistent across a class. One would expect all students who have experienced the class orientation on the first day, or even experienced the class throughout the semester, to have similar perceptions of how the class operates. The varying perceptions among students based on whether they prefer less/equal/more time spent on an activity might be an issue of communication, as these students have different understandings of the class time. Connecting this back to the instructional communication framework, instructors should better communicate their expectations to the students to ensure proper understanding among all students. If instructors are perfectly clear about their expectations, there should be no room for misunderstandings among the students. Communication research has several recommendations to ensure instructors clearly communicate their ideas. Chesebro (2002) states that there are two components of clear communication: verbal and structural clarity. To achieve verbal clarity, instructors should communicate with fluency (no stammering or pauses) and use explanations and examples. Structural clarity "enables students to develop a clear schema for course material," (Chesebro, 2002, p. 97) or in this context, the class format. Elements of structural clarity in instructional communication include previewing the content of the discussion, properly organizing the discussion, reviewing the content, and providing an outline of the main points (for example by clearly describing in the syllabus how important activities are and how much time will be spent on them).

8.1.1.2 What Specific Class Activities Do Students Properly or Improperly Understand?

Many students had misunderstandings of "lecture", "step-by-step labs, "investigative labs", "small group work", and at times "individual work" (see Figure 12) as compared to what the instructor said would be the time spent on these activities. For these activities, students who preferred less time had a greater misunderstanding than the other students, except for the “small group work” activity and in several classes, the “step-by-step labs” activity. At times, students did have the correct understandings for lecture, but when they did not, they had large misunderstandings ranging up to 40 percentage points difference between what the students perceived would or did occur and what the instructor perceived. Students often underestimated the amount of time that would be or was spent on step-by-step labs, with misunderstandings up to 40 percentage points less time spent on step-by-step labs than what the instructor stated. Many instructors indicated the investigative labs would not and did not occur very frequently, less than 5% of the class time. Student in these classes often overestimated the time spent on investigative labs by up to 20 percentage points. Students underestimated small group work by roughly 10 to 30 percentage points. When students misunderstood individual work, the difference between what they perceived and what the instructor perceived reached up to 20 percentage points. Students had fairly correct understandings of the time spent on the remaining activities: “quiz/tests”, “student presentations”, “demonstrations”, and “whole class discussion”.

Considering that there is some dependence of student perceptions of the class on what they would prefer in their class, it may not be surprising that students are incorrectly perceiving activities like lecture and small group work. Less lecture in a studio class is one of the features that differentiates studio from traditional. Similarly, studio classes incorporate more group work than a traditional class. Given these results, it appears that students’ perceptions might be biased towards the features of a traditional class.

8.1.1.3 How Well Do Students Agree with the Class Format?

Student agreement with the overall class format can be found in Tables 29 and 30. The average class agreement to each of the four overall agreement measures did not vary from class to class. Student agreement with the importance placed on class activities and the time spent on out of class activities, was typically in the 50-65% range, meaning that students agreed with the importance placed on 50-65% of class activities or they agreed with the time spent on 50-65% of the out-of-class activities. These results show that there is much room for improvement in student agreement with these aspects of the studio class. Student agreement with the time spent on in-class activities was given a score on a scale of 0-100 with larger values indicating stronger agreement. Across all classes, the average in-class time agreement score ranged between scores of 70-85, which is a fairly high level of agreement but still with some room for improvement.

Although each class has a similar class average agreement to the four aspects of the class format, student agreement within a class varies greatly. The standard deviation of student agreement levels for the importance agreement measures and the out-of-class time agreement measures were around 20, which is about one third to one half of the average score. The standard deviations of the in-class time agreement measures were sometimes smaller, because it is a more precise measure. However, the standard deviations were still on the same order of magnitude as the average, indicating a significant amount of variation with each class. This means that from class to class student agreement is on average the same, but from student to student there is a wide variety of agreement levels.

It may be expected that all students have different ideas of what makes their most effective physics class, thus resulting in a wide range of student level agreement scores. The similarities in the class averages across each class suggests that the strategies these instructors used were ineffective as they used different strategies but got the same results. Before drawing conclusions from these findings, I should look at student agreement to the specific activities.

8.1.1.4 What Specific Class Activities Do Students Agree or Disagree With?

The proportion of students who agree with the importance placed on or time spent on each individual activity can be found in Tables 32 and 33. The two most important findings from these tables have to do with student disagreement with the time spent on lecture and disagreement with the “extra reading” activity. The time spent on lecture was the most disagreed with activity by a decent margin (between 6 and 12 percentage points lower than the next activity). Only a third of the students agreed with this activity. One might expect that students in a studio course would prefer more time spent on lecture but this is not actually the case. Students are fairly uniformly distributed over the agreement groups; roughly equal proportions of students prefer less lecture as students who prefer more lecture. For instructors, this means that their discussion about lecture expectations should not only be focused on getting students to prefer less lecture time than what they initially had wanted.

Tables 32 and 33 also show that students often disagree with the importance placed on and time spent on the “extra reading” activity. This activity was always one of the least agreed with activities. Most students who disagreed would prefer less importance or time spent on the “extra reading” activity. This might be due to students correlating “extra reading” with the general concept of learning outside of class that is a feature of the studio style. Instructors who focus on getting agreement with “extra reading,” or the idea of learning outside of class, can go a long way to increasing their students’ overall class agreement.

Both of these disagreements should not be too surprising. Lecture is an activity that students probably expect but generally do not experience in a studio class room. “Extra reading”, if connected to the idea of learning outside of class, is more work for the student to do outside of class time and so they would prefer less importance and time on that activity. What is surprising is that not all students who disagree with lecture prefer more time spent on it. In fact, there are roughly equal proportions of students who want less lecture as there are students who want more. This is perhaps because students find value in the other activities that occur for the class, such as group work or labs. These students may want less lecture to have more time for those activities. These findings are placed at the intersection of the educational psychology and communication component of my theoretical framework. From the instructor’s communication of the class format, students develop agreement or disagreement with aspects of the class and specific class activities. The somewhat low agreement to the importance placed on class activities and time spent on out-of-class activities suggest that instructors may not communicate well enough how the class will operate and/or why students should agree with the planned operation.

8.1.1.5 Do These Things Change Between the Beginning and End of the Semester?

I found that three of the four overall agreement measures changed to a statistically significant degree over the course of a semester. In the Fall semester, student agreement with the importance of out-of-class activities increased throughout the semester. Also, student agreement with the time spent on in-class activities decreased. However, neither of these effects were significant in the Spring semester. Rather, in the Spring semester the increase in agreement to the time spent on out-of-class activities was significant. Although these effects were statistically significant, they each had a small effect size.

Agreement to several specific activities changed throughout the semesters. For example, agreement with "extra reading" increased in both semesters, despite remaining one of the least agreed with activities. Agreement with the time spent on step-by-step labs and whole class discussion decreased, as did agreement with the importance placed on quizzes or tests. Finally, agreement with the importance of assigned reading/annotating activities increased in the Fall semester. One could interpret these as signs of students gaining agreement with the studio style as the semester goes on. Agreement shifts towards students preferring the studio style; agreement with reading outside of class increases while agreement with the traditional step-by-step labs decreases.

8.1.2 What Effect Does Student Agreement Have on Student Performance?

8.1.2.1 Does Student Agreement with the Class Format Predict Their Performance on Concept Tests?

The findings for these questions are summarized in Table 35. I constructed separate hierarchical linear models for the FCI and the CSEM to determine if students' agreement scores predicted their post-score on the concept tests. For both the FCI and the CSEM, no student agreement measures were significant predictors of student post-scores.

8.1.2.2 Does Student Agreement with the Class Format Predict Their Performance on Attitudinal Surveys?

Only in the Fall semester did student agreement significantly predict students' CLASS post-scores. Specifically, student agreement with the time spent on in-class activities had a significant effect. A one-unit increase in a student's in-class time agreement score corresponded to a 0.17-unit increase in their CLASS post-score. This means that a student whose in-class time agreement score is 14 percentage point higher than another student will have one more expert-like response on the CLASS than the other student. When looking at the individual in-class activities, student agreement with the time spent on student presentations significantly predicted their post-score. However, these effects were very small (Ferguson, 2009), with effect sizes of 0.02 for the overall agreement score, and 0.01 for student presentation agreement. Although these are very small effects, when one considers that students typically either have the same attitudes or worse attitudes at the end of a semester than they did at the start, any positive effect should be desirable.

8.1.2.3 Does Student Agreement with the Class Format Predict Their Final Grade?

In both semesters, student agreement with the time spent on out-of-class activities significantly predicted the student's expected final grade. When looking at the individual out-of-class activities, "extra reading" in the Fall semester and assigned reading/annotating in the Spring semester significantly predicted expected final grade. All of these effects had a positive relationship with the expected final grade, meaning if students agreed more, they were more likely to expect a higher grade. If I connect the concept of reading outside of class to the idea of having to learn outside of class (as opposed to being taught in a classroom), then this result is unsurprising. Students agreeing with this idea are likely coming to class prepared as they should be, putting themselves in the best position to learn in the class and thus getting better grades. All of these effects, however, were also either very small or small, with the largest effect size being 0.04.

These findings that student agreement with the time spent on class activities has some effect, albeit small, on students' attitudes and expected final grades are solidly connected to the educational psychology component of the instructional communication framework. This component relates to the cognitive processes of students and their relation to student learning. I have found that student agreement does indeed have a connection, though small, to how well they do in the class.

8.1.3 What Do Instructors Do in An Attempt to Gain a High Level of Student Buy-In?

8.1.3.1 What Class Activities Do Instructors Commonly Talk About with Their Students to Set Their Expectations and Get Their Agreement with Those Expectations?

The strategies instructors used when discussing the class format and/or why students should agree with the format are found in Tables 36-39. The activities that instructors discussed with their students were similar at the beginning and end of a semester and across both semesters. Instructors often discussed lecture, quizzes or tests, labs, small group work, and out-of-class work. At the start of the Fall semester, many instructors also discussed the “individual work” activity, but this did not continue throughout the rest of the semester. This makes sense as many instructors said individual work would either not happen or rarely happen in class. So, they mention that at the start of the semester and do not need to discuss it any further. In the Spring semester, several instructors also discussed the “whole class discussion” activity. The activities instructors did not discuss may not have been talked about because they were either not a part of the class or would happen very infrequently. However, instructors might talk about an activity that would not occur precisely *because* it would not occur, going against what the instructor might think the students expect. This may be the case for individual work, which was not planned to happen frequently but was discussed by numerous instructors in the beginning of the Fall semester.

8.1.3.2 What Methods or Topics Do Instructors Commonly Use to Get Student Agreement?

The most commonly used methods were “advantages of engagement”, “comparison between reformed and traditional classes”, and “discussion about students”. These methods were commonly used both at the start and throughout both semesters. These methods were used by many instructors for each activity, but more so for some activities than others. For example, the "comparison between reformed and traditional classes" method seems to have been used more regarding the "lecture" and "small group work" activities. This is perhaps because these are two defining activities of a studio class. The "discussion about students" method appears to be used more often with student-centered activities like "small group work" and "out-of-class work". This makes sense as it is a student-centered method.

"Evidence-based discussion", "extended example", and "try and see" were methods not used often by instructors. It may be that instructors did not use the "evidence-based discussion" because they did not have such materials at the ready to present to their students.

8.1.3.3 Through Which Formats Do Instructors Commonly Have These Discussions?

"Lecture", "class discussion/activity" and "reading/syllabus" were the most commonly used formats. This is not surprising as lecture may be the easiest way to convey information and placing such information in a syllabus is commonplace in university classes. "Class discussion/activity" was used to a lesser extent than the other two, but it did seem to occur more often with "small group work". Like the "comparison" method, this might be explained by the fact that it is a student-centered format being used to discuss a student-centered activity. Instructors also used the "reminders" and "rewarding" formats fairly consistently across each activity, but to a lesser degree than the three previously mentioned formats. The "audio/video" format was rarely used.

8.1.3.4 Do These Things Change from the Beginning to the End of the Semester?

For the most part, the discussed activities, methods used, and formats used remained fairly consistent from the beginning to the end of a semester. "Individual work" was discussed less at the end of the Fall semester compared to the start, probably because that activity wasn't planned into the class format. Also in the Fall semester, the use of the "try and see" method decreased as the semester went on. In the Spring semester, the use of an evidence-based discussion decreased. In both semesters, the use of the "reading/syllabus" format decreased throughout the semester. This is likely because every instructor indicated that they used the syllabus for this format. Considering the syllabus is typically something every class goes over on the first day, it is not surprising that instructors used this format so much at the start of the semester and less so throughout the semester.

This exploration into the strategies the instructors in my research used to gain buy-in is the pedagogy research component of the instructional communication framework. This research investigates teaching and teaching practices. Considering that the practices that I am exploring, strategies to gain buy-in, are all discussion-based, there is overlap with communication research as well.

8.1.4 What Effect Do Instructors' Discussions of the Class Format Have on Student Buy-In?

8.1.4.1 Is the Use of Certain Methods or Formats to Get Student Agreement Associated with Higher Levels of Student Agreement?

I took two approaches to answer this question: a quantitative and a qualitative approach. First, I attempted to determine effective strategies for gaining agreement by analyzing the SIMBA and the PUMBA responses together. I formed hierarchical models to determine if specific strategies significantly predicted student agreement to a specific activity. The findings of these analyses are summarized in Tables 41-44. Ultimately, I found that, in general, a discussion about an activity usually increases agreement with that activity, especially if that discussion's aim is to gain agreement rather than just explain expectations. In some cases, however, a discussion about the class format (without discussing agreement) led to decreased agreement. This may be due to breaking the students' ignorance. For certain activities, like demonstrations, students may default to agreement because they do not know how important demonstrations are to the class or how much time will be spent on them. By discussing demonstrations, but not trying to get agreement, students might realize that demonstrations are not as important or time consuming as they initially figured they would be, thus they began to disagree with it. Another possible explanation for this flips the direction of causality. Perhaps instructors begin to discuss an activity because they notice low buy-in from the students. The discussion may be ineffective and so as a result, these strategies become associated with low buy-in.

I wanted to do more than just determine if a discussion is effective or not,; I wanted to determine the best strategies to use in those discussions. I did find some methods and formats that significantly predict student agreement. For example, I found that "rewarding appropriate behavior" had a positive effect on student agreement with the time spent on "extra reading". I also found that using the "class discussion/activity" format or the "rewarding" format has a negative effect on student agreement with the importance of assigned problem-solving homework. These effects, however, did not appear in both semesters, meaning that they may have appeared as significant by chance. Considering that for out-of-class activities, I ran eight tests on the same sets of data, I increased the likelihood of finding a significant effect by chance. Applying the Bonferroni correction to the p-values (which is multiplying the p-value by the number of tests, a method used specifically for this purpose), only the "comparison" method for group projects and the "reminders" format for "extra reading" remain as specific strategies that effect student agreement. But again, neither of these strategies appeared in both semester, casting doubt towards their actual influence. Another issue with this data is that the instructor strategy variables were often collinear with other strategies, a result of few instructors in the study and a low variation in strategies used. When strategies are collinear, I cannot determine if the found effect is due to one activity or the other. On top of that, for most of the significantly predictive strategies, only one or two instructors differed from the others in terms of using or not using those strategies. Thus, the found effect might not be from the strategy but instead from some feature of the one or two classes that differed from the rest.

To supplement the findings of the quantitative approach, and attempt to answer this question in another way, I interviewed students and instructors to determine what were effective strategies used in the class and what could be effective strategies. I found that very little the instructor did actually influenced a student's agreement at the start or throughout the semester, supporting the quantitative findings. Agreement was predominantly influenced by students' experiences and opinions developed before class and then developed from the class. At the start of the semester, a student would agree or disagree with the class format if the way in which it was described matched what they previously believed made a good class. This also agrees with my findings that a student perceived discussion of the activity significantly affected student agreement. That is to say that instructor strategies to gain agreement did not change what the students believed made a good class. All that the instructors did that influenced agreement was describe the class so that students could decide if it aligns with their prior beliefs. As the semester progressed, student agreement with the class was dependent on how they were doing in the class and which aspects they liked or disliked just by experiencing them. Little that the instructor might have done throughout the semester to affect agreement had an influence. This is similar to the work done by Henderson and Dancy that found difficulties in instructors adopting RBISs (Henderson & Dancy, 2008). Like instructors, it may be difficult to get students to buy-in to the RBISs because they may see them as dogmatic, based upon questionable research, and not developed with the students' particular situation (e.g. learning style, background) in mind.

Then what could instructors do to influence student agreement? Given the rhetorical theory framework, instructors must first establish their ethos, or in other words, their credibility as an instructor. No instructor had an issue doing that as each student interviewed felt that their instructor was competent, caring, and of good character. Some students felt that their instructor was so credible that no other strategies were necessary to get them to agree to the class format; they trusted their instructor's decisions. Beyond ethos, I asked students what other strategies (that may be classified as pathos or logos) would be effective at gaining their agreement. Students and instructors agree that an evidence-based discussion, a logos argument, should have a positive effect on agreement. However, instructors do not seem to have this kind of argument prepared, which may be why they used this method so infrequently. Instructors may not use an evidence-based argument because they question education research (Henderson & Dancy, 2008). Andrews and Lemons (2015) find that biology instructors favor personal experience over evidence when deciding to implement active learning in their class. So the instructor responses from my study, which state that an evidence-based discussion might be effective for influencing student agreement, are somewhat in conflict with previous literature stating that instructors themselves may not be convinced by evidence. However, Hora, Bouwma-Gearhart, and Park (2014) found that many faculty in their study used data to drive their instructional decisions. There were some faculty that did not, but most of the faculty in their study used data to some degree. Some used formal, statistical analyses to inform their instructional decisions while others used exam scores and student feedback (which is a form of evidence) to inform their decisions. The literature is not in consensus on how instructors use evidence in their instructional decisions. However, even if it is true that instructors do not feel evidence is effective at changing their minds, they may have a different opinion about evidence changing their students' minds. My research indicates that instructors do believe that an evidence-based discussion could influence students' agreement with reformed classes. Future work should explore if such a discussion does have a positive effect on student agreement.

In the rhetorical theory of argumentation, students and instructors agree that a logos type argument would be a persuasive way to gain student agreement. Some students also stated that a student focused discussion, an argument resembling a pathos argument, would be effective. These methods could work because the instructors had already established their ethos. Student responses to the interview indicate that a pathos or logos argument may be effective in gaining their agreement once instructors establish ethos (which they had). Students agreed that the "try and see" method would be ineffective at gaining their agreement. Although trying out the class and seeing whether it works for the students was the primary influence on student agreement throughout the semester, students said that being told 'try and see" from the instructor would not work to gain their agreement. There is a difference between being told "try and see" and actually experiencing the class and seeing its benefits. Students also agreed that the "class discussion/activity" would be the most effective way to discuss the class format. The "class discussion/activity" format describes how to have a conversation and so it is a form of communication. Recall that in communication research, communication may occur in three ways: one-way, interaction, or a transaction (Mottet et al., 2006). Based on the students' responses, gaining student agreement using transactional communication would be most effective. This is when information is shared and built upon between multiple people. Students would prefer to have a class discussion, building upon the knowledge from their instructor or from other students. They feel it is more personal and allows the students to understand and digest the discussion with the aid of their peers.

This final research question, what strategies influence student agreement, is represented by the three components of instructional communication. The strategies, what instructors should talk about and how they should have the conversation, are subjects of communication and pedagogy research. How these strategies affect student agreement with the class and subsequently their performance in class brings the educational psychology component into the mix so that this final question is truly representative of instructional communication research.

8.2 Implications

8.2.1 Instructors Should Strive to Gain Student Buy-in

Instructors should strive to achieve student buy-in for their reformed classes because student agreement does have some effect on their attitudes about physics and their performance in the class. My research has shown that when students agree with the time that they spend on activities for their physics class, they will have more expert-like attitudes about physics and expect to receive a higher grade by the end of the semester. When they agree with the time spent in a studio physics class, student attitudes about physics are more expert-like. When students agree with the time spent outside of class, particularly on reading activities, they will expect to receive a higher grade, perhaps because students correlate reading outside of class with the general concept of learning outside of class. If students agree with the concept of learning outside of class time then they are likely more prepared for and receptive of the activities that occur in class and so perform better in the class. For these reasons, it is necessary for instructors to make efforts to increase their students' buy-in to the reformed strategies that they employ in a studio physics classroom.

From the classes that participated in this study, it was clear that student agreement could be improved for several aspects of a studio class. Students only agreed with the importance placed on a little more than half of the class activities. They also agreed with the time spent on just over half of the out-of-class activities. These are three areas where instructors should work to improve student agreement. Instructors should especially work on improving agreement with the time spent on out-of-class activities as that was shown to predict higher student grades. There were a few specific activities that instructors should devote the most attention to when trying to gain student agreement. Only one-third of students agreed with the time spent on lecture. Similar proportions of students would prefer more or less time spent on lecture so an instructor should be sure to appeal to both sides when discussing lecture with their students. Assigned reading, and reading beyond what is assigned were two activities that often appeared as one of the least commonly agreed with activities. These are two of the major disagreements in a studio class, and when you consider the idea that students might equate these activities to learning outside of class in general, it becomes clear that instructors should work to achieve buy-in to reading outside of class.

8.2.2 Instructors Should Have an Evidence-Based Class Discussion When Introducing the Class Format

My research also suggests how instructors might try to gain student buy-in and what strategies might be effective. Knowledge of the strategies instructors have used or have not used can aid with the development of that first-day discussion to orient students to the class and gain their agreement. For example, seeing that student-centered strategies like the "student focused discussion" method and the "class discussion/activity" format have been used more often regarding student-centered activities may inspire an instructor to try a similar approach. Additionally, finding that instructors do not often use an evidence-based discussion, despite both instructors and students believing it would be an effective way to gain agreement, should prompt instructors or departments to prepare evidence-based materials for themselves or their faculty.

I have quantitative evidence that implies having a discussion about the class format and/or why students should agree with it does indeed increase agreement. There are a few exceptions where discussing and activity without talking about why students should agree with it negatively affects agreement, including labs and demonstrations. It may be that for these activities, students have an idea of what they expect and they agree with it, but if the instructor discusses the activity, the students might realize they are not getting what they had expected. Thus, they no longer agree with it. However, if the discussion does include topics about why students should agree with the activity, it would indeed increase their agreement.

I have qualitative evidence of the strategies that students say would be effective for gaining their agreement. Many of the students I interviewed agreed that an evidence-based discussion would help them agree with the class. Instructors also believed that this method would be effective. So clearly, it is important that we provide instructors with the evidence of the benefits and advantages of a studio classroom. I also found from the student interviews that students believe a class discussion or activity would be effective at gaining their agreement with the class format. To encourage the use of this format among studio instructors, we need to make sure instructors have planned activities or discussions regarding the class format. Instructors would benefit from learning how other studio instructors have successfully used a class discussion or activity when explaining the class format.

8.2.3 Instructors Should Highlight the Positive Aspects of Class Experiences

Finally, my research, specifically the student interviews, suggests that much of student agreement is influenced by experiences in the classroom. Student agreement throughout the semester was based on if students were doing well in the class and if they were enjoying their experience. To gain student agreement, instructors would do well to highlight the parts of the class that students are doing well on and how the class format leads to that positive experience. For example, a studio class might have an exam with two portions, one completed by the individual and one completed with a group. If the group portion received higher grades, the instructor could highlight that as a benefit of working in groups. As another example, an instructor might ask students from a previous studio class to come in and talk about all of the good things they got from the studio class. That should, in a way, provide a glimpse of their future experiences to new studio students. As student agreement was largely based on experiences in the class, this should be an effective way to gain agreement. This practice of getting advice from past students is supported in the literature (McClure, Combrink, Foor, Walden, & Trytten, 2006; Seymour, 2000). These are also, in a sense, evidence-based methods. Test scores are evidence that group work is working well. Testimonials from previous students are evidence of studio's successes.

<u>8.3 Limitations</u>

There are several limitations to this study that should be addressed. Firstly, the SIMBA and the PUMBA are both self-report surveys. Self-report surveys are prone to issues of forgetting, misunderstanding, dishonesty, and bias (Stone, Bachrach, Jobe, Kurtzman, & Cain, 1999). Students may have forgotten how much time was spent in their actual class or instructors may have forgotten if they used a certain method or format. Secondly, despite finding evidence for the validity of my survey through interviews, other students may have misunderstood the survey. The fact that students were mistakenly indicating an activity would be less important in their most effective class than their actual class, even though they later say it doesn't happen in their actual class, is proof that students were not fully understanding the survey. Students may also have rushed through the survey to earn extra credit. Although I had numerous checks in place to remove these responses, some may still have slipped through. Lastly, survey takers may have biased responses. For example, if a student was in a bad mood, they may have indicated more disagreement in their responses than if they were in a better mood. Many of these limitations were addressed to the best of my ability. For example, I tried to emphasize the most important parts of the question prompts to reduce misunderstandings. Also, I created checks so I could eliminate dishonest responses. Some limitations, like forgetting and bias, cannot be accounted for and must be considered when interpreting the results.

A second limitation to the survey is that it was only validated by students at UCF. Other students at other universities may interpret activities differently than students at UCF. They might also understand the questions differently. For example, it is possible that at some universities there is less of a distinction between activities that occur in class and activities that occur out of class. If so, students might have difficulty separating the activities as in-class and out-of-class activities. This limitation could be overcome simply in future studies by having students or instructors from other universities take a look at the survey and discuss any misunderstandings.

Lastly, and perhaps most limiting, the small sample size of classes and instructors limited my findings. Having few different instructors and little variations in the strategies that they used made it difficult for me to quantitatively determine differences between the classes. Because only one or two instructors would differ from the rest for most of the specific methods and formats, it makes it difficult to confidently say that an effect found from a strategy is in fact due to that strategy. This research would be much improved by a larger sample of instructors who use various methods and formats to discuss the class format.

8.4 Future Work

This study suggests several avenues for future research. One such avenue is to take a closer look into instructor perceptions of student resistance. An instructor might gain student agreement, but if they still perceive resistance, whether or not that resistance is accurate, they would still have difficulty implementing the research-based instructional strategies. Future research could look into how exactly instructors perceive resistance and if that changes when student agreement changes.

Other avenues of future work focus on the student. First, motivated by the "differing perspectives phenomenon", why do students' perspectives vary depending on what they would prefer in the class? Also, why did only the students who preferred less of an activity differ in their perspectives? The next line of future research continues in the instructional communication framework and stems from the student interviews. Why do students prefer certain strategies over others? It was fairly clear that students preferred an evidence-based discussion, but why did less prefer a student focused discussion and why did they believe the "try and see" method would be ineffective? These are questions in the educational psychology realm of the instructional communication framework. What is it about the evidence-based discussion and the class/discussion format that both students and instructors felt it would be effective? This question has elements of communication, pedagogy, and educational psychology. Answering this question could further improve how an instructor tries to gain buy-in at the start of the semester.

Another avenue of future research pursues my finding that an evidence-based discussion should be an effective method for gaining student agreement. Research should help instructors collect and develop evidence-based materials to present to their students. Then, one can see if these methods do in fact increase student agreement as was suggested by both students and instructors. Research could look into what elements of the evidence-based discussion students respond best to.

Lastly, and what I feel would be the most immediate next step in this research, future work should look into if instructor strategies change what students think is the most effective way of learning physics. This also falls under all three elements of instructional communication research. What is it about what the instructors communicate or how they communicate (communication and pedagogy) that influences what students feel is an effective physics class (educational psychology)? One can potentially gain student agreement by describing the class format and that format coincidentally being what the student had wanted. One can also gain student agreement by changing what the student wanted so that it is now the class format. When I began this research, the latter is what I expected to be occurring, but from the student interviews it is clear that what students believe is the most effective class is not being changed by the discussions of the class format. I could answer this question by measuring the importance placed and time spent on class activities for a student's most effective class before and after the class orientation. If these measures change, I know they would be due to the orientation.

8.5 Closing Remarks

My goal for this research was to provide research-proven strategies to gain student buy-in to a reformed classroom. I believed that if an instructor could gain student buy-in, then they would have an easier time using the research-based instructional strategies used in a reformed classroom.

If these strategies were implemented without resistance, it would lead to greater student performance. I have been able to pinpoint the specific aspects of a studio class and the specific activities that instructors need to work on gaining student agreement with. Through interviews, I have found that using an evidence-based discussion via a class discussion or activity should be an effective way to gain agreement. Once students agree more with the class, they may have more expert-like attitudes about physics and expect to receive a higher grade. If these strategies are successfully implemented and student agreement is increased, students will be happier and instructors will be able to teach in the appropriate reformed manner which will lead to higher performing students, making the studio style successful for both student and instructor.

APPENDIX A:
THE PERCEPTIONS OF UNDERGRADUATES MEASURE OF BUY-IN ACHIEVEMENT (PUMBA)

This survey works for both mobile and non-mobile devices though it is best taken on a non-mobile device.

If you are willing to volunteer for this research, please click “YES” below. Otherwise, if we may not use your survey response for research purposes, but you still want to take the survey, click “NO” below.

O YES

O NO

Select your institution.

O University A

O University B

O University C

Enter your ID. (ex: 1234567)

____

The following are various types of activities for a studio physics class.

For which of the items below did your instructor explain its role for this course, such as how it will work and/or how much time will be spent on it.
Some examples:

- Stating that there will be little lecture.
- Including in the syllabus that most of the class will be group work involving a certain number of students working on problems together.
- Explaining that successful students read before class and come prepared with questions.
- Stating that labs will be mixed in throughout the class and will be done in your small groups.

☐ Lecture

☐ Quizzes or tests

☐ Lab group work

☐ Non-lab group work (e.g. problem solving group work)

☐ Individual work

☐ Student Presentations

☐ Physical demonstrations and simulations

☐ Whole class discussions

☐ Out-of-class work

The activities you selected on the previous screen are listed below. For which activities did your instructor explain why that activity and the time spent on it in this class is important for your learning?

Some examples:

- Providing a research article about the benefits of labs mixed in class.
- Stating that what is expected of you in the class are the same things that will be expected of you in your occupation.
- Explaining that hearing information from your peers makes it easier for you to understand.

☐ Lecture

☐ Quizzes or tests

☐ Lab group work

☐ Non-lab group work (e.g. problem solving group work)

☐ Individual work

☐ Student Presentations

☐ Physical demonstrations and simulations

☐ Whole class discussions

☐ Out-of-class work

You will be asked to estimate the amount of time spent inside and outside of class on the following class actions. Before answering, read through each action and their definitions. If your device has a mouse or track-pad, throughout the survey, you may hover over the text of each action to reveal its definition.

In class actions:

Lecture	The instructor is talking to the whole class. Actions include going over PowerPoint slides or explaining/working through the solution to a physics problem.
Quiz/Test	Formal graded assessment of students' abilities.
Step-by-step labs	Labs in which you are told what to do, what to measure, and what to calculate. You are typically following a procedure and looking for the right result.
Investigative labs	You are given a research question and you have to determine what to do to answer it, what you need to measure, and what you need to calculate.
Other small-group work	Any work that you do with the help of your classmates that has not yet been covered by the other listed aspects. Small-group work should include, but is not limited to, working on worksheets with your group and discussing clicker questions with your peers.
Individual work	Students are working individually without the help of their classmates to complete assigned work.
Student presentation	Students present their work to the entire class. This includes you giving a presentation and you listening to another student presentation.
Viewing or engaging in physical demonstrations or simulations	You are taking part in, or watching others (including the instructor), engage in a brief activity with the purpose of being able to observe a physical phenomenon.
Whole class discussion of scientific concepts	Discussing scientific concepts with the whole class. Students contribute to the majority of the conversation while the instructor may be guiding the conversation.

Out of class actions:

Reading/annotating assigned text	Reading or annotating (adding comments or questions to an online textbook) texts that your instructor has assigned to you.
Reading beyond what is assigned	Reading texts that your instructor has not assigned to you.
Completing assigned conceptual/problem solving homework	Completing conceptual or problem solving homework that your instructor has assigned to you.
Completing conceptual/problem solving work beyond what is assigned.	Completing conceptual or problem solving work that your instructor has not assigned to you. This includes problems from a worksheet, an online source, or problems found in your textbook.
Writing lab reports	Completing a lab report as homework for a lab activity done in class.
Working on group projects	Working on projects with a group out of class. This does not include doing homework, reading, or studying with a group which should be considered in the above out of class aspects.
Using other online resources or viewing instructional videos	Using online resources such as Wikipedia or Hyperphysics to supplement your learning. Viewing any videos with the purpose of gaining knowledge appropriate for class. This includes your instructor's lecture videos, online educational videos (such as Khan Academy), or videos connected to your textbook.
Getting help from your instructor/TA/tutor	Getting additional help from someone else, not including classmates. This includes, but is not limited to, attending office hours, going to the physics help room, and going to a tutor.

O I have carefully read through each action.

O I have not carefully read through each action.

Throughout this survey you will be asked to respond based on your most effective physics class. Your most effective physics class is a hypothetical class that is designed so that it is the best class possible for teaching you physics.

Carefully read through every prompt before responding. **You can hover over the text of each response to reveal its definition.**

For each of the following in-class actions, indicate whether that in-class action would be less/equally/more important in your most effective class compared to the importance it seems it will have in your actual class. If an action seems that it will not occur in your actual class, consider it to have no importance in your actual class (i.e. your response should be either "Equal..." or "More...")..

Do not confuse importance with time. Something could seem important to the class but not occur very often.

"In my most effective class, *compared to my actual class*, this in-class action would be..."

	Less important than my actual class.	Equally important as my actual class.	More important than my actual class.
Lecture	O	O	O
Quiz	O	O	O
Step-by-step labs	O	O	O
Investigative labs	O	O	O
Other small-group work	O	O	O
Individual work	O	O	O
Student presentation	O	O	O
Viewing or engaging in physical demonstrations or simulations	O	O	O
Whole class discussion of scientific concepts	O	O	O

For each of the following out-of-class actions, indicate whether that out-of-class action would be less/equally/more important for your most effective class compared to the importance it seems it will have in your actual class. If an out-of-class action is not something done for your class (e.g. you may not have group projects in your class), consider it to have no importance in your actual class (i.e. your response should be either "Equal..." or "More...")..

Do not confuse importance with time. Something could seem important to the class but not ccur very often.

"For my most effective class, *compared to my actual class*, this out-of-class action would be..."

	Less important than my actual class.	Equally important as my actual class.	More important than my actual class.
Reading/annotating assigned text	O	O	O
Reading beyond what is assigned	O	O	O
Completing assigned conceptual/problem solving homework	O	O	O
Completing conceptual/problem solving work beyond what is assigned	O	O	O
This is an attention check. Click "Equally important".	O	O	O
Writing lab reports	O	O	O
Working on group projects	O	O	O
Using other online resources or viewing instructional videos	O	O	O
Getting help from instructor/TAs/tutor (office hours, physics help room, etc.)	O	O	O

I

In an average week of your actual studio physics class, what percentage of time will be spent on each of the following actions in the classroom?
In an average week of your most effective physics class, what percentage of time would be spent on each of the following actions in the classroom?

Use the "Total" at the bottom to make sure your percentages sum to 100%.

Do not confuse importance with time. Something could seem important to the class but not occur very often.

	Percentage of time that will be spent **in the classroom** for your **actual** class	Percentage of time that would be spent **in the classroom** for your **most effective** class
Lecture		
Quiz		
Step-by-step labs		
Investigative labs		
Other small-group work		
Individual work		
Student presentation		
Viewing or engaging in physical demonstrations or simulations		
Whole class discussion of scientific concepts		
Total		

For each of the following out-of-class actions, would you spend a less/equal/more amount of time on this action for your most effective class compared to the time you feel your actual class intends for you to spend on that action? If your actual class does not intend for you to spend any time on an action, consider it to have no time spent on that action in your actual class (i.e. your response should be either "Equal..." or "More...").

Do not confuse importance with time. Something could seem important to the class but not occur very often.

"For my most effective class, *compared to my actual class*, this out-of-class action would spend..."

	Less time than my actual class intends.	Equal amount of time as my actual class intends.	More time than my actual class intends.	I have no feeling as to how much time my actual class intends for me to spend on this action.
Reading/annotating assigned text	O	O	O	O
Reading beyond what is assigned	O	O	O	O
Completing assigned conceptual/problem solving homework	O	O	O	O
Completing conceptual/problem solving work beyond what is assigned	O	O	O	O
This is an attention check. Click "More time".	O	O	O	O
Writing lab reports	O	O	O	O
Working on group projects	O	O	O	O
Using other online resources or viewing instructional videos	O	O	O	O
Getting help from instructor/TAs/tutor (office hours, physics help room, etc.)	O	O	O	O

To which gender identity do you most identify?

- O Man
- O Woman
- O Transgender Man
- O Transgender Woman
- O Gender Variant/Non-Conforming
- O Not Listed ____
- O Prefer Not to Answer

Which of the following best represents your racial heritage? (Check all that apply.)

- ☐ American Indian or Alaska Native
- ☐ Asian
- ☐ Black or African American
- ☐ Native Hawaiian or Other Pacific Islander
- ☐ White
- ☐ Prefer Not to Answer

Are you of Hispanic or Latino/a origin?

O Yes

O No

O Prefer Not to Answer

What is your age?

____

Have you taken a studio style class before this semester?

O Yes

O No

Which of the following are you currently classified as?

O Freshman

O Sophomore

O Junior

O Senior or higher

APPENDIX B:
PUMBA VALIDITY INTERVIEW PROTOCOL

Ice Breaker:
Why are you taking this physics course this semester?

I'm going to ask you about your thoughts on your current studio physics course. When responding, please think about both what you do in class and what you do out of class.

1. Before the semester began, what did you expect this course was going to be like?

    a) What were those expectations based on?

2. How is taking a studio course different than taking a traditional or lecture type course?

    b) How has taking the studio course been different than you expected?

3. What do you like about the studio course?

4. What do you not like about the studio course?

5. Some people would say all physics courses should be taught in studio mode. How would you respond?

6. What advice would you have for a student planning to take this course next semester?

7. How much would you say you agree with the format of the class?

Now, I am going to ask you about the survey you took at the beginning of the semester that asked you about the time you expect to take on several actions inside and out of class. It also asked you to rank the importance of these actions and do these same things for a class that is the most effective at teaching you physics. Please look over this copy of the survey to refresh your memory.

8. Explain in your own words what each question is asking you.

9. Were there any parts of the survey that were confusing or that you did not understand?

10. Were there any activities (lecture, student presentation, etc.) that you do in class that are not represented on the survey?

11. Were there any activities that you found difficult to answer? Why?

12. What do you interpret the different classroom activities to include? Were there any activities that you think overlap with other activities?

13. To respond to several of the questions you need to have some idea of the importance of various activities for your actual class. What were your ideas about this based on?

14. Do you feel you could have responded with the time spent on the out-of-class actions, similar to how you were asked to respond with the in-class actions?

15. Explain one or two of the things your instructor did to explain the role of those items.

16. Can you explain your response for any major differences between expected and most effective?

17. Can you explain your response for any (nearly) identical answers between expected and most effective?

APPENDIX C:
THE SURVEY OF INSTRUCTOR METHODS FOR BUY-IN ACHIEVEMENT (SIMBA)

IRB Consent Form

Title of Project: Student views and expectations for reformed instruction physics classes

Principal Investigator: Jacquelyn Chini

Other Investigators: Matthew Wilcox

You are being invited to take part in a research study. Whether you take part is up to you.

The purpose of this study is to investigate students' views and expectations for reformed physics classrooms such as the SCALE-UP physics class, as well as investigating how those views and expectations are formed. The ultimate goal is to provide a better means of gaining students' understanding and agreement of the class design.

Participation in this research involves taking this survey. The survey should take approximately 15-20 minutes to complete. Your decision to participate in the research is voluntary and at any time you may withdraw your consent to participate. Your responses will be stripped of any identifying information before being shared outside of the research group.

Study contact for questions about the study or to report a problem: If you have questions, concerns, or complaints contact Dr. Jackie Chini, Assistant Professor, University of Central Florida, Department of Physics, (407) 823-3607 or jchini@ucf.edu. You may also contact Matthew Wilcox at mwilcox1@knights.ucf.edu.

IRB contact about your rights in the study or to report a complaint: Research at the University of Central Florida involving human participants is carried out under the oversight of the Institutional Review Board (UCF IRB). This research has been reviewed and approved by the IRB. For information about the rights of people who take part in research, please contact: Institutional Review Board, University of Central Florida, Office of Research & Commercialization, 12201 Research Parkway, Suite 501, Orlando, FL 32826-3246 or by telephone at (407) 823-2901.

If you are willing to volunteer for this research, please click "YES" below. Otherwise, click "NO" below.

- O Yes
- O No

Enter your name. Your name will <u>only</u> be used to connect your responses to your students' responses on their survey. We will never report your responses with your name.

____

Your thought process throughout the survey should be as follows: *Am I discussing with the students how much of a class activity they should expect, how it will work, and/or how I expect them to behave during that activity?* (Yes or No). If yes, *what am I doing, if anything, to get them to agree that adhering to this expectation is the best way to learn?* Finally, *through which format(s) am I having this discussion about those expectations*?

The definitions and examples for all possible answer choices are found here. Note that there are 3 tabs with definitions. Also, in the survey, **hovering over the <u>text</u> of each topic or format will reveal more information about it.**

A **sample** question is shown below. Responding "Yes" on the first question reveals the two follow up questions shown.

Did you discuss **expectations about lab group work** in order to set students' expectations and understandings?

Yes

No

What did you discuss to try to get students to understand/agree with the lab group work expectations for the class? (Choose all that apply)

Advantages of engagement

Comparison between reformed and traditional classes

Discussion about students

Evidence based discussion

Extended example

"Try it and you will see."

Other; Please briefly describe

Through which format(s) did you discuss the expectations for and/or why students should agree with this topic? (Choose at least one)

Lecture

Class discussion/activity

Reading/syllabus

Audio/Video

Reminders throughout the semester

Rewarding appropriate behavior

An "expectation" refers to a type of activity for class that you are talking to the students about (e.g. lecture or labs). "Discussing an expectation" is defined as **discussing how much of it there will be, how it will typically operate, and/or how students are expected to behave for that expectation**.
Note: Only discussing how an activity/expectation in class will be graded is not sufficient to be considered as discussing that activity/expectation for the purposes of this survey.

Hovering your mouse pointer over the <u>text</u> of an expectation or option will reveal more information about it.

Did you discuss **expectations about lecture** in order to set students' expectations and understandings?

O Yes

O No

What did you discuss to try to get students to understand/agree with the lecture expectations for the class? (Choose all that apply)

- ☐ Advantages of engagement
- ☐ Comparison between reformed and traditional classes
- ☐ Discussion about students
- ☐ Evidence based discussion
- ☐ Extended example
- ☐ "Try it and you will see."
- ☐ Other; Please briefly describe

____

Through which format(s) did you discuss the expectations for and/or why students should agree with this topic? (Choose at least one)

- ☐ Lecture
- ☐ Class discussion/activity
- ☐ Reading/syllabus
- ☐ Audio/Video
- ☐ Reminders throughout the semester
- ☐ Rewarding appropriate behavior

Did you discuss **expectations about quizzes or tests** in order to set students' expectations and understandings?

- O Yes
- O No

Did you discuss **expectations about lab group work** in order to set students' expectations and understandings?

- O Yes
- O No

Did you discuss **expectations about non-lab group work (e.g. problem solving group work)** in order to set students' expectations and understandings?

- O Yes
- O No

Did you discuss **expectations about individual work** in order to set students' expectations and understandings?

- O Yes
- O No

Did you discuss **expectations about student presentations** in order to set students' expectations and understandings?

- O Yes
- O No

Did you discuss **expectations about physical demonstrations and simulations** in order to set students' expectations and understandings?

- O Yes
- O No

Did you discuss **expectations about whole class discussions** in order to set students' expectations and understandings?

- O Yes
- O No

Did you discuss **expectations about out-of-class work** in order to set students' expectations and understandings?

- O Yes
- O No

Did you discuss **other expectations about the class format** in order to set students' expectations and understandings?

- O Yes
- O No

Use the definitions below to respond to the following question.

In class actions:

Lecture	The instructor is talking to the whole class. Actions include going over PowerPoint slides or explaining/working through the solution to a physics problem.
Quiz/Test	Formal graded assessment of students' abilities.
Step-by-step labs	Labs in which you are told what to do, what to measure, and what to calculate. You are typically following a procedure and looking for the right result.
Investigative labs	You are given a research question and you have to determine what to do to answer it, what you need to measure, and what you need to calculate.
Other small-group work	Any work that you do with the help of your classmates that has not yet been covered by the other listed aspects. Small-group work should include, but is not limited to, working on worksheets with your group and discussing clicker questions with your peers.
Individual work	Students are working individually without the help of their classmates to complete assigned work.
Student presentation	Students present their work to the entire class. This includes you giving a presentation and you listening to another student presentation.
Viewing or engaging in physical demonstrations or simulations	You are taking part in, or watching others (including the instructor), engage in a brief activity with the purpose of being able to observe a physical phenomenon.
Whole class discussion of scientific concepts	Discussing scientific concepts with the whole class. Students contribute to the majority of the conversation while the instructor may be guiding the conversation.

In an average week of your studio physics class, what percentage of time do you **intend** to spend on each of the following actions in the classroom? Remember, you can hover over the text of each option for more information.

Use the "Total" at the bottom to make sure your percentages sum to 100%.

Do not confuse importance with time. Something could seem important to the class but not occur very often.

	Percentage of time that you **intend** to spend **in the classroom**.
Lecture	
Quiz	
Step-by-step labs	
Investigative labs	
Other small-group work	
Individual work	
Student presentation	
Viewing or engaging in physical demonstrations or simulations	
Whole class discussion of scientific concepts	
Total	

Which of the following will you be able to provide for my research? (Choose all that apply)
Note: I can accept raw responses and score them myself so do not worry about going through the trouble of scoring them if you do not have to.

- ☐ Concept test pre/post-scores (FCI, FMCE, CSEM, etc.)
- ☐ Attitudinal test pre/post-scores (CLASS)
- ☐ Concept test pre/post-scores with identified data (So I may match a student's scores with their survey responses)
- ☐ Attitudinal test pre/post-scores with identified data (so I may match a student's scores with their survey responses)
- ☐ Final student grades

To which gender identity do you most identify?

- O Man
- O Woman
- O Transgender Man
- O Transgender Woman
- O Gender Variant/Non-Conforming
- O Not Listed ____
- O Prefer Not to Answer

Which of the following best represents your racial heritage? (Check all that apply.)

- ☐ American Indian or Alaska Native
- ☐ Asian
- ☐ Black or African American
- ☐ Native Hawaiian or Other Pacific Islander
- ☐ White
- ☐ Prefer Not to Answer

Are you of Hispanic, Latino/a or of Spanish origin?

- O Yes
- O No
- O Prefer Not to Answer

What is your age?

____

Prior to this semester, how many semesters have you taught in either a studio style or a largely reformed instruction style.

____

APPENDIX D:
SIMBA VALIDITY INTERVIEW PROTOCOL

I am going to ask you several questions about the interview you took regarding the things you did to discuss the format of the course to your students in your studio class. Please look over the survey and your responses to refresh your memory.

1. Explain in your own words what each question is asking you.
2. Were there any parts of the survey that were confusing or that you did not understand?
3. What do you interpret the different classroom activities to include? Were there any activities that you think overlap with other activities?
4. Were there any activities (lecture, student presentation, etc.) that you do in class that are not represented on the survey?
5. What do you interpret the different agreement methods to include?
6. Are there any agreement methods that you feel are missing from the survey or that overlap with others?
7. What do you interpret the different formats to include?
8. Are there any formats that you feel are missing from the survey or that overlap with others?
9. Were there any response options that you found difficult to answer? Why?
10. Explain some of the ways in which you discussed the course format and how that led to the way you responded on the survey.

APPENDIX E:
INSTRUCTOR POST INTERVIEW PROTOCOL

My first set of questions will be in regard to your responses on the survey. The purpose of this portion of the interview is to ensure you are understanding the survey correctly and that I am interpreting your responses correctly.

1. Explain in your own words what each question is asking you.
2. Can you describe what might be considered as each of the response options?
3. Explain some of the ways in which you discussed the course format and how that led to the way you responded on the survey.

For the remainder of the interview, I will be asking you about any student resistance you may experience, the ways you may try to get students to agree with the course format, and your students' agreement to certain class activities.

4. When you initially started teaching a studio course, did you expect student resistance to the course format? If so, what did you expect?
5. Did these expectations change in newer semesters as a studio instructor? If so, why did they change?
6. What kinds of student resistance have you experienced in your studio class?
7. Does this resistance change throughout the semester? Does it decrease or increase?
8. What types of things do you do to try to get students to agree to the course format? (Interviewer can refer to survey responses here)
9. How do the students respond to these methods? Is it successful?
10. Students seem to agree (less/more) with (some activity). Do you have any idea why that might be? Did you do something to change their opinion of it?
11. With the methods that you used to get student agreement with the course format, did you get the idea for those methods from another source or did you come up with it yourself?
12. Did your instruction change over the course of the semester? Did it change from what you originally planned it would be? If so, what is the cause of the change?
13. Your students indicated (some activity) occurred (more/less) often. Any idea why?
14. Do you have any questions for me?

APPENDIX F:
STUDENT POST INTERVIEW PROTOCOL

I am going to ask you several questions about your experiences and opinions of your introductory studio physics course. I will also be asking you several questions about your instructor and what he or she did to introduce you to the course format.

1. Think back to the first week of class. At that point in time and from what the instructor explained, what did you think the class would be like?
2. At the start of the semester, what did the instructor do to explain to you the class format?
3. Did your instructor do anything to get you to agree with the format at the start of the semester? Did this help you agree with the class format?
4. Did the class end up being different than what you were initially told to expect? How so?
5. What class activities do you agree or disagree with?
6. Are there any examples of activities that you initially did not agree with but agreed with at the end? What changed?
7. Are there any examples of activities that you initially agreed with but did not agree with at the end? What changed?
8. Throughout the semester, did your instructor do anything to get you to agree with the class format?
9. Did you instructor's actions affect your opinion of the class format?
10. What could the instructor do to help you agree with the class format?
11. *After student responds with their own ideas. Referring to SIMBA response options* Which of these methods and formats would work best in convincing you to buy-in to the class?
12. At the start of the semester, did you feel that your instructor was competent? Caring? Had good character?
13. Did this change over time?
14. What is more likely to convince you of something, a logical argument (facts, research evidence) or an argument appealing to you personally (your experiences, your feelings, etc.)?

APPENDIX G:
STUDENT AND INSTRUCTOR RESPONSES FOR POTENTIALLY EFFECTIVE STRATEGIES FOR GAINING BUY-IN

Table 47. An outline of the potential influences on student agreement.

Instructor	Student	Potential Effective Methods	Potential Ineffective Methods	Potential Effective Formats	Potential Ineffective Formats
1	1	M-EVIDENCE, M-ENGAGE, M-TRY, TRUST	M-COMPARISON	F-LECTURE, F-READ, TRUST	F-AV, F-DISCUSSION
1	2*	-	-	-	-
2	3	M-EVIDENCE, M-COMPARISON, M-EXAMPLE	M-TRY	F-LECTURE, F-READ, F-REMIND	F-AV, F-DISCUSSION
2	4	M-STUDENT, M-COMPARISON, M-EXAMPLE	-	F-DISCUSSION, F-LECTURE, F-READ, F-REWARD	-
4	5	M-EVIDENCE+ M-STUDENT, M-ENGAGE, M-COMPARISON, M-TRY	M-STUDENT	F-DISCUSSION, F-REMIND, F-REWARD	-
4	6	M-STUDENT, M-ENGAGE, TRUST	M-TRY	F-DISCUSSION, F-REMIND, F-REWARD	F-AV, F-READ
7	7	M-EVIDENCE, M-COMPARISON	M-TRY	F-DISCUSSION, F-REMIND, F-REWARD	F-AV
7	8	M-EVIDENCE, M-STUDENT, M-ENGAGE	-	F-DISCUSSION, F-REWARD	F-READ
8	9	M-EVIDENCE, M-COMPARISON	M-TRY, M-EXAMPLE	F-DISCUSSION, F-LECTURE, F-READ	F-AV

*Due to time constraints, this portion of the interview did not occur with student 2.

Table 48. Instructor responses to effective and ineffective methods and formats.

Instructor	Effective formats	Ineffective Formats	Effective methods	Ineffective methods
1	Class Discussion/ Activity	-	Evidence-based Discussion	"Comparisons", Advantages of Engagement
2	Lecture Format, Audio/Video	Reading/Syllabus, "Reminders", "Reward"	Evidence-based Discussion, Student Focused Discussion	Advantages of Engagement
4	-	-	-	-
7	Lecture Format	Reading/Syllabus	Evidence-based Discussion	"Try and See"
8	"Reminders", "Reward"	Lecture Format	Evidence-based Discussion	-

APPENDIX H:
INTERNAL REVIEW BOARD (IRB) PERMISSION LETTERS

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Jacquelyn J Chini**

Date: **February 04, 2018**

Dear Researcher:

On 02/04/2018 the IRB approved the following modifications to human participant research until 11/14/2018 inclusive:

Type of Review:	IRB Addendum and Modification Request Form Expedited Review Category #6 & 7
Modification Type:	Added study personnel: Erin Scanlon, Elijah Ibadlit, Brian Zamarripa, Constance Doty, and Alyssa Johnson; added interview protocol; uploaded revised protocol and consent forms.
Project Title:	Examining algebra-based SCALE-UP at many universities
Investigator:	Jacquelyn J Chini
IRB Number:	SBE-15-11623
Funding Agency:	National Science Foundation
Grant Title:	Collaborative Research: Investigating Institutional Success at Overcoming Challenges in Algebra-based Studio Physics
Research ID:	1055924

The scientific merit of the research was considered during the IRB review. The Continuing Review Application must be submitted 30days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

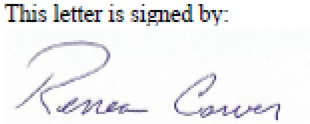
If continuing review approval is not granted before the expiration date of 11/14/2018, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a signed and dated copy of the consent form(s).

All data, including signed consent forms if applicable, must be retained and secured per protocol for a minimum of five years (six if HIPAA applies) past the completion of this research. Any links to the identification of participants should be maintained and secured per protocol. Additional requirements may be imposed by your funding agency, your department, or other entities. Access to data is limited to authorized individuals listed as key study personnel.

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

This letter is signed by:



Signature applied by Renea C Carver on 02/04/2018 03:27:34 PM EST

Designated Reviewer

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Exempt Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Jacquelyn J. Chini** and Co-PIs: **Jarrad W. Pond & Matthew Wilcox**

Date: **July 27, 2016**

Dear Researcher:

On 07/27/2016, the IRB approved the following activity as human participant research that is exempt from regulation:

Type of Review:	Exempt Determination
Modification Type:	Addition of a new interview, removal of Shima Gholam and addition of Matthew Wilcox. A new recruitment and revised protocol was uploaded and a new consent was approved for use.
Project Title:	Collaborative Research: Investigating Institutional Success at Overcoming Challenges in Algebra-based Studio Physics
Investigator:	Jacquelyn J. Chini
IRB Number:	SBE-13-09567
Funding Agency:	National Science Foundation
Grant Title:	Collaborative Research: Investigating Institutional Success at Overcoming Challenges in Algebra-based Studio Physics
Research ID:	1347510

This determination applies only to the activities described in the IRB submission and does not apply should any changes be made. If changes are made and there are questions about whether these changes affect the exempt status of the human research, please contact the IRB. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

Kamille Chap

Signature applied by Kamille Chaparro on 07/27/2016 09:16:49 AM EDT

IRB Coordinator

REFERENCES

Adams, W. K., Perkins, K. K., Podolefsky, N. S., Dubson, M., Finkelstein, N. D., & Wieman, C. E. (2006). New instrument for measuring student beliefs about physics and learning physics: The Colorado Learning Attitudes about Science Survey. *Physical Review Special Topics-Physics Education Research, 2*(1), 010101.

AERA, APA, & NCME. (1999). *Standards for educational and psychological testing*. Washington, DC.

Agresti, A. (1996). *An introduction to categorical data analysis* (Vol. 135): Wiley New York.

Andrews, T. C., & Lemons, P. P. (2015). It's personal: biology instructors prioritize personal evidence over empirical evidence in teaching decisions. *CBE-Life Sciences Education, 14*(1), ar7.

Appleton-Knapp, S. L., & Krentler, K. A. (2006). Measuring student expectations and their effects on satisfaction: The importance of managing student expectations. *Journal of marketing education, 28*(3), 254-264.

Aristotle, O. (1991). Rhetoric: A Theory of Civic Discourse, trans. George A. Kennedy. *New York: Oxford University Press, 1*, 3-5.

Bagozzi, R. P., & Fornell, C. (1982). Theoretical concepts, measurements, and meaning. *A second generation of multivariate analysis, 2*(2), 5-23.

Barthelemy, R. S., Hedberg, G., Greenberg, A., & McKay, T. (2015). The climate experiences of students in introductory Biology. *Journal of microbiology & biology education, 16*(2), 138.

Bautista, N. U., Schussler, E. E., & Rybczynski, S. M. (2014). Instructional experiences of graduate assistants implementing explicit and reflective introductory biology laboratories. *International Journal of Science Education, 36*(7), 1184-1209.

Beichner, R. J. (2011). SCALE-UP FAQ. Retrieved from http://scaleup.ncsu.edu/FAQs.html

Beichner, R. J., Bernold, L., Burniston, E., Dail, P., Felder, R., Gastineau, J., . . . Risley, J. (1999). Case study of the physics component of an integrated curriculum. *American journal of physics, 67*(S1), S16-S24.

Beichner, R. J., Saul, J. M., Abbott, D. S., Morse, J. J., Deardorff, D., Allain, R. J., . . . Risley, J. S. (2007). The student-centered activities for large enrollment undergraduate programs (SCALE-UP) project. *Research-based reform of university physics, 1*(1), 2-39.

Beichner, R. J., Saul, J. M., Allain, R. J., Deardorff, D. L., & Abbott, D. S. (2000). Introduction to SCALE-UP: Student-Centered Activities for Large Enrollment University Physics.

Belcher, J. W. (2003). Improving student understanding with TEAL. *Faculty Newsletters, 16*(8).

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives, handbook I: The cognitive domain* (Vol. 19): New York: David McKay Co Inc.

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological bulletin, 110*(2), 305.

Bowerman, B. L., & O'connell, R. T. (1990). *Linear statistical models: An applied approach*: Brooks/Cole.

Brennan, R. L. (2010). Generalizability theory and classical test theory. *Applied Measurement in Education, 24*(1), 1-21.

Breslow, L. (2010). Wrestling with pedagogical change: the TEAL initiative at MIT. *Change: The Magazine of Higher Learning, 42*(5), 23-29.

Burnard, P. (1991). A method of analysing interview transcripts in qualitative research. *Nurse education today, 11*(6), 461-466.

Chasteen, S. (2017). How do I help students engage productively in active learning classrooms?

Chesebro, J. L. (2002). Teaching clearly. *Communication for teachers*, 93-103.

Chini, J. J., Gaffney, J. D., & Al-Rawi, A. (2013). *Expectancy Violation in Traditional and Studio-mode Introductory Physics Courses.* Paper presented at the Physics Education Research Conference, Portland, OR.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement, 20*(1), 37-46.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences . Hilsdale. *NJ: Lawrence Earlbaum Associates, 2*.

Cohen, J. (1992). A power primer. *Psychological bulletin, 112*(1), 155.

Conley, N. A., & Ah Yun, K. (2017). A survey of instructional communication: 15 years of research in review. *Communication Education, 66*(4), 451-466.

Cooper, K. M., Ashley, M., & Brownell, S. E. (2017). Using Expectancy Value Theory as a Framework to Reduce Student Resistance to Active Learning: A Proof of Concept. *Journal of microbiology & biology education, 18*(2).

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*: ERIC.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *psychometrika, 16*(3), 297-334.

Dancy, M. H., & Henderson, C. (2005). *Beyond the individual instructor: Systemic constraints in the implementation of research-informed practices.* Paper presented at the AIP Conference Proceedings.

Dancy, M. H., & Henderson, C. (2010). Pedagogical practices and instructional change of physics faculty. *American journal of physics, 78*(10), 1056-1063.

DeBeck, G., Settelmeyer, S., Li, S., & Demaree, D. (2010). *TA Beliefs in a SCALE-UP Style Classroom.* Paper presented at the 2010 PHYSICS EDUCATION RESEARCH CONFERENCE.

DeMonbrun, M., Finelli, C. J., Prince, M., Borrego, M., Shekhar, P., Henderson, C., & Waters, C. (2017). Creating an instrument to measure student response to instructional practices. *Journal of Engineering Education, 106*(2), 273-298.

Diamantopoulos, A., & Winklhofer, H. M. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing research, 38*(2), 269-277.

Dori, Y. J., Belcher, J., Bessette, M., Danziger, M., McKinney, A., & Hult, E. (2003). Technology for active learning. *Materials Today, 6*(12), 44-49.

Eccles, J. S. (1983). Expectancies, values, and academic behaviors.

Elby, A., Scherr, R., Goertzen, R., & Conlin, L. (2008). Open-Source Tutorials in Physics Sense Making.

Ene, M., Leighton, E. A., Blue, G. L., & Bell, B. A. (2015). *Multilevel models for categorical data using SAS® PROC GLIMMIX: the basics.* Paper presented at the SAS Global Forum 2015 Proceedings.

England, B. J., Brigati, J. R., & Schussler, E. E. (2017). Student anxiety in introductory biology classrooms: Perceptions about active learning and persistence in the major. *PloS one, 12*(8), e0182506.

Etkina, E., & Heuvelen, A. (2001). Investigative Science Learning Environment: Using the processes of science and cognitive strategies to learn physics.

Etkina, E., & Van Heuvelen, A. (2007). Investigative science learning environment–A science process approach to learning physics. *Research-based reform of university physics, 1*.

Feldman, K. A. (1977). Consistency and variability among college students in rating their teachers and courses: A review and analysis. *Research in Higher Education, 6*(3), 223-274.

Feldman, K. A. (1993). College students' views of male and female college teachers: Part II—Evidence from students' evaluations of their classroom teachers. *Research in Higher Education, 34*(2), 151-211.

Ferguson, C. J. (2009). An effect size primer: a guide for clinicians and researchers. *Professional Psychology: Research and Practice, 40*(5), 532.

Field, A. (2009). *Discovering statistics using SPSS*: Sage publications.

Foote, K. T., Neumeyer, X., Henderson, C., Dancy, M. H., & Beichner, R. J. (2014). Diffusion of research-based instructional strategies: the case of SCALE-UP. *International Journal of STEM Education, 1*(1), 1.

Fornell, C., & Bookstein, F. L. (1982). Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. *Journal of Marketing research*, 440-452.

Framing the Interactive Engagement Classroom, Boulder: Science Education Initiative at the University of Colorado, Boulder. (2013). Retrieved from http://www.colorado.edu/sei/fac-resources/framing.html

Froyd, J. E., Borrego, M., Cutler, S., Henderson, C., & Prince, M. J. (2013). Estimates of use of research-based instructional strategies in core electrical or computer engineering courses. *Education, IEEE Transactions on, 56*(4), 393-399.

Frymier, A. B., & Houser, M. L. (1998). Does making content relevant make a difference in learning? *Communication Research Reports, 15*(2), 121-129.

Frymier, A. B., & Shulman, G. M. (1995). “What's in it for me?”: Increasing content relevance to enhance students' motivation. *Communication Education, 44*(1), 40-50.

Frymier, A. B., Shulman, G. M., & Houser, M. (1996). The development of a learner empowerment measure 1. *Communication Education, 45*(3), 181-199.

Frymier, A. B., & Thompson, C. A. (1992). Perceived teacher affinity-seeking in relation to perceived teacher credibility. *Communication Education, 41*(4), 388-399.

Gaffney, J. D., Gaffney, A. L. H., & Beichner, R. J. (2010). Do they see it coming? Using expectancy violation to gauge the success of pedagogical reforms. *Physical Review Special Topics-Physics Education Research, 6*(1), 010102.

Gaffney, J. D., Gaffney, A. L. H., & Chini, J. J. (2013, 2013). *Using Expectancy Violation To Investigate Student Dissatisfaction In Studio Physics.* Paper presented at the Physics Education Research Conference, Portland, OR.

Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The qualitative report, 8*(4), 597-606.

Graneheim, U. H., & Lundman, B. (2004). Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. *Nurse education today, 24*(2), 105-112.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American journal of physics, 66*(1), 64-74.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in quantitative methods for psychology, 8*(1), 23.

Heilman, J. D., & Armentrout, W. (1936). The rating of college teachers on ten traits by their students. *Journal of Educational Psychology, 27*(3), 197.

Heller, K., & Heller, K. (2001). *Cooperative group problem solving in physics*: Brooks/Cole Publishing Company.

Heller, P., & Hollabaugh, M. (1992). Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups. *American Journal of Physics, 60*(7), 637-644.

Heller, P., Keith, R., & Anderson, S. (1997). Teaching problem solving through cooperative grouping (Part 1): Group versus individual problem solving. *MAA NOTES*, 159-172.

Henderson, C., & Dancy, M. H. (2007). Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics. *Physical Review Special Topics-Physics Education Research*, *3*(2), 020102.

Henderson, C., & Dancy, M. H. (2008). Physics faculty and educational researchers: Divergent expectations as barriers to the diffusion of innovations. *American journal of physics, 76*(1), 79-91.

Henderson, C., Dancy, M. H., & Niewiadomska-Bugaj, M. (2012). Use of research-based instructional strategies in introductory physics: Where do faculty leave the innovation-decision process? *Physical Review Special Topics-Physics Education Research, 8*(2), 020104.

Hestenes, D., & Halloun, I. (1995). Interpreting the force concept inventory: A response to March 1995 critique by Huffman and Heller. *The physics teacher, 33*(8), 502-502.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The physics teacher, 30*(3), 141-158.

Hogan, T. P., Benjamin, A., & Brezinski, K. L. (2000). Reliability methods: A note on the frequency of use of various types. *Educational and psychological measurement, 60*(4), 523-531.

Hora, M. T., Bouwma-Gearhart, J., & Park, H. J. (2014). *Exploring data-driven decision-making in the field: How faculty use data and other forms of information to guide instructional decision-making*. Retrieved from

Krathwohl, D. R. (1964). *Taxonomy of educational objectives: The classification of educational goals* (Vol. 2): Longmans, Green.

Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). Handbook II: affective domain. *New York: David McKay*.

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *psychometrika, 2*(3), 151-160.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159-174.

Laws, P. W. (1991). Calculus-based physics without lectures. *Physics today, 44*(12), 24-31.

Laws, P. W., Sokoloff, D., & Thornton, R. (1999). Promoting active learning using the results of physics education research. *UniServe Science News, 13*, 14-19.

Madsen, A., McKagan, S. B., & Sayre, E. C. (2015). How physics instruction impacts students' beliefs about learning physics: A meta-analysis of 24 studies. *Physical Review Special Topics-Physics Education Research, 11*(1), 010115.

Maloney, D. P., O'Kuma, T. L., Hieggelke, C. J., & Van Heuvelen, A. (2001). Surveying students' conceptual knowledge of electricity and magnetism. *American journal of physics, 69*(S1), S12-S23.

Martin, M. M., Chesebro, J. L., & Mottet, T. P. (1997). Students' perceptions of instructors' socio-communicative style and the influence on instructor credibility and situational motivation. *Communication Research Reports, 14*(4), 431-440.

Mazur, E. (1999). Peer instruction: A user's manual: AAPT.

Mazur, E., King, G., Lukoff, B., & Miller, K. (Producer). Perusall. Retrieved from www.perusall.com

Mazur, E., & Somers, M. (1999). Peer instruction: A user's manual. *American journal of physics, 67*(4), 359-400.

McClure, L. S., Combrink, T. S., Foor, C. E., Walden, S. E., & Trytten, D. A. (2006). *I WISH SOMEONE WOULD'VE TOLD ME: UNDERGRADUATE ENGINEERING STUDENTS OFFER ADVICE TO INCOMING STUDENTS.* Paper presented at the Proceedings of the 2006 American Society for Engineering Education Annual Conference and Exposition.

McCroskey, J. C. (1968). *An introduction to rhetorical communication: The theory and practice of public speaking*: Prentice-Hall.

McCroskey, J. C., & Teven, J. J. (1999). Goodwill: A reexamination of the construct and its measurement. *Communications Monographs, 66*(1), 90-103.

McCroskey, L., Richmond, V., & McCroskey, J. (2002). The scholarship of teaching and learning: Contributions from the discipline of communication. *Communication Education, 51*(4), 383-391.

McDermott, L. C., & Shaffer, P. S. (1998). *Tutorials in introductory physics*: Prentice Hall.

McPherson, M. B., Kearney, P., & Plax, T. G. (2003). The dark side of instruction: Teacher anger as classroom norm violations. *Journal of Applied Communication Research, 31*(1), 76-90.

Messick, S. (1989). Validity.

Mottet, T. P., Richmond, V. P., & McCroskey, J. C. (Eds.). (2006). *Handbook of instructional communication: Rhetorical and relational perspectives*: Allyn & Bacon.

Myers, R. H. (1990). *Classical and modern regression with applications*. Retrieved from

Nave, C. R. (2000). *Doing It by the Numbers: JavaScript Calculations in Web-Based Physics Instructional Material.* Paper presented at the 121st AAPT National Meeting, Guelph, Ontario, Canada.

Nguyen, K., Borrego, M. J., Finelli, C., Shekhar, P., DeMonbron, M., Hendersen, C., & Waters, C. (2016). *Measuring Student Response to Instructional Practices (StRIP) in Traditional and Active Classrooms.* Paper presented at the ASEE Annual Conference and Exposition, New Orleans, Louisiana.

Nguyen, K., Husman, J., Borrego, M., Shekhar, P., Prince, M., Demonbrun, M., . . . Waters, C. (2017). Students' Expectations, Types of Instruction, and Instructor Strategies Predicting Student Response to Active Learning. *International Journal of Engineering Education, 33*(1), 2-18.

Pinder-Grover, T. A. (2013). Active Learning in Engineering: Perspectives from Graduate Student In-structors.

Popham, W. J. (2009). Assessing Student Affect. *Educational Leadership, 66*(8), 85-86.

Potter, W., Webb, D., Paul, C., West, E., Bowen, M., Weiss, B., . . . De Leone, C. (2014). Sixteen years of collaborative learning through active sense-making in physics (CLASP) at UC Davis. *American journal of physics, 82*(2), 153-163.

Potvin, G., Hazari, Z., Tai, R. H., & Sadler, P. M. (2009). Unraveling bias from student evaluations of their high school science teachers. *Science Education, 93*(5), 827-845.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1): Sage.

Redish, E. F. (2009). 121-122 Facilitating in Tutorial. Retrieved from http://umdperg.pbworks.com/w/page/10511167/121-122%20Facilitating%20in%20Tutorial

Redish, E. F., Saul, J. M., & Steinberg, R. N. (1998). Student expectations in introductory physics. *American journal of physics, 66*(3), 212-224.

Reeves, T. D., & Marbach-Ad, G. (2016). Contemporary test validity in theory and practice: a primer for discipline-based education researchers. *CBE-Life Sciences Education, 15*(1), rm1.

Rogers, E. M. (1995). Diffusion of innovation theory: New York: Free Press.

Rose, S., Spinks, N., & Canhoto, A. I. (2014). *Management research: Applying the principles*: Routledge.

Rosenthal, J. A. (1996). Qualitative descriptors of strength of association and effect size. *Journal of social service Research, 21*(4), 37-59.

Saul, J., & Redish, E. (1997). Final evaluation report for FIPSE grant# P116P50026: Evaluation of the Workshop Physics dissemination project. *University of Maryland*.

Schroeder, C. M., Scott, T. P., Tolson, H., Huang, T. Y., & Lee, Y. H. (2007). A meta-analysis of national research: Effects of teaching strategies on student achievement in science in the United States. *Journal of Research in Science Teaching, 44*(10), 1436-1460.

Seymour, E. (2000). *Talking about leaving: Why undergraduates leave the sciences*: Westview Press.

Shekhar, P., Demonbrun, M., Borrego, M., Finelli, C., Prince, M., Henderson, C., & Waters, C. (2015). Development of an observation protocol to study undergraduate engineering student resistance to active learning. *International Journal of Engineering Education, 31*(2), 597-609.

Snijders, T., & Bosker, R. (2012). Multilevel analysis: An introduction to basic and advanced multilevel analysis. Thousand Oaks, California: SAGE Publications Inc.

Sokoloff, D. R., & Thornton, R. K. (1997). *Using interactive lecture demonstrations to create an active learning environment.* Paper presented at the AIP Conference Proceedings.

Sorensen, C., Churukian, A., Maleki, S., & Zollman, D. (2006). The New Studio format for instruction of introductory physics. *American journal of physics, 74*(12), 1077-1082.

Stewart, C. T., & Malpass, L. F. (1966). Estimates of achievement and ratings of instructors. *The Journal of Educational Research, 59*(8), 347-350.

Stone, A. A., Bachrach, C. A., Jobe, J. B., Kurtzman, H. S., & Cain, V. S. (1999). *The science of self-report: Implications for research and practice*: Psychology Press.

Thompson, B. (1988). Common methodology mistakes in dissertations: Improving dissertation quality.

Thornton, R. K. (1997). *Learning physics concepts in the introductory course: microcomputer-based labs and interactive lecture demonstrations.* Paper presented at the Proc Conf on Intro Physics Course.

Turpen, C., Dancy, M. H., & Henderson, C. (2016). Perceived affordances and constraints regarding instructors' use of Peer Instruction: Implications for promoting instructional change. *Physical Review Physics Education Research, 12*(1), 010116.

Webb, N. L. (2006). Identifying content for student achievement tests. *Handbook of test development*, 155-180.

Webb, N. M., Shavelson, R. J., & Haertel, E. H. (2006). Reliability Coefficients and Generalizability Theory. *Handbook of statistics, 26*, 81-124.

West, E. A., Paul, C. A., Webb, D., & Potter, W. H. (2013). Variation of instructor-student interactions in an introductory interactive physics course. *Physical Review Special Topics-Physics Education Research, 9*(1), 010109.

Wilcox, M., Yang, Y., & Chini, J. J. (2016). Quicker method for assessing influences on teaching assistant buy-in and practices in reformed courses. *Physical Review Physics Education Research, 12*(2), 020123.

Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American psychologist, 54*(8), 594.

Wilson, J. M. (1994). The CUPLE physics studio. *The physics teacher, 32*(9), 518-523.

Yerushalmi, E., Henderson, C., Heller, K., Heller, P., & Kuo, V. (2007). Physics faculty beliefs and values about the teaching and learning of problem solving. I. Mapping the common core. *Physical Review Special Topics-Physics Education Research, 3*(2), 020109.