University of Nevada, Reno

The effects of instructor questioning behavior on voluntary student engagement and academic performance in an introductory psychology course

A thesis submitted in partial fulfillment of the requirements for the degree

of Master of Arts in Psychology

by

Carolyn A. Brayko

Dr. Ramona Houmanfar/Thesis Advisor

May 2013

University of Nevada, Reno
Statewide • Worldwide

THE GRADUATE SCHOOL

We recommend that the thesis
prepared under our supervision by

CAROLYN BRAYKO

entitled

The Effects Of Instructor Questioning Behavior On Voluntary Student Engagement And Academic Performance In An Introductory Psychology Course

be accepted in partial fulfillment of the
requirements for the degree of

MASTER OF ARTS

Ramona Houmanfar, Ph.D., Advisor

Mark Alavosius, Ph.D., Committee Member

Maryann Demchak, Ph.D., Graduate School Representative

Marsha H. Read, Ph. D., Dean, Graduate School

May, 2013

Abstract

The current study investigated the effects of different instruction styles on student active responding and academic performance in an introductory psychology course at a university. A specific protocol called Personalized Instructor Engagement (PIE) was developed where instructors called on students to answer specific questions, while asking additional questions for students to voluntarily respond. Using a between group design, we exposed one group of students to the PIE condition and the other group to a control condition where all the instructor-prompted questions asked were answered voluntarily by students. The data suggest that students in the PIE condition were significantly more likely to actively participate during class than students in the control group. Although no difference in academic performance was perceived at the group level, identified “low performing” students in the PIE condition performed significantly higher on their quizzes over the course of the semester than students in the control condition. A regression analysis also revealed that the independent variable manipulation played a larger role than active responding alone in relation to academic performance. Implications and future research directions for the active responding literature are discussed.

Table of Contents

List of Tables

List of Figures

The effects of instructor questioning behavior on voluntary student engagement and academic performance in an introductory psychology course

As the current economic crisis ravages American budgets, and the importance of education in our culture is rising, educators, administrators, and researchers alike are striving to find effective yet frugal improvements to the school system. Policies on education such as No Child Left Behind coupled with a lack of funding for both public and private educational institutions are very strong contingencies that force educators to change their teaching methods to maintain their funding and providing a quality learning experience, as best they are able. No longer the era of Project Follow Through, where the U.S. federal government spent around one billion dollars to discover the most effective form of instruction, it is more likely that schools will require smaller, more impactful interventions, rather than lengthy, complex system overhauls. The following discussion reviews the history and current trends of effective educational technology in the literature then prompts the reader to consider the relation between the efficacy of the research and the practicality of dissemination and implementation. While the review of the literature attempts to acknowledge the seminal works pertinent to the current study, there are other, more exhaustive, articles that better summarize the history and breadth of these concepts (Benjamin, 1988; Kulik, Cohen, & Ebeling, 1980; Kulik, Kulik & Bangert-Drowns, 1988; Kulik, Kulik & Cohen, 1980).

The science of behavior analysis has a strong history in educational research. In fact, understanding the processes of learning in some form is a primary goal for most behavior researchers. Understandably, B.F. Skinner was one of the first behavioral

scientists to directly speak to the analysis of behavior within the context of a classroom (Skinner, 1954, Sparzo, 1992). While Skinner did not create the first "teaching machine," he designed a machine that provided reinforcement based on successive approximations, rather than simply based on trial and error (Benjamin, 1988; Skinner, 1958). Since then, behaviorally-based education interventions have greatly expanded to such notable technologies as Personalized System of Instruction (Keller, 1968), precision teaching (Lindsley, 1971), interteaching (Boyce & Hineline, 2002), as well as a variety of other procedures involving active responding, such as guided notes and electronic voting systems (Austin, 2000). While each technology stands alone as a unique teaching process, for the purpose of the current investigation, we will consider them all to be various methods for promoting active responding in classroom settings.

The term active responding refers to any process in which individuals engage with new learning material by responding publicly (e.g. taking notes, asking questions, etc.). As it is a broadly scoped concept, there are numerous various applications of active responding. Much of academic success in a class has been attributed to different opportunities for the students to actively respond to the learning material (Michael, 2006). In fact, very few educators today would disagree with its importance. There is a strong and diverse body of educational research which supports active student responding while learning new material (Boyce & Hineline, 2002; Gauci, Dantas, Williams, & Kemm, 2009; Fernald & Jordan, 1991; Griffin & Griffin, 1998; Haggas & Hantula, 2002; Hake, 1997; Keough, 2012; Malanga & Sweeney, 2008; McMichael & Corey, 1969; Miller & Mallot, 1997; Neef, Perrin, Haberlin, & Rodrigues, 2011; Neef, Van Norman, & Ferreri, 2006; Russell, Caris, Harris, & Hendricson, 1983; Saville, Cox, O'Brien, &

Vanderveldt, 2011; Semb, 1974; Semb, Hopkins, & Hursh, 1973; Sutherland & Wehby, 2001; Tudor & Bostow, 1991; Vargas & Vargas, 1991). Giving students an opportunity to interact with new information in a different way than simply reading or listening has been widely supported by research as well as most educators. Ultimately, most if not all of active responding technologies rely upon the use of questions in a classroom; yet, there is virtually no research that has probed the different approaches to questions in the classroom and their potential effects on student behavior and academic outcome.

As evidenced by the breadth of research on active responding, researchers and teachers alike have been very creative in employing new ways to engage their students. This creativity, however, comes at a cost. Some forms of active responding, such as electronic voting systems (or EVS, also referred to as “clickers”) and personalized system of instruction can have a large initial monetary and time cost. Purchasing the materials needed for these systems is a burden felt either by the educational institution or the student; and even if the process costs relatively little once running, it is the initial set-up that deters many classrooms from employing such evidence based programs. Or with larger system changes, like with a Personalized System of Instruction (PSI), many educators are daunted by the large amount of work to create and maintain such a system (Pear & Crone-Todd, 1999). Cost can also be recognized in terms of the time spent preparing the materials needed for class. For example, if guided notes or study guides, the instructor must create them and perhaps even correct the student responses, which would be in addition to the other student work that is being assigned and graded. Additionally, most active responding protocols promote anonymity in the classroom, which reduces the personal connection between the student and instructor. Choral responding, unmonitored

note taking, and EVS reduce individualized interaction between student and teacher, which reduces the opportunity to provide immediate, specific feedback.

Examining the potential benefits and disadvantages of the numerous active responding programs eventually leads to the question of what component of these active responding technologies is actually having the effect on student learning. There is a growing body of research to identify the specific processes or mechanisms that reinforce student engagement or any learning behavior. Starting with Skinner, behavior researchers have long known that creating more opportunities to respond makes for a faster shaping process (Sparzo, 1992, Sutherland & Wehby, 2001). In fact a common limitation in the active responding literature is that results favoring the intervention could be due to the fact that students had more opportunity to respond in the active responding condition than in the control (Keislar & McNeil, 1962; Kellum, Carr, & Dozier, 2001, Miller & Mallot, 1997). Other more recent studies (Gauci, Dantas, Williams, & Kemm, 2009; Malanga & Sweeney, 2008) kept the number of questions constant for all participants and still demonstrated a greater learning gain for more actively engaged students.

Although increasing the opportunities to respond in class is linked with better academic performance, the opportunity is not often capitalized upon equally among the students. The advantage of response cards or EVSs is that all students may respond at once, not excluding one another. Practically speaking, however, this is not how most classrooms are run (for the reasons mentioned above). For most classes, an instructor asks a question and one student answers, either because the instructor called upon that student or because the student self-selected to respond. What is surprising is that despite the fact that educators have been conducting their classes in this “traditional” fashion for

many years, there is no empirical work to show that the way instructors call on students has any impact on student engagement and learning.

When considering new programs for helping students succeed, many behavioral researchers have neglected to manipulate the most important part: teaching behavior. Focusing on the student behavior and the tools with which the student can actively respond has consequently ignored the influences of teaching behavior. The technology alone in some cases is not sufficient to increase student performance (Neef, McCord, & Ferreri, 2006); not to mention that simply knowing the curriculum material does not make an effective teacher (Michael, 2006). Martinez and Martinez (1988) report that "the contribution of teachers to student learning is widely undervalued" and state that the majority of educational research, specifically pertaining to Personalized System of Instruction (PSI) and mastery-learning, has ignored the potential influences of teachers (p. 26). Considering the teacher variable also has a social practicality. Researchers will inevitably bring their interventions into the schools, and the receiving teachers will be skeptical if there is no appeal to their skills as professionals (read Thompson, 1980 for the social validity measures of instructors using PSI).

Much psychological research that addresses the instructor's specific impact on student learning uses subjective and unreliable measures, such as student evaluations. In many cases, research has examined qualities of the instructor rather than behaviors (Murray, Rushton, & Paunonen, 1990; Solomon, Rosenberg, & Bezdek, 1964). Especially when teacher effectiveness is measured by students, a phenomenon known as "educational seduction" may result, where even incoherent instructors may receive an

excellent rating if they are attractive and charismatic (Abrami, Leventhal, & Perry, 1982; Naftulin, Ware, & Donnelly, 1973).

There are a few factors to be considered in terms of the utility of the above mentioned studies. First, identifying traits or qualities of the instructor may inform the educator, but gives no insight as to how to improve their situation. In the Solomon, Rosenberg, and Bezdek study, the teacher behaviors identified by the questionnaire, "energy," "clarity," "flamboyance," etcetera, are not behaviors; rather they are nouns and adjectives. From a behavior analytic perspective, the only way to change a situation is by altering an observable behavior, which is objective and described as a verb. Secondly, even if personality traits or attractiveness could be manipulated as an independent variable, the descriptions of these variables are vague and subjective. For example, Murray and colleagues (1990) conducted a large study examining instructor personality types in relation to teacher effectiveness using traits like "impulsive" (defined as: "spontaneous, impetuous, acts on spur of the moment") and "fun-loving" (defined as: "easygoing, playful, does things just for fun"). Even with definitions, these traits are difficult to observe objectively since one student may see a class activity as being impulsive, while another describes it as doing something "just for fun."

An alternative to researching unalterable variables as a factor for student success, the behavioral account pinpoints specific observable behaviors that enhance the learning environment. Instructors hold the key to establishing reinforcing contingencies for engagement and academic performance. Few articles directly speak to teacher behavior outside of contrived, and sometimes complicated, active responding technologies. The intent of this paper is to propose a simple, effective behavioral intervention that will

impact student success in a college-level course. Past research has shown that active responding is beneficial for student learning; however, many active responding techniques require technology that takes money to purchase, or time to make. Additionally, active responding techniques like guided notes, response cards, or electronic voting systems (Austin, 2000) are simply tools which can be used ineffectively as much as they can be used to produce a positive outcome. I intend to show that when using no other technology than the instructor to provide the opportunities to actively respond, there can also be a significant effect on student learning, without the time or money spent.

One distinct advantage to identifying simpler methods for increasing student engagement and performance is that such a method would be much easier to disseminate and implement in various classroom settings. Training and manuals cost a considerable amount of time and money, especially for overburdened school districts, as such, it is not overly surprising that only a few behavioral education interventions have been widely implemented (Axelrod, 1992). Thus, the current study also attempted to validate a simple classroom technique that would be efficacious, effective, and cost effective. Embry and Biglan (2008) emphasize the importance of the evidence-based "kernel," which they describe as being the most basic unit of behavioral instruction wherein one variable influences one aspect of behavior and demonstrates a functional outcome for the consumer. By utilizing behavioral kernels in the classroom, students will experience greater success in their class at very little cost to teachers and administrators. In fact, Embry and Biglan reference many kernels that would be applicable to educational settings, although calling on students is not mentioned.

Recently there have been a few educational studies that have referenced Embry and Biglan's work and have studied a kernel-based intervention in the interest of a more feasible dissemination process. In 2010, Hopson and Holleran-Stelker published findings on a substance abuse program for alternative high school students and directly speak to the necessity, yet difficulty, with getting helpful programs into "real world settings." Understandably, if a community program is too intricate, or requires materials that the program would need to create or purchase, even with strong convincing evidence it will be challenging to implement it in the communities that need it most. Applying for grants is a lengthy process and never guaranteed, so cost-effective programs are highly desired. In 2011, Cooper conducted a lengthy review of articles supporting various behavior-based teaching strategies for intervening with students with social, emotional and behavioral issues (SEBD). In his review, Cooper cites Embry and Biglan's work on kernels as providing multiple viable strategies for teachers working with SEBD. Research is still needed to identify and test the applicability of the various kernels in various settings; nevertheless, using evidence-based kernels appear to be a pragmatic solution to a host of "real world" problems.

After reviewing the active responding literature and recognizing its various benefits and a few common disadvantages (like dissemination issues), we found a commonality among the various active responding techniques that, when using a kernel approach, may potentially be easier to implement in any classroom on any budget. The simplest way to introduce active responding into a classroom setting is by asking questions. In my current study, the instructors interacted with their students on a personalized level, such that almost every student was asked to respond and had

opportunities to respond throughout the course. We called this type of specific student-teacher interaction “Personalized Instructor Engagement,” or PIE, which essentially means calling directly on students while still providing opportunities for them to volunteer responses and will be further defined below. Using a between-groups design, the current study aimed to: (a) determine the effects of PIE on student engagement, and (b) determine the effects of PIE on student academic performance. Our analyses included both individual and group data to reflect the potential effects of PIE at both the micro and macro level within the classroom. Finally a regression analysis was conducted to determine to what degree the variance in academic performance was accounted for both the independent variable and active responding.

Method

Participants & Setting

The participants for this study consisted of 276 students from both sections (1 and 2) of Dr. Houmanfar’s Psychology 101 course at the University of Nevada, Reno with an approximate balanced distribution of participants between sections (n = 144, 132, for Section1 and 2 respectively, sample after applying inclusion criteria; n = 185, 156, respectively, sample before applying inclusion criteria). The aim of the course is to teach university students a general overview of the different fields in the science of psychology. Individualized data were only collected for the participants who signed a consent form; however some de-identified data collected at the aggregate (class-wide) level included all students. For individualized data collection, there were strict inclusion criteria. The study only included participants who: (a) attended at least eight of the 10 class lectures, and (b) were not moving at the faster pace at any point during the semester (because they will not

be exposed to the same variables as students attending a larger class). Any student who attended a lecture more than 5 minutes late or attended a discussion for the other section was counted as “absent” and his/her data were not collected.

Although exhaustive demographic data were not collected, there is some evidence that the two sections of students shared some similarity. Students in this Introductory Psychology course were primarily freshman (77% and 68%, respectively for Sections 1 and 2) and the majority of students registered for this course to fulfill a university social science requirement (72% and 82%, respectively).

After completing a pretest before instruction, the average score for both sections was 46.6%. The study also compared quiz scores across sections for the top and bottom 10% performers (see “Data Analysis” for a more elaborate description). Among the top performers, the average pretest scores were 47.0% snd 55.6%, for Section 1 and Section 2 respectively. For the bottom performers, the average pretest scores were 46.3% and 44.5%, for Section 1 and Section 2 respectively. While there was an initial modest difference in groups for the top performers before instruction, there was no measurable difference between the bottom performers.

The course in which the study was conducted consisted of two large enrollment sections which followed a form of PSI. With this particular PSI structure, the students are required to proceed through the course at a minimum pace of one chapter a week; however, if the students reach a mastery criterion, they have the option to move through the material more quickly and have the option to finish the course early. As noted earlier, students moving at the faster pace were not included in the study thus their requirements and progress will not be described in this paper.

For students progressing at the regular pace, they were required to attend one lecture a week in Edmund J. Cain Hall, room 253, also known as the Learning Lab, on the campus of the University of Nevada, Reno. For each section there was a selection of 12 different class times that were evenly dispersed across the week and time of day. This room consisted of a classroom where the students attended a weekly lecture, an attached computer lab where the students were required to take all of their online quizzes and exams, and an office where the undergraduate proctors and teaching assistants manage the administrative duties for the course. The classroom was large enough to accommodate up to 28 students, and equipped with a computer, projector, and screen on which the instructors presented standardized PowerPoint presentations for each chapter. The computer lab is accessible only to Dr. Houmanfar's Psychology 101 students. As a part of the PSI structure of this course, students had access to a scheduling website, for making appointments for their classes, quizzes, and exams, in addition to interacting with the course webpage set up through the university's online course management system, WebCampus Learn. As students enter the Learning Lab, they are asked to stop at the office window. There, they are able to check-in with a proctor for either their lecture time or quiz appointment.

The instructors consisted of four teaching assistants currently enrolled in a graduate psychology program at the University of Nevada, Reno. These teaching assistants were not only responsible for presenting the course material and review sessions, but also for implementing the independent variable manipulation and recording student engagement data. Of the four teaching assistants, two had taught this course material at least twice before, and two taught the material for the first time during this

project. All instructors were expected to be familiar with the provided lecture material taken from the course's textbook, Peter Gray's *Psychology* (2011).

Independent Variable

The independent variable for this study was a manipulation of the type of instruction to which students were exposed. In addition to the structured overview of the content, students were exposed to either the presence or absence of the independent variable: Personalized Instructor Engagement (PIE) versus no-PIE. PIE simply refers to the process of the instructor individually prompting students in addition to providing opportunity for students to volunteer responses. PIE is a two-step process. First, the instructor calls on a student from a preselected list of random desk numbers, then asks the prepared question to the selected student. Instructors constructed PIE questions in this manner to maintain consistency of question-asking as well as reduce the likelihood that being called on in class would become an aversive condition. For example, if the instructor asked the question first and then called on a student, the student may find the attention aversive if he or she was not paying attention, or did not hear the question. This method may also have an effect on student engagement; however, the current study required instructors to call on the student before asking the question.

In the absence of PIE, students were still asked questions by the instructor, but the instructor did not select any student to respond. After the question was asked to the class, students were given a maximum of 10 seconds to respond, if they chose. If no student answered the question, the instructor answered the question and moved onto the next slide. In this sense, we will refer to these questions as "voluntary" on behalf of the

student. For both conditions, students were always allowed to make additional questions or comments in class. Such responses will be referred to as “unprompted.”

The same 12 chapter-specific questions were asked in all classes for both sections. For the PIE classes, eight of the 12 questions were PIE questions, or pre-marked on a sheet to indicate that the instructor should call on a student (see Appendix A). The remaining four questions were voluntary in that the question was asked generally to the classroom and only a student who selected him/herself to respond would be called on. While the number of questions asked remained constant across groups, the nature of PIE (calling on students), precluded the study from also keeping an equal number of voluntary questions across groups. The PIE protocol only assigned eight of the 12 questions to call on students, because we did not want to establish a contingency that voluntary responses were not allowed or acceptable. For the no-PIE classes, all 12 questions were voluntary questions. By holding the number of questions asked constant, as well as the questions themselves constant, individually prompting students was the only aspect of the course that varied across sections. From this, a reasonable claim can be made that any meaningful differences in engagement and academic performance between sections may be due to the instructor soliciting responses or student self-selecting to respond. Instructors were restricted to only asking the 12 prepared questions; and the only other questions asked during the class session consisted of a six multiple-choice question quiz at the end of the lecture (see Appendix B).

Dependent Variables & Measurement System

The study identified two dependent variables to examine in relation to the presence or absence of PIE: voluntary student engagement and academic performance.

There are many ways that students engage themselves with course material, yet pragmatically speaking, we were limited to only observing public behavior occurring during the lecture. In addition, due to technology limitations, instances of "engagement" could only include behaviors observable and recordable by the instructor. This study defined student engagement as a student verbally responding in relation to the course material. Acceptable, recordable responses included students' responses to PIE (individually prompted by the instructor) questions, voluntary (self-selection) questions, or unprompted questions or comments. An unprompted response included any verbal interaction between student and instructor that was initiated by the student.

Student Engagement. To measure student engagement, the instructors used seating charts (see Appendix C), to capture two different levels of data: individual and class-wide. The seating charts, filled out by the undergraduate proctoring staff, included the names of students and where they sat for a particular lecture. The instructors marked on the seating chart using different symbols to indicate the number and which students verbally responded, under what condition (PIE, voluntary, or unprompted), and the number of responses occurred per class session. By coding each of the seating charts, we were able to track the participants throughout the semester and note their level and quality (topography) of their verbal engagement during classes. In addition to individualized data, we also measured class-wide, aggregate data to track the ratio of student responding, that is, the number of students who responded at least once during a class divided by the total number of student responses during that class.

Academic Performance. To measure academic performance, we used the scores for the two types of quizzes regularly administered for each chapter (referred to as "end-

of-lecture quizzes" and "chapter quizzes"). The end-of-lecture quizzes consisted of six multiple choice questions, each worth one point, which tested primarily definitional information provided in the PowerPoint for the relevant chapter (see Appendix B). Quiz questions were the same for both conditions. Either the instructors or the undergraduate proctors manually corrected each quiz, and then entered the grades into the electronic grade-center via the university provided online course management system, WebCampus Learn.

All other quiz scores were automatically entered because the course required students to take their assessments online in the supervised computer lab. We obtained the chapter quiz data by using WebCampus Learn. Each chapter quiz was 15 multiple-choice questions, randomly selected by WebCampus Learn from a test bank, worth a total of 15 points. The content of the chapter quizzes only covered material from that particular textbook chapter. Due to the special nature of this course's PSI structure, students were allowed to retake their chapter quizzes up to nine times during the week. Monday through Thursday, the students had the opportunity to take their chapter quiz two times a day, and on Friday they could only take it once.

Because the aim of this study was to focus on the impact of the instructor style, the data analysis only considered the initial quiz attempt. It should be noted, however, that students were also allowed to take their chapter quizzes before they attended class. Although it is rare for students to do this, it is possible that the data may include scores from quizzes taken before the participant attended a class, where they contacted the experimental conditions. It was not feasible to obtain these data for 276 participants for 10 separate quizzes, so all initial chapter quiz scores were collected regardless of timing.

All scores were calculated by number of correct answers in the assessment. The course worked on a point-based system, where each question answered correctly counted as one point toward their final cumulative grade. A key for each of the chapter quizzes and exams had been established from previous semesters; whereas, the experimenter wrote the end of-lecture questions and answer key. Interobserver agreement was not explicitly measured for grading quizzes for two reasons. Although end-of-lecture quizzes were graded by hand, all students had access to their grades through WebCampus Learn and could notify the staff about any discrepancies, which serves as an informal, functional IOA. Additionally, for the chapter quizzes, no IOA was conducted because quizzes were taken online and graded automatically through WebCampus Learn.

Experimental Design

There were two levels of the independent variable, the presence of PIE in the classroom and the absence of PIE in the classroom. Due to the constraints of the university course sections, students were not randomly assigned to the groups; however, the conditions were randomly assigned to the sections. Furthermore, due to the slow emergence of PIE's effects on classroom engagement, the study did not attempt to reverse the conditions within groups. Therefore, the study had a quasi-experimental, between-subject design. This between-group design with a pretest has the inherent weakness that only one measurement was collected before instruction started to demonstrate the initial similarity of the groups (Shadish, Cook, & Campbell, 2002). While a between groups design is not the ideal for experimental rigor, time constraints and limitation in the number of lectures students are required to attend excluded other design options, such as a counter-balanced reversal design. The experimenters found it

more valuable to see the potential long term effects of PIE, or no PIE, on engagement and academic performance rather than switching conditions prematurely. Had we switched conditions during the semester, it is possible that the group differences that emerged could have been masked. The data were collected repeatedly on the same measures throughout the semester, which gives the study a time-series design element (Shadish, Cook, & Campbell, 2002). In addition to the single pretest, the repeated measures taken show the potential effects of the independent variable on the dependent variables over time.

To ensure that the experimental design was implemented consistently, students could only attend the lectures that were held for their section. To minimize student confusion, both the syllabus and the WebCampus course page posted section-specific schedules (see Appendix D for the master class schedule). To minimize instructor confusion, the master schedule, displaying both Section 1 and Section 2 classes, was posted next to the computer from which the instructors taught. By limiting students to solely attend lectures specific to their section, they were only exposed to one condition. For Section 1, students were exposed to the PIE variable throughout the entire semester. For Section 2, students were exposed only to a no PIE, or volunteer only, condition. Overall, the study was conducted over 10 weeks, where one chapter was covered each week. In addition to that, there were 3 weeks for student orientation to the course, during which the students took a pre-test, 1 week between lecture weeks 5 and 6, allotted for the students to take their midterm, and 1 week at the end of the semester for the students to take their final. Due to the time constraints of studying a college course, the experimental design was limited to the course's timeline.

In order to obtain social validity data, the study analyzed data from the university required student evaluation form. Each student was asked to complete a course evaluation form that was compiled by past administrators of this course with a few additional questions specific to the current study. The aim of the survey was to capture the students' overall satisfaction with the course, the content, and the instructors. There will also be an opportunity for them to comment on whether or not they enjoyed the PIE manipulation, whether they felt that method of instruction helped them learn the material, and if they could, would they choose to take a course with a similar instruction style again.

Procedure

To minimize as much instructor variability as possible, a thorough protocol was administered and practiced by all instructors before the start of the semester. As stated before, all instructors used a seating chart (Appendix C), and after confirming at the beginning of class that all students were seated in the correct seat, the instructors were trained to call on students by their name during class instruction. In addition to training instructors on the experimental protocol, research assistants sat in as observers for roughly one third of the discussions for the last five weeks of instruction to calculate program integrity.

It was also vitally important that instructors gave consistent feedback similarly to the other instructors. If some students received a negative or discouraging response from the instructor, responding in class could have become an aversive situation in which the student would have responded less than they would have with a reinforcing consequence. To help with this potential confounding variable, instructors were given a formulaic

protocol to standardize instructor reactions to student responses whether they answer correctly, incorrect, or partially correct (Appendix E).

In attempt to maximize experimental control within the classroom, we denied students from using laptops and other electronics during class. Although we cannot prevent students from entirely disengaging from the presented material, the study did constrain students from participating in more putatively reinforcing behaviors like surfing the Internet. Prohibiting electronic devices during class also helped to ensure that students did not cheat on the various quizzes and assessments.

In both conditions, the instructors used the same PowerPoint presentation for each chapter, to ensure that the material covered was consistent across instructors, lectures, and conditions. Across both sections, all instructors taught directly from the prepared presentation by reading the information on the slide and then elaborating with examples. If at any point a student had a question or comment about the material, the instructor called upon the student using their name, referencing the seating chart if necessary. If no student raised their hand or made a voluntary comment, the instructor continued moving from slide to slide. Each of the 10 questions was principally definitional and was asked while the lecture material was still presented on the screen (see Appendix A for examples of lecture questions). Every instructor asked the same sets of questions for both sections, which were printed out and left at the computer the instructors use for their lectures.

Number prompts for the instructors were embedded in the PowerPoint slides to enhance the likelihood that all twelve questions were asked at approximately the same points within the lecture. On average, a question was asked every one to two slides during the PowerPoint presentation, and the numbered icon presented on the slide

corresponded to the numbered question to be asked. Specific to PIE questions, there was another printed list of desk numbers that was taped to the computer monitor indicating which students to call on and in what order. The numbers were randomly selected from the website random.org without repeating any desk number. A new list of desk numbers was presented every week. Using random numbers helped to ensure that all areas of the classroom and a greater diversity of students were selected for PIE questions. If in the event that no student was sitting in one of the indicated desks, instructors moved down the row to the next student who had not previously been called on for a previous PIE question. It should be noted, however, that in the occasion that there were fewer than eight students, some students were asked PIE questions more than once within a single class session.

For every overt verbal response, instructors indicated on the seating chart which student responded and the response topography (see Appendix C). For all ten weeks, the protocol instructed the teachers to distinguish PIE responses from all others; however, halfway through the study, we included the added discrimination between the voluntary response (instructor initiated question and student initiated response) and the unprompted response (student initiated question or comment). Thus, the students exposed to the PIE condition could respond in all three forms (PIE, voluntary, or unprompted), whereas, students in the no PIE condition only had the opportunity to respond voluntarily or make their own question or comment.

To maintain control, instructors were not permitted to ask any additional questions of their own, including general questions such as, “Do you have any questions?” during the lecture. Once the lecture slides were completed, instructors were

free to speak and transition as they normally would into the end-of-lecture quiz. Instructors likewise ceased marking the seating chart once the class had moved onto the quiz. For example, if a student asked for the instructor to repeat one of the quiz questions, that response would not be marked as a response within the class session.

Data Analysis

To examine engagement, we analyzed both individualized data following the participants over the course of the semester and aggregated data to examine how the two conditions could be affecting the class as a whole. Based on the seating charts, quantity and topography (PIE, voluntary, or unprompted) of responses were noted for each participant for each week. Also, for every class each week, we recorded the total number of responses, the number of responses according to topography, and the number of different students who responded. These data allowed for us to calculate the ratio of responding for each class session.

Academic performance data were collected using the university's course management system, WebCampus Learn, which the course used for entering and posting grades for students. The WebCampus Learn Grade Center allows for both individual and group analysis (by section).

For both dependent variables, the study conducted both descriptive statistics, to allow for visual inspection of the data, and inferential statistics, to determine the group differences between sections in relation to the independent variable. We presented descriptive data in graphs for visual analysis and used Cohen's *d* to calculate effect sizes. Analyses of variance (ANOVA) were run to determine whether the mean differences between sections were statistically significant both on a weekly basis and total

performance for the whole semester. When the analyses of variance were significant, eta-squared (η^2) was also calculated to show the effect size.

A second analysis was performed to examine between-subject differences at the individual level. For student engagement, overall participation for the semester was calculated for all participants and compared across sections. For academic performance, 20% of the participants were retroactively selected based on their overall performance on end-of-lecture quizzes and chapter quizzes. The participants' data whose scores were among the top 10% were analyzed, as well as the data from the bottom 10% of participants (these groups will further be referenced as "top performers" and "bottom performers," respectively). These secondary analyses are essential for better understanding the degree to which the independent variable exerts a contingency for increased student responding and improved academic performance for all levels of students. When considering the utility of an educational intervention, it is helpful to know how they will impact specific demographics, like top and bottom performers, rather than only considering the average impact for the hypothetically average student.

A third analysis was completed to investigate the difference between the PIE and no-PIE condition when considering both dependent variables: student engagement and academic performance. To do this, only the bottom 10% of performing participants was used. The sample was restricted to the bottom performers due to the fact that it was only within that smaller sample that significant differences in academic performance were detected. Additionally, there is utility to considering the potential effects on the lowest performing students when implementing a new intervention in a classroom. Low performers are most likely to be susceptible to subtle shifts in active responding

contingencies since they presumably have a weaker repertoire for achieving classroom success. For this analysis, we conducted a multiple linear regression which analyzed which of the two variables, the independent variable manipulation or actually responding in class, correlated more strongly with academic performance. The purpose of using a multiple linear regression model in this case gives a more descriptive account of the effects of PIE: if the PIE contingency actually has a unique impact on classroom academic performance, or if PIE functions solely by virtue of actively responding in class.

Finally, in line with university policy, the course administered an end of the semester course evaluation for both sections (see Appendix G). The course evaluation was modified to included questions relevant to the study (questions 16 through 19). Class-wide data were collected and compared across groups to evaluate the social validity of the PIE protocol. Even with the potential benefits of PIE, practically speaking, it is important to consider the perceptions of the recipients of the PIE, or no PIE, conditions.

Program integrity data were collected during the last five weeks of the study. For 34% of the classes offered during weeks 6 through 10, trained observers attended the class lectures for each of the instructors. All observers used a standardized checklist indicating the protocol's important features (e.g., starting on time, calling on students in the correct manner, etc.). For every correct application of the protocol throughout the class, the observer would mark the corresponding box (see Appendix F). To calculate integrity data, total correctly executed items were divided by the total of correct and incorrect items, multiplied by 100. The instructors were near perfect in protocol implementation at 97% and 96% accuracy for Section 1 and Section 2 protocols

respectively, with the most often neglected item was finishing within 10 minutes of the scheduled end time of the class. Total program integrity for all instructors for both sections was 97%, with the accuracy ranging from 95%-98% among the instructors.

Results

Student Engagement

The primary dependent variable was student engagement within the context of the students' weekly class lectures. The average number of responses made per class per week compared across sections is represented in Figure 1. Although the data are variable across sections, the average number of responses per class is higher in seven out of ten weeks for Section 1 students who received the PIE variable (the difference between sections during week nine is minimal). The difference between sections simply based on number of responses, when considering the difference between the means over all ten weeks of instruction, is not significant, $F(1, 230) = 0.94$, $p = 0.3$.

To get a more descriptive account of student participation between sections, the average ratio of student responding was also calculated for each class session across the whole semester (see Table 1 and Figure 2). The ratio indicates the number of students who responded at least once during a class divided by the total number of student responses during that class. At the beginning of the semester, a similar number of students responded across sections. This is not surprising considering that it takes time for students to come under the control of the classroom contingencies when they only attended one 90 minute class per week. By week four, the average ratio of responding started to diverge and continued to grow. Considering the average number of responses from week to week in Figure 1, Figure 2 shows that despite the fact that the total number

of responses do not greatly differ between sections, fewer students in Section 2 were responding, compared to Section 1 students. When considering overall differences between the two sections over ten weeks of instruction, the difference of ratio responding is strongly significant with a moderate effect size, $F(1,229) = 16.04$, $p < 0.001$, $\eta^2 = .07$, $d = 0.46$.

Part of this significant difference is due to the nature of PIE which requires a large diversity of student responding. For the eight PIE questions asked during the class, the instructors called on eight different students. Looking at the ratio of responding data (Figure 2), it appears to validate the process of PIE. The unique quality of PIE as an active responding technique is that instructors are calling on specific students, not just providing the opportunity to respond. The between-section comparison of ratio of responding confirms that PIE was successful in its aims to select students to respond, and as a result, maintained a high ratio of responding, unlike in the no PIE classes. Additionally, all 12 questions were more likely to be answered in Section 1 than Section 2 (Figure 3). Again, one must consider that 66% (8 out of 12) of the questions in the PIE condition guaranteed a student response; however, there was no reason to expect that all four voluntary questions would be answered.

The current study was also interested in whether the independent variable had an effect of voluntary engagement, that is, student responding outside of instructor initiated interaction. Because there was an uneven number of “voluntary” questions built into the two different conditions (PIE = 4, no-PIE = 12), the study measured and analyzed unprompted student responses as a purer indication of voluntary engagement. Figures 4 and 5 take different approaches to analyzing unprompted responding between groups.

Figure 4 displays the average number of unprompted responses that occurred for each week between groups. Figure 5 indicates an overall increase in percentage of classes with unprompted responses over the course of the semester. Looking more specifically at section differences, with the exception of weeks 2, 3, and 6, there was a higher percentage of Section 1 classes with unprompted responses than Section 2. Over ten weeks, unprompted responses occurred for 56% of Section 1 classes, and 43% of Section 2 classes. The difference in classes with unprompted responses is significant across sections with a small effect, $F(1, 230) = 3.91$, $p < .05$, $\eta^2 = .02$, $d = 0.25$.

Our final evaluation of the independent variable's effects on student engagement examined individual student engagement over the course of the semester. By tracking each participant throughout the ten weeks of instruction, we could determine the percentage of the participants responding at different levels. The data in Figure 6 depict three ranges of responding to instructor prompted questions (PIE or voluntary questions) as well as the percentage of students who made at least one unprompted response between weeks 6 and 10. In the middle of the semester, an update to the instructor protocol was made so instructor coding on the seating chart discriminated between voluntary (instructor prompted) questions from unprompted (student initiated) responses. Figure 6 shows a greater percentage of Section 1 participants who responded one to four times than Section 2 participants (34%, 27%, respectively). There is a similar difference between Section 1 and 2 participants for responding five or more times during the semester (62%, 53%, respectively). What is most astounding is the large difference in the percentage of participants who never responded. A little more than 2% of Section 1 participants were never called on by the instructor and never initiated their own comment.

For Section 2 however, over 18% of participants never had any over instructor interaction during ten weeks of instruction. Finally, for the final five weeks of instruction, a slightly larger percentage of Section 1 students responded without a prompt from the instructor than Section 2 students (18.2% versus 15.4%, respectively).

Academic Performance

The study's secondary dependent variable, academic performance, used end-of-lecture discussion quizzes and weekly chapter quizzes to investigate the degree to which the independent variable affected the participants' grades. Taken at the largest level of analysis, a between-section comparison for all participants, there is not much to suggest that there is any significant difference between groups on end-of-lecture or chapter quizzes (see Tables 2 and 3; Figures 7 and 8). The overall grade distributions were calculated for all students in both sections. Again, the differences between sections are minimal with 84% of Section 1 students earning a C or above, compared to 81% of Section 2 students earning a C or greater (see Figure 9). In particular, it is also worth noting that roughly the same difference between sections for percentage of students earning C's (Section 1, 24.1%, and Section 20.3%) is similar to the added section difference for students earning D's and F's (Section 1, 6.5%, 7.8% and Section 2, 10.0%, 8.0%, for D's and F's respectively).

If PIE had any effect on academic performance, it is completely masked at the group level. Analyzing the data for all participants in two large groups assume homogeneity of subjects, which is not the case when considering that our sample included students of all ability levels and unique learning histories with respect to classroom engagement and academic success. Thus the primary analyses for academic

performance used smaller groups, based on performance level, potentially providing a more descriptive analysis of how a PIE classroom might differently affect certain individuals more than a no-PIE classroom.

After all the data had been collected, participants were identified and grouped according to academic performance. Taking into account both the scores for the end-of-lecture quizzes and the final chapter quizzes, we selected the participants who scored within the top 10% of the sample (top performers), and the bottom 10% of the sample (bottom performers). Once the individuals were identified, their performance data were compared according to performance level across sections for both end-of-lecture quizzes and chapter quizzes over the course of the semester. For both sections, the top performers on average scored very well throughout the semester (See Tables 4 and 5; Figures 10 and 11). There is more differentiation between sections for the chapter quizzes than for the end-of-lecture quizzes, which may suggest there was a ceiling effect on a quiz that only consisted of six questions taken directly from the lecture. There were no significant differences between the sections among the top performers at the week to week level, or when considering the semester as a whole. Although, using visual inspection for chapter quizzes (see Figure 11), Section 1 participants did perform as well or better on average than Section 2 students for 7 of the 10 weeks.

Larger differentiations were found among the bottom 10% of participants when compared across sections (See Tables 6 and 7; Figures 12 and 13). In Figure 12, the difference in average end-of-lecture quiz scores between sections increases throughout the semester, with the exception of week 5 where bottom performers performed about equally on their end-of-lecture quiz. Particularly at the end of the semester, the

differences between the averages of end-of-lecture quiz scores for Section 1 versus Section 2 bottom performers is statistically significant. Refer to Table 6 for specific effect sizes and ANOVA results for all weeks for the lowest 10% performance group. When considering the mean difference in end-of-lecture quiz performance for the whole semester, the bottom performing participants in Section 1 scored significantly higher than Section 2 bottom performing participants with a moderate effect, $F(1, 240) = 18.04$, $p < .001$, $\eta^2 = .07$, $d = .53$.

There is a weaker trend shown in Figure 13 when average chapter quiz scores were compared between sections among the bottom performers. While the difference of academic performance between groups from week to week was only statistically significant for week 5 of instruction, Section 1 bottom performer students typically earned higher grades on their chapter quiz than Section 2 bottom performer students (9 out of 10 weeks; refer to Table 7 for specific effect sizes and ANOVA results). When considering the chapter quiz performance between Section 1 and 2 for all ten weeks, the difference between the average scores was significant with a small effect, $F(1, 242) = 4.78$, $p = .03$, $\eta^2 = .02$, $d = .28$. The weaker trend in chapter quiz scores, compared to end-of-lecture quizzes is not overly surprising since students take their end-of-lecture quizzes immediately at the end of class, thus immediately after coming into contact with either the PIE or no-PIE contingency. Additionally, since students were allowed to take their chapter quizzes anytime during the week, before or after they attend class, it is certain that other environmental influences and contingencies are also influencing academic performance for chapter quizzes.

As a final analysis of the academic measures, we conducted ANOVAs considering the potential influence of instructors on academic performance (see Tables 8 and 9 for full list of data). In an effort to show further internal validity in this study, the ANOVAs compare the means of the quiz scores for the students that attended a particular instructor's discussion for one week. Due to the fact that students were allowed to change class times and instructors throughout the semester, ten separate analyses were conducted for both end-of-lecture and chapter quizzes. No significant difference implies that regardless of condition, quiz performance was not significantly impacted due to the instructor variable alone. Out of 20 ANOVAs, there were only two weeks for end-of-lecture quizzes and one week for chapter quizzes where instructor differences were statistically significant (refer to Tables 8 and 9). Overall, it can be said that there is no known evidence to show that the differences in instructor had any impact on academic performance.

Interaction between Dependent Variables

The study also investigated the potential interaction between the dependent variables, student engagement and academic performance, in relation to the differences in the independent variable. It does appear, especially for certain demographics within the sample (bottom 10% performers), that being exposed to the PIE condition increases engagement in class and academic performance; however, it is also prudent to see if there is any relation between engagement and academic performance. Since the greatest differences between groups existed among the low performers, this individual data analysis continued with the same bottom performer sample.

When simply correlating the relation between individual end-of-lecture quiz scores with the number of responses the individual emitted throughout the semester, there was a weak to moderate effect, $r(242) = .20$, $p = .001$. This correlation includes both section for the purpose of demonstrating that there is at least a general relation between responding in class and performance on the end-of lecture quiz.

A multiple linear regression analysis was conducted to see which variable had greater predictive value for academic outcome on end-of-lecture quizzes: the experimental condition or the number of times the student responded. To determine the effectiveness of PIE as a kernel, it is important to distinguish the difference between the effects of PIE and simply responding in class (active responding). That is, we wanted to investigate if PIE was solely functioning as a result of implementing a contingency for active responding or if PIE has an additional influence on performance outside of active responding. Both the independent variable (PIE) and the number of responses (engagement) played a significant role in academic performance, $R^2 = .11$, adjusted $R^2 = .11$, $F(2, 239) = 15.20$, $p < .001$.

Additionally, the section the student was in bore a more significant influence over academic performance than responses alone (see Table 10). Simply put, roughly 10% of the quiz score variance is explained by both independent variable (PIE versus no PIE) and degree of student engagement. Furthermore, the difference between Section 1 and Section 2 is -.52 points on the end-of-lecture quiz (with the average Section 1 student earning 0.52 points more on the quiz than the average Section 2 student) regardless of the number of times a student responses. Considering that the end-of-lecture quizzes were only worth 6 points, a half point difference equates to about eight percentage points. So,

for the bottom performers, being in the PIE classroom (Section 1) suggests that on average, they scored almost a letter grade higher on their end-of-lecture quizzes than their no PIE classmates (Section 2). In the same vein, responding once in class could increase the quiz around .21 points in either condition. In all, the data suggest that active responding may have partially mediated the PIE contingency; however, PIE's greater correlation to academic performance suggests that there was a unique element to this intervention that had an effect regardless of student overt responding.

Another way to consider these data are seen in Figure 14 where the end-of-lecture quiz grades of low performers are analyzed according to amount of responding during a single class. Students who responded tended to score better on their quiz than students who did not respond at all (average quiz scores: m = 4.99, 5.01, 5.31, for "0 responses", "1 response", and "2+ responses" respectively). The data suggest that the independent variable had less of an effect on student performance, when the participant responded two or more times in class. This is evidenced by the fact that there is very little difference in average quiz scores between Section 1 and Section 2 participants who responded two or more times in class. For instances when participants did not respond during class, there is a noticeable difference between Section 1 and Section 2 participants' quiz scores. Section 1 bottom performing participants scored an average of 5.25 points (out of 6) without ever talking in class; yet, Section 2 bottom performing participants who did not talk in class scored an average of 4.58 points. The difference in quiz scores between PIE and no PIE for no responses during class is statistically significant for the low performers, $F(1,135) = 17.50$, $p < .001$. The significance of these data are that the PIE variable may be establishing a contingency, reinforcing students for learning the class material whether or

not they are overtly responding. On the other hand, bottom performing students not who did not come into contact with the PIE variable did not engage and learn the material as well when there was no possibility of the instructor calling them without their self-selection.

Finally, we analyzed the differences in the cumulative active responding with respect to the identified top and bottom performers for participants in both sections. The data in Figure 15 represent the average cumulative record for top and bottom performers for both sections. Visualizing the data in this manner allows for visual inspection of both within section and between section differences. For Section 1 participants, both top and bottom performers responded at a similar pace throughout the semester; whereas, Section 2 top performers performed more frequently than Section 2 bottom performers, such that by the end of the semester the average Section 2 top performer participated almost twice as much as the bottom performer (9 versus 5.77, respectively). Between sections Section 2 top performers responded more frequently than Section 1 top performers; however, Section 1 bottom performers responded more than Section 2 bottom performers.

Social Validity Analysis

An additional analysis was conducted to determine the social validity of the PIE intervention. Especially since it is common for instructors to worry that calling on students would be aversive, the study explored the students' evaluations of the class in terms of enjoyment, engagement, and learning. See Appendix G for the full social validity questionnaire and Table 11 for a more detailed description of social validity data.

Students who encountered PIE reported enjoying this course more than other introductory level courses they have taken more than students who did not encounter PIE

(32.1%, 29.9% respectively). Additionally, more Section 1(PIE) students reported that the questioning style increased their level of engagement than Section 2 students (59.1%, 49.8% respectively), as well as marking that classes were helpful for quiz preparation (60.4%, 56.7% for Section 1and Section 2 respectively). Not only would Section 1 students be more interested in learning in the same manner again (63.5, 54.0%), but also would be more likely to recommend this course to a friend (80.4%, 71.3%).

Discussion

The main purpose for the current study was to determine the effectiveness of Personalized Instructor Engagement (PIE) on student performance and academic performance in an introductory psychology course. While there are many effective behavioral technologies available for classrooms, disseminating those techniques has proven difficult. This study proposed an active responding technology as a proposed kernel to further uncover the mechanisms of active responding and to show evidentiary support for a simple classroom intervention (randomly calling on students). The data suggest that PIE is consistently effective at maintaining a high ratio of student responding as well as promoting academic performance among students will weaker repertoires for succeeding in class (bottom 10% of student performers). When considering the results of the current study, PIE's effectiveness provides several directions for future research in addition to some important implications for disseminating this technology to wider audiences.

Overall, the results suggest that students who were exposed to a classroom where the instructor randomly selected eight students per class to answer questions (PIE) were more likely to respond in class, but not significantly (Figure 1). It should be noted,

however, that students in both sections were exposed to 12 questions during each class. The fact that average responding exceeded 12 responses for both sections for every single week, suggested that the supportive structure of the Personalized System of Instruction within the course minimized any potential effect from the independent variable at this level of analysis.

For classes where instructors did not individually prompt students to answer questions, over the course of the semester fewer students participated during the class session (see Table 1 and Figure 2). There is an important distinction between looking at average number of responses during a class and the number of students responding. While roughly the same number of responses occurred during both Section 1 and 2 classes, a greater diversity of students responded in Section 1 than in Section 2. This was shown by the fact that the ratio of responding remained consistently high for Section 1 students but decreased over time for Section 2 students. Additionally, potentially as a result of the different classroom contingencies (PIE versus no PIE), a small proportion of Section 1 participants never spoke up during class, as opposed to a much larger percentage of Section 2 participants never overtly responded over 10 weeks. If the instructor's goal is to engage a larger proportion of the class during lectures, using PIE is significantly more effective than asking questions without selecting the students to answer them.

One potential criticism of PIE is that it artificially inflates student participation since instructors call on students to respond, whereas in the no PIE condition all responses were voluntary. To address this concern, we included unprompted responses as a measure of student engagement (refer to Figures 4 and 5). Unprompted responses

included any student initiated comment or question. Since the number of questions and amount of lecture material was held constant for all classes for each chapter, it is reasonable to assume that there was an equal opportunity for unprompted responding for both conditions. A higher percentage of PIE classes occasioned unprompted responses than no PIE classes. The data at the individual level also support the claim that unprompted responses were more likely to occur in PIE classes rather than no PIE classes. Thus, not only does PIE promote active responding in relation to instructor prompted responses (PIE or voluntary) but also unprompted responding. It is possible that the contingency of PIE reinforces the responding behavior of all students, not just the ones who are being called upon by the instructor.

In relation to the study's secondary dependent variable, academic performance, PIE had a weaker effect on the class-wide level; yet, when considering various ability levels of the participants, there were some important distinctions between groups. For example, for both top and bottom performers, Section 1 students scored higher on average than Section 2 top and bottom performers on the chapter quizzes (refer to Figures 11 and 13). Particularly among the bottom performers, participants in the PIE condition performed significantly higher than participants in the no PIE condition over the whole semester for end-of-lecture quizzes and chapter quizzes (Figures 12 and 13). Especially when visually inspecting the end-of-lecture quiz data for the bottom performers, Section 2 participants performed consistently from week to week whereas Section1 participants' average performance increased steadily over the course of the semester. This differentiation is particularly interesting because new material was covered each week, thus the increased quiz scores cannot be due to practice effects. When considering the

potential cumulative effects of PIE, the grade distributions suggest that the same lower performing students in Section 1 were more likely to move their grade from a D to a C, than Section 2 students (see Figure 9).

By considering all three variables in relation to one another, the regression analysis suggests that a small portion of the academic performance variance can be explained by the manipulation of both the independent variable and active responding. Considered separately, the PIE condition accounted for about 27.0% percent of the variance in academic performance, when holding active responding constant. Active responding, on the other hand, only accounted for just 4.8% of the variance in quiz performance when experimental conditions were held constant. Clearly there are other variables influencing a student's performance, and there was not expectation that a kernel like PIE would overcome more significant factors of academic success (e.g., whether the student studied, read the textbook, contacted the material in other classes).

Looking at the regression as a component analysis, the data suggest that there is a noticeable difference between simply responding in class and learning in a PIE classroom. Future research could further explore the form versus the function of certain active responding technologies. While active responding was still significantly correlated with academic performance, the independent variable manipulation had a large effect. This could mean that the added contingency in a PIE classroom (from the students' perspective, the instructor could call on them at any time) was just as effective, if not more so than answering questions in class. Our analysis only includes data from the bottom performers, who showed the most significant reaction to the independent variable.

Future research could also examine to what extent the PIE contingency influenced other populations.

With regard to the dependent variables, the use of PIE appears to have enhanced student engagement and academic performance. Furthermore, based on self-report, the majority students who were exposed to PIE indicated that they found the classes helpful in increasing their interest and preparing them for their quiz (see Appendix G for the full list of questions, and Table 11 for the data). A greater proportion of students in Section 1 (59.1%) indicated that the instructor's questioning style helped with their engagement with the material than Section 2 students (49.8%). One potential benefit that can come from this study is to inform instructors anxious to call on students lest it be aversive. On the contrary, students reported favorably about being called upon and that they perceived PIE as having an effect on their performance in the course. It should be noted that the instructor feedback was relatively controlled. Students were given praise or acknowledgement when they answered correctly, and encouragement to reach the correct answer if they answered incorrectly. It is possible given a less amicable classroom environment calling on students randomly could become aversive.

Although some research has been done to examine the personality traits of instructors in relation to their effectiveness (e.g., Murray, Rushton, & Paunonen, 1990), more work is still required to identify objective teaching behaviors to influence instructor effectiveness. There is no doubt that some instructors, given their unique history, have a much better relationship with their students and as a result are more effective at reinforcing student engagement and performance. The unfortunate truth is that not all instructors have that quality, and need instruction on better classroom management and

establishing effective, positive contingencies to stimulate student participation. PIE as an intervention may be one step in that direction. Not only does PIE require a set amount of instructor prompted interaction, it also allows for unprompted responses on behalf of the student. The data suggest that in the majority of instances, the differences in instructor bore no significance in academic performance which could indicate that the PIE protocol promotes consistent teaching across instructors. Additionally, if instructors were to follow the same protocol, it would give them more specific guidelines on how to respond to student answers in a way that putatively reinforces their active responding (see Appendix E).

Using PIE is a relatively simple procedure involving little preparation on behalf of the instructor, no systematic rearrangements, and requires no budget. Besides using a seating chart and preparing questions and desk numbers to select students during class, the instructor's regular teaching routine would be largely uninterrupted. For the purpose of maintaining experimental control in the study, instructors were asked to refrain from asking their own additional questions, but there need not be such restriction when implementing this procedure outside of the experimental setting. In fact, the positive correlation between active responding and academic performance suggests that including more questions could potentially enhance the effect on outcome. The PIE intervention is easily adjustable to any classroom covering any material, making it a viable application for enhancing class lectures.

One may argue that class lectures should be eliminated given the research showing the benefits of such technologies as Personalized System of Instruction (PSI) and interteaching (Boyce & Hineline, 2002; McMichael & Corey, 1969; Pear & Crone-

Todd, 1999; Saville, Cox, O'Brien, & Vanderveldt, 2001). While strong empirical evidence supports the efficacy of many behavioral systems of instruction, pragmatically speaking, many instructors and educational institutions are either unable or unwilling to make drastic changes from the status quo of lecturing. Particularly in secondary education and higher education, instructor training is more focused on the subject material that they teach rather than teaching skills (Michael, 2006). Training thousands of teachers to completely change the way they conduct their classes is only possible with strong institutional contingencies (e.g., governmental policy). Thus, providing a kernel sized intervention, like PIE, would be a more reasonable approach to changing teacher behavior that would in turn enhance the student learning.

Disseminating PIE would not even necessarily require formal training. Future studies could examine how different methods of training instructors on an active responding kernel, like PIE, effects protocol integrity and student outcomes. Possibly, having instructors read a three page protocol may be just as effective as providing an hour long training. Implications from such a study would make the cost-benefit analyses of current active responding techniques more visible and powerful.

Not only could PIE be helpful for shaping effective instructor behaviors (e.g., calling on students, providing positive feedback), but the data suggest that PIE is effective at engaging a diverse population of students. There is an increasing diversity of students within the university system, and there is a greater variance of academic history and competence that is unprecedented (Merisotis & Phipps, 2000; Porter & Polikoff, 2012). As a larger body of students with varying ability levels enters universities, instructors need to be prepared to instruct in a manner that benefits all students. The

active responding literature has demonstrated learning gain in the classroom (Boyce & Hineline, 2002; Haggas & Hantula, 2002; Malanga & Sweeney, 2008; McMichael & Corey, 1969; Miller & Mallot, 1997; Neef, Perrin, Haberlin, & Rodrigues, 2011; Saville, Cox, O'Brien, & Vanderveldt, 2011; Tudor & Bostow, 1991). The current study adds to the active responding literature by specifying levels of performer within the classroom to identify how more or less successful students do in a PIE class versus a no PIE class.

There is a fine line between creating an environment that stimulates engagement for high-achieving students, while still providing enough support for students with weaker academic skills to participate. As mentioned earlier, the top performing participants in Section 1 actively responded about half as much as Section 2 top performers (Figure 15); however, academic performance was about equal for both groups (Figure 10 and 11). PIE demonstrates this "happy medium" in that the top 10% of performing participants in Section 1 performed just as well or better than Section 2 top performing participants for both end-of-lecture quizzes and chapter quizzes. Additionally, the bottom 10% of performing participants in Section 1 performed significantly better on end-of-lecture quizzes than Section 2 bottom performing participants (Figure 12); and over the course of the semester scored significantly higher on the chapter quizzes (Figure 13). Perhaps the active responding contingency that was established in the PIE classroom limited the excessive responding of the top performers, and increased the likelihood that low performers would respond in class. Such differences are important to consider when instructors are seeking a technology that would enhance bottom performer responding while not obstructing the learning progression of the top performing students.

Limitations

There are several limitations to this study that are worth noting. Firstly, the current study was limited in methodological rigor. Since the study was conducted in the applied setting of a university course, the students could not be randomly assigned to the two sections. Although the pretest conducted at the beginning of the semester revealed near identical averages between sections, such a limited baseline does not functionally demonstrate equal groups Furthermore, a between groups design cannot adequately confirm the degree to which the differences in outcome were due to the independent variable manipulation. Though the original research plan was to run a counterbalanced reversal design, due to the slow emergence of the meaningful differences between groups, there was no time to reverse the conditions within the constraints of the semester. With a counterbalanced reversal design, or something similar, there would be an opportunity to also examine within subject differences according to the presence or absence of PIE. Although we did find several significant differences between groups for both student engagement and academic performance, future studies should attempt to analyze PIE with a within subject design with a more extended baseline, or a reversal element, to demonstrate that differences in the dependent variable were due to the independent variable manipulation rather than preexisting differences between individuals.

One benefit to this study which helped the integrity of the data was that we took repeated measures of engagement and performance throughout the semester instead of having only a pretest, posttest comparison. Having multiple measures of the dependent variables shows a distinct pattern of emerging differences between groups. For example,

at the beginning of the semester, for both ratio of responding and the end-of-lecture quiz scores for the bottom performers, the data clearly show little to no difference between groups (see Figures 2 and 12). Over time, however, the differences between groups become increasingly apparent. Because all other controllable factors were held constant, the growing difference between groups can be attributed to the presence or absence of PIE, rather than initial group differences.

Another limitation was that students were simultaneously exposed to another behavioral technology to improve student performance. The introductory psychology course in which the study was run used a version of PSI which limited may have minimized potential effects of the independent variable for a few reasons. For one, students had minimal exposure to the independent variable. Students were required to attend only one 90 minute lecture each week for 10 weeks. Had students attended more or longer classes, more data could have been collected, and the different classroom contingencies (PIE or no PIE) may have established control sooner.

Secondly, because students were allowed to choose their classes on a weekly basis, students could potentially take classes on different days of the week, at different times of day, and with different instructors. Some did, and some stayed with the same class throughout the semester. Although we could not control for student attendance for specific classes, they were restricted to only attend classes for their section, meaning they would only be exposed to their condition. The schedule attempted to control for confounding variables by spreading classes for each section throughout the day and throughout the week and among all the instructors (see Appendix D).

Thirdly, there was no contingency within the course to restrict students from taking their chapter quizzes before they attended lecture, thus contacting the experimental condition. As such, the data for the chapter quizzes should be examined with some caution. It was not feasible to only track initial chapter quiz scores taken after a class session for 276 participants for each week. It is likewise unknown how frequently this occurs; but it has been informally observed that most students take their chapter quiz for the first time after they attend class. Nevertheless when examining the potential effects of PIE on academic performance, the end-of-lecture quiz scores are likely to be more descriptive than the chapter quizzes. Every student took the same end-of-lecture quiz at the time relative to attending class which is much more controlled than when students took their chapter quizzes.

Finally, this course has used a version of PSI for over 15 years and historically shown that students perform much better with this structure than other large enrollment courses (compared to class data before the switch to PSI). As a result of students already being immersed in a supportive classroom environment, there could have been a possible ceiling effect when manipulating the presence or absence of PIE. That is, the PSI contingencies already in place were likely having a stronger impact on student performance and are likely stronger than the contingencies surrounding PIE. Of course, both sections of students were exposed to the same PSI contingencies, so the structure of the course does not necessarily confound the data; however, greater effects may have emerged in a more “traditional” college classroom with fewer systematic supports to keep students engaged (e.g., fewer weekly quizzes, no formal contingency for attendance).

Future Directions

When considering possibilities for future research, it is important to include evaluations about the immediate efficacy of the intervention itself, the practicality of implementing the research, and developing the research area as a whole. In the following discussion of future research directions we recommend conducting studies to: (a) continue to validate PIE as a generalizable intervention, (b) evaluate the effectiveness of PIE compared with other active responding technologies, and (c) refine the existing knowledge base of active responding.

In light of the current study, it would be ideal to replicate and extend the PIE protocol in a more typical college classroom. Or, in order to parse out the various functional components of PIE, future studies could be conducted through an analogue study using a smaller sample size. Given that we conducted the study in a very particular educational setting using a version of PSI, future research in this area will need to establish the PIE protocol's generalizability to other settings. Doing so would help to determine how the PIE classroom influences engagement and performance without the distraction of potentially more potent contingencies (e.g. taking classes at different times every week, taking chapter quizzes before class). Using the current methodology in a more typical classroom would allow for greater experimental control over certain variables that was not possible with this study.

As stated earlier, any extensions of this study may find a faster impact of the PIE condition in a class that meets multiple times a week and/or for longer periods of time. Constrained to the course and university schedule, the present study was unable to reverse the conditions within each section for within subject analyses. Perhaps in an

analogue study or a year-long course may prove more conducive to working with a relatively slow acting variable and would allow for reversing conditions at least once in a counterbalanced design.

It is possible that the "slow acting" nature or weak stimulus control of the PIE condition may be strengthened by increasing the number of questions asked during a single class period. Increasing the number of questions asked in each condition may emphasize the independent variable's effect. In most classes for both sections, all 12 questions (PIE and voluntary) were answered. Our classroom seated 28 students, and when only asking 12 questions, the methodology may have led to students generating rules about the likelihood that they would be called upon.

Despite the wide array of active responding research, little has been said on the topic of practicality and cost-effectiveness. Clearly educational institutions, particularly in this country, are consistently underfunded and operate under tight budgets. If behavior scientists have any aspiration of making a meaningful impact on education with efficacious programs, further research is needed to identify the treatments that are the least disruptive to a current system. Ideally, if classrooms could grossly rearrange their system to incorporate the technologies already developed, then hypothetically vast improvements would occur in student outcomes. Practically speaking, teachers and students need reasonably effective treatments that can be quickly installed, sustainable, and inexpensive. Those interested in enhancing the impact of behavior science in schools may want to consider conducting cost-benefit analyses for various active responding technologies and various systems of instruction. Once known and clearly defined, it may not only alert educators to the extreme value of such interventions, but also guide

decision makers to consider only the most cost-effective programs to fit their needs. As mentioned earlier, taking a kernel-based approach to educational interventions, may be more effective for dissemination; however more discussion and research is needed before PIE, or other various active responding programs, can be considered a kernel.

Taking a broader view of the active responding research area as whole, much can still be done to find the specific kernels, mediators, or processes that are most effective under a specific set of circumstances. As previously noted, active responding literature is significant with almost every published account of active responding reports improved academic outcome over a control condition, or an approximated "traditional" classroom setting. The results of this study continue this trend, and make an attempt at analyzing the potential mediator, or process, by which these technologies function.

Now that there is a robust collection of effective education programs, further mediational analyses should be conducted to better understand the similarities between interventions. It has been previously discussed that by empirically determining functionally similar mediators among treatments, then the individuals will more effectively apply an intervention that is relevant to the particular problem (Kazdin, 2008). Although Kazdin made this argument in relation to clinical psychology, a similar argument may be made for educational programs. For example, the results of this study suggest that PIE is most effective at engaging the participation of lower performing students and may as a result have a greater impact on their academic performance. If a certain instructor has no problem with getting each student discussing issues, or potentially works with a smaller number of students, then implementing PIE protocol may not be effective as another intervention like interteaching.

Along this same line, more research is needed to determine under what conditions are specific active responding technologies effective. Especially since institutions of higher learning are increasingly adopting online systems to allow for distance learning courses, it stands to reason that certain interventions would not be conducive to an online format. PIE, for example may work for synchronized classes, where the instructors and students have a live class over webcams, but may not be effective for a-synchronized courses, where there is no live interaction. Future studies in active responding could also begin considering active responding with online courses.

With so many options for instructors and administrators to consider, and so little data to describe the process responsible for making an intervention effective, the educator's struggle is not surprising. Identifying processes responsible for the function of an intervention is critically important. For example, the effectiveness of PIE for bottom performers was partially mediated by active responding, but much of the variance of academic performance is largely unaccounted for. Future studies may consider evaluating potential mediators that have a potent effect on student engagement and academic performance. Conducting such studies may reveal that many effective interventions are mediated by the same process. Recognizing the functional similarities among education programs would promote parsimony and precision, yet another factor that could aid the dissemination of behavioral technologies.

Conclusions

The findings of the current study were an initial evaluation of a kernel based intervention, PIE, in relation to student engagement and academic performance. In terms of engagement, or active responding, PIE was very effective at not only maintaining a

high level of student participation during class, but also necessarily included a greater proportion of the students than in the no PIE, or voluntary, condition. These results can largely be attributed to the PIE protocol that required different students to answer questions within a class session. There were also a few significant effects demonstrated on academic performance, but only among the bottom 10% of performing participants. When considering the interaction between active responding and quiz scores, again for the bottom performers, the effects of PIE were most noticeable for student who did not talk during class. In the absence of PIE, the end-of-lecture quiz scores for students who did not say anything during class were significantly lower than the scores of students who did not vocally participate in the presence of PIE.

The PIE protocol appears to be most beneficial for students with a weaker behavioral repertoire for participating in class in a way that helps them succeed. As is, the PIE protocol is an adaptable kernel-based intervention that would be easy to implement at little to no cost on behalf of the instructor or educational institution. The small to medium effects on student outcome on behalf of PIE become more meaningful when taking the potential for dissemination into account.

Further studies are needed to better understand the efficacy of PIE, particularly within a more traditional classroom setting. The investigators also highly recommend that extensions in this area utilize more rigorous methodologies to establish clearer datasets reflecting the effects of PIE.

References

Abrami, P.C., Leventhal, L., & Perry, R.P. (1982). Educational seduction. *Review of educational research, 52*(3), 446-464.

Austin, J. (2000). *Behavioral approaches to college teaching.* Unpublished manuscript, Department of Psychology, Florida State University, Tallahassee, Florida.

Axelrod, S. (1992). Disseminating an effective educational technology. *Journal of Applied Behavior Analysis, 25*(1), 31-35.

Benjamin, L.T. (1988). A history of teaching machines. *American Psychologist, 43*(9), 703-712.

Boyce, T.E., & Hineline, P.H. (2002). Interteaching: A strategy for enhancing the user-friendline;ss of behavioral arrangements in the college classroom. *The Behavior Analyst, 25*(2), 215-226.

Brophy, J.E. (1979). Teacher behavior and student learning. *Educational Leadership, 37*(1), 33-45.

Cooper, P. (2001). Teacher strategies for effective intervention with students presenting social, emotional and behavioural difficulties: An international review. *European Journal of Special Needs Education, 26*(1), 71-86. doi: 10.1080/08856257.2011.543547

Embry, D.D., & Biglan, A. (2008). Evidence-based kernels: Fundamental units of behavioral influence. *Clinical Child and Family Psychology Review, 11,* 75-113. doi:10.1007/s10567-008-0036-x

Fernald, P.S., & Jordan, E.A. (1991). Programmed instruction versus standard text in introductory psychology. *Teaching of Psychology, 18*(4), 205-211.

Gauci, S.A., Dantas, A.M., Williams, D.A., & Kemm, R.E. (2009). Promoting student-centered active learning in lectures with a personal response system. *Advances in Physiology Education, 33*, 60-71. doi: 10.1152/advan.00109.2007

Griffin, M.M., & Griffin, B.W. (1998). An investigation of the effects of reciprocal peer tutoring on achievement, self-efficacy, and test anxiety. *Contemporary Educational Psychology, 23*, 298-311.

Gray, Peter. (2011). *Psychology*. New York, NY: Worth Publishers.

Haggas, A.M., & Hantula, D.A. (2002). Think or click? Student preference for overt vs. covert responding in web-based instruction. *Computers in Human Behavior, 18(2),* 165-172. doi:10.1016/S0747-5632(01)00041-3

Hake, R.R. (1997). 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. doi: 10.1119/1.18809

Hopson, L.M., & Holleran-Stelker, L.K. (2010). The effectiveness of adapted versions of an evidence-based prevention program in reducing alcohol use among alternative school students. *Children and Schools, 32*(2), 81-92.

Kazdin, A.E. (2007). Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology, 3,* 1-27. doi: 10.1146/annurev.clinpsy.3.022806.091432

Keislar, E.R., & McNeil, J.D. (1962). A comparison of two response modes in an autoinstructional program with children in the primary grades. *Journal of Educational Psychology, 53*(3), 127-131.

Kellum, K. K., Carr, J. E., Dozier, C. L. (2001). Response-card instruction and student learning in a college classroom. *Teaching of Psychology, 28*(2), 101–104.

Keller, F. S. (1968) Good-bye teacher. *Journal of Applied Behavior Analysis, 1,* 78-89.

Keough, S.M. (2012). Clickers in the classroom: A review and a replication. *Journal of Management Education, 36,* 822-847. doi: 10.1177/1052562912454808

Kulik, J.A., Cohen, P.A., & Ebeling, B.J. (1980). Effectiveness of programmed instruction in higher education: A meta-analysis of findings. *Educational Evaluation and Policy Analysis, 2*(6), 51-64.

Kulik, C-L.C., Kulik, J.A., & Bangert-Drowns, R.L. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research, 60*(2), 265-299.

Kulik, J.A., Kulik, C-L.C., & Cohen, P.A. (1980). Effectiveness of computer-based college teaching: A meta-analysis of findings. *Review of Educational Research, 50*(4), 525-544.

Lindsley, O. R. (1971). From Skinner to precision teaching: The child knows best. In J. B. Jordan and L. S. Robins (Eds.), *Let's try doing something else kind of thing: Behavior principles and the exceptional child.* Reston, VA: The Council of Exceptional Children.

Malanga, P.R., & Sweeney, W.J. (2007). Increasing active student responding in a university applied behavior analysis course: The effect of daily assessment and response cards on end of week quiz scores. *Journal of Behavioral Education, 17,* 187-199. doi:10.1007/s10864-007-9056-8

Martinez, J.G.R., & Martinez, N.C. (1988). “Hello, Teacher”: An argument for reemphasizing the teacher’s role in PSI and mastery learning. *American Journal of Education, 97,* 18-33.

McMichael, J.S., & Corey, J.R. (1969). Contingency management in an introductory psychology course produces better learning. *Journal of Applied Behavior Analysis, 2*(2), 79-83.

Merisotis, J.P., & Phipps, R.A. (2000). Remedial education in colleges and universities: What’s really going on? *The Review of Higher Education, 24,* 67-85. doi: 10.1353/rhe.2000.0023

Michael, J. (2006). Where’s the evidence that active learning works? *Advances in Physiology Education, 30,* 159-167. doi: 10.1152/advan.00053.2006

Miller, M.L., & Malott, R.W. (1997). The importance of overt responding in programmed instruction even with added incentives for learning. *Journal of Behavioral Education, 7,* 497-503. doi:10.1023/A:1022811503326

Murray, H.G., Rushton, J.P., &Paunonen, S.V. (1990). Teacher personality traits and student instructional ratings in six types of university courses. *Journal of Educational Psychology, 82*(2), 250-261.

Naftulin, D.H., Ware, J.E., & Donnelly, F.A. (1973). The Doctor Fox lecture: A paradigm of educational seduction. *Journal of Medical Education, 48,* 630-635.

Neef, N.A., McCord, B.E., & Ferreri, S.J. (2006). Effects of guided notes versus completed notes during lectures on college students’ quiz performance. *Journal of Applied Behavior Analysis, 39*(1), 123-130. doi: 10.1901/jaba.2006.94-04

Neef, N.A., Perrin, C.J., Haberlin, A.T., & Rodrigues, L.C. (2011). Studying as fun and games: Effects on college students' quiz performance. *Journal of Applied Behavior Analysis, 44*(4), 897-901. doi: 10.1901/jaba.2011.44-897

Neef, N.A., Van Norman, R., & Ferreri, S.J. (2006). Comparison of interteaching and lecture in the college classroom. *Journal of Applied Behavior Analysis, 39*(1), 49-61. doi: 10.1901/jaba.2006.42-05

Pear, J.J., & Crone-Todd, D.E. (1999). Personalized system of instruction in cyberspace. *Journal of Applied Behavior Analysis, 32*(2), 205-209.

Porter, A.C., & Polikoff, M.S. (2012). Measuring academic readiness for college. *Educational Policy, 26,* 394-417. doi: 10.1177/0895904811400410

Russell, I.J., Caris, T.N., Harris, G.D., & Hendricson, W.D. (1983). Effects of three types of lecture notes on medical student achievement. *Journal of Medical Education, 58*(8), 627-636.

Saville, B.K., Cox, T., O'Brien, S., & Vanderveldt, A. (2011). Interteaching: The impact of lectures on student performance. *Journal of Applied Behavior Analysis, 44*(4), 937-941. doi: 10.1901/jaba.2011.44-937

Semb, G. (1974). The effects of mastery criteria and assignment length on college-student test performance. *Journal of Applied Behavior Analysis, 7*(1), 61-69.

Semb, G., Hopkins, B.L., & Hursh, D.E. (1973). The effects of study questions and grades on student test performance in a college course. *Journal of Applied Behavior Analysis, 6*(4), 631-642.

Shadish, W.P., Cook, T.D., & Campbell, D.T. (2002). *Experimental and Quasi-experimental Designs for Generalized Causal Inference.* Houghton Mifflin Co.: Boston.

Skinner, B. F. (1954). The science of learning and the art of teaching. *Harvard Educational Review, 24,* 86-98.

Skinner, B. F. (1958). Teaching machines. *Science, 128,* 969-977.

Solomon, D., Rosenberg, L., & Bezdek, W.E. (1964). Teacher behavior and student learning. *Journal of Educational Psychology, 55*(1), 23-30.

Sparzo, F.J. (1992). B.F. Skinner's contributions to education: A retrospective appreciation. *Contemporary Education, 63*(3), 225-233.

Sutherland, K.S., & Wehby, J.H. (2001). Exploring the relationship between increased opportunities to respond to academic requests and the academic and behavioral outcomes of students with EBD: A review. *Remedial and Special Education, 22*(2), 113-121.

Thompson, S. B. (1980). Do individualized mastery and traditional instructional systems yield different course effects in college calculus?. *American Educational Research Journal, 17*(3), 361-375. doi:10.2307/1162621

Tudor, R.M., & Bostow, D.E. (1991). Computer-programmed instruction: The relation of required interaction to practical application. *Journal of Applied Behavior Analysis, 24(2), 361-368.* doi:10.1901/jaba.1991.24-361

Vargas, E.A., & Vargas, J.S. (1991). Programmed instruction: What it is and how to do it. *Journal of Behavioral Education, 1*(2), 235-251.

Tables

Table 1

Average ratios of responding for each instructor for each week over ten weeks of instruction

	Section	1	2	3	4	5	6	7	8	9	10
TA1	1	1.00	0.92	0.75	0.92	0.92	0.83	[a]	[a]	0.83	0.77
	2	0.96	0.86	0.72	0.65	0.67	0.64	0.87	0.79	0.83	0.81
TA2	1	0.84	0.86	0.81	0.87	0.83	0.92	0.84	0.85	0.96	0.70
	2	0.87	0.85	0.87	0.83	0.78	0.81	0.70	0.81	0.60	0.67
TA3	1	0.74	0.85	0.84	0.72	0.81	0.81	0.86	0.85	0.83	0.80
	2	0.78	0.66	0.74	0.68	0.57	0.54	0.58	0.70	0.55	0.47
TA4	1	0.90	0.84	0.85	0.73	0.58	0.80	0.77	0.77	0.61	0.80
	2	0.88	0.89	0.89	0.67	0.57	0.64	0.59	0.71	0.51	0.57
Total	1	0.85	0.87	0.83	0.78	0.73	0.83	0.83	0.83	0.74	0.79
	2	0.87	0.82	0.82	0.72	0.67	0.66	0.68	0.76	0.61	0.63

Note. Ratio of responding was calculated by dividing the number of students who responded during one class session by the number of total responses for that class session. Data above are the average ratios for each instructor for each week.

[a] Instructor was absent for the section 1 sessions for weeks 7 and 8.

Table 2

ANOVAs Comparing Section 1 and 2 End-of-Lecture Quiz Scores over Ten Weeks of Instruction

Week		SS	df	MS	*F*	*p*	η^2	*d*
1	Between	0.23	1	0.23	0.30	0.59	0.00	0.3
	Within	209.91	271	0.78				
2	Between	0.90	1	0.90	1.01	0.32	0.00	0.3
	Within	239.64	271	0.88				
3	Between	5.96	1	5.96	4.26	0.04*	0.02	0.32
	Within	379.14	271	1.40				
4	Between	0.66	1	0.66	0.54	0.46	0.00	0.3
	Within	330.35	271	1.22				
5	Between	0.92	1	0.92	0.65	0.42	0.00	0.86
	Within	383.02	271	1.41				
6	Between	1.28	1	1.28	0.72	0.40	0.00	0.34
	Within	482.00	271	1.78				
7	Between	6.83	1	6.83	3.13	0.08	0.01	-0.48
	Within	592.13	271	2.19				
8	Between	6.51	1	6.51	2.93	0.09	0.01	0.28
	Within	601.48	271	2.22				
9	Between	1.70	1	1.70	0.99	0.32	0.00	0.27
	Within	464.55	271	1.71				
10	Between	0.63	1	0.63	0.57	0.45	0.00	0.66
	Within	299.29	271	1.10				

Note. SS = Sum of Squares, MS = Mean Square. * Significance at $p < .05$

Table 3

ANOVAs Comparing Section 1 and 2 Chapter Quiz Scores over Ten Weeks of Instruction

Week		SS	df	MS	*F*	*p*	η^2
1	Between	1.00	1	1.00	0.20	0.66	0.00
	Within	1363.71	271	5.03			
2	Between	3.16	1	3.16	0.42	0.52	0.00
	Within	2052.71	270	7.60			
3	Between	4.48	1	4.48	0.69	0.41	0.00
	Within	1730.22	266	6.51			
4	Between	2.82	1	2.82	0.50	0.48	0.00
	Within	1496.48	266	5.63			
5	Between	4.01	1	4.01	0.53	0.47	0.00
	Within	2007.80	267	7.52			
6	Between	0.41	1	0.41	0.09	0.77	0.00
	Within	1262.21	264	4.78			
7	Between	0.28	1	0.28	0.04	0.84	0.00
	Within	1829.09	261	7.01			
8	Between	5.56	1	5.56	0.66	0.42	0.00
	Within	2219.22	264	8.41			
9	Between	12.47	1	12.47	2.19	0.14	0.01
	Within	1500.24	264	5.68			
10	Between	26.86	1	26.86	3.67	0.06	0.01
	Within	1955.71	267	7.33			

Note. SS = Sum of Squares, MS = Mean Square.

Table 4

ANOVAs Comparing Section 1 and 2 End-of-Lecture Quiz Scores among the Top 10% of Performing Participants over Ten Weeks of Instruction

Week		SS	df	MS	*F*	*p*	η^2
1	Between	0.00	1	0.00	0.00	1.00	0.00
	Within	1.86	26	0.07			
2	Between	0.00	1	0.00	0.00	1.00	0.00
	Within	6.86	26	0.26			
3	Between	0.04	1	0.04	0.18	0.68	0.01
	Within	5.21	26	0.20			
4	Between	0.04	1	0.04	0.18	0.68	0.01
	Within	5.21	26	0.20			
5	Between	0.57	1	0.57	0.68	0.42	0.03
	Within	21.86	26	0.84			
6	Between	0.00	1	0.00	0.00	1.00	0.00
	Within	8.43	26	0.32			
7	Between	0.04	1	0.04	0.15	0.70	0.01
	Within	6.07	26	0.23			
8	Between	0.00	1	0.00	0.00	1.00	0.00
	Within	3.43	26	0.13			
9[a]	Between						
	Within						
10	Between	0.00	1	0.00	0.00	1.00	0.00
	Within	3.43	26	0.13			

Note. SS = Sum of Squares, MS = Mean Square.

[a]. Test could not be computed for Week 9 because there was no variance between groups.

Table 5

ANOVAs Comparing Section 1 and 2 Chapter Quiz Scores among the Top 10% of Performing Participants over Ten Weeks of Instruction

Week		SS	df	MS	*F*	*p*	η^2
1	Between	1.29	1	1.29	0.34	0.57	0.01
	Within	98.57	26	3.79			
2	Between	0.89	1	0.89	0.18	0.67	0.01
	Within	128.36	26	4.94			
3	Between	2.29	1	2.29	0.43	0.52	0.02
	Within	139.14	26	5.35			
4	Between	0.32	1	0.32	0.06	0.81	0.00
	Within	136.64	26	5.26			
5	Between	7.00	1	7.00	1.12	0.30	0.04
	Within	162.00	26	6.23			
6	Between	1.75	1	1.75	0.36	0.55	0.01
	Within	126.36	26	4.86			
7	Between	4.32	1	4.32	0.71	0.41	0.03
	Within	158.64	26	6.10			
8	Between	0.89	1	0.89	0.13	0.72	0.01
	Within	176.07	26	6.77			
9	Between	0.00	1	0.00	0.00	1.00	0.00
	Within	87.86	26	3.38			
10	Between	12.89	1	12.89	3.49	0.07	0.12
	Within	96.07	26	3.70			

Note. SS = Sum of Squares, MS = Mean Square.

Table 6

ANOVAs Comparing Section 1 and 2 End-of-Lecture Quiz Scores among the Bottom 10% of Performing Participants

Week		SS	df	MS	*F*	*p*	η^2	*d*
1	Between	0.00	1	0.00	0.00	1.00	0.00	0
	Within	14.71	26	0.57				
		14.71	27					
2	Between	0.36	1	0.36	0.87	0.36	0.03	0.36
	Within	10.31	25	0.41				
		10.67	26					
3	Between	1.19	1	1.19	1.20	0.29	0.05	0.46
	Within	20.73	21	0.99				
		21.91	22					
4	Between	1.50	1	1.50	2.20	0.15	0.09	0.59
	Within	15.00	22	0.68				
		16.50	23					
5	Between	0.15	1	0.15	0.13	0.73	0.01	0.14
	Within	29.23	24	1.22				
		29.39	25					
6	Between	1.52	1	1.52	1.13	0.30	0.05	0.44
	Within	28.30	21	1.35				
		29.83	22					
7	Between	2.13	1	2.13	1.95	0.18	0.10	0.62
	Within	19.67	18	1.09				
		21.80	19					
8	Between	4.51	1	4.51	3.82	0.06†	0.15	0.77
	Within	24.80	21	1.18				
		29.30	22					
9	Between	4.55	1	4.55	4.90	0.04*	0.20	0.87
	Within	18.55	20	0.93				
		23.09	21					
10	Between	7.54	1	7.54	9.88	0.00**	0.29	1.06
	Within	18.31	24	0.76				

Note. SS = Sum of Squares, MS = Mean Square. † = near significant; * = significant at $p < .05$; ** = significant at $p < .001$

Table 7

ANOVAs Comparing Section 1 and 2 Chapter Quiz Scores among the Bottom 10% of Performing Participants

Week		SS	df	MS	*F*	*p*	η^2	*d*
1	Between	3.57	1	3.57	0.63	0.44	0.02	0.30
	Within	148.43	26	5.71				
2	Between	5.54	1	5.54	0.59	0.45	0.02	0.30
	Within	224.00	24	9.33				
3	Between	1.36	1	1.36	0.54	0.47	0.03	0.32
	Within	50.50	20	2.53				
4	Between	2.04	1	2.04	0.60	0.45	0.02	0.30
	Within	84.48	25	3.38				
5	Between	16.67	1	16.67	5.19	0.03*	0.19	0.86
	Within	70.67	22	3.21				
6	Between	2.69	1	2.69	0.69	0.41	0.03	0.34
	Within	85.32	22	3.88				
7	Between	6.83	1	6.83	1.32	0.26	0.06	-0.48
	Within	108.91	21	5.19				
8	Between	2.91	1	2.91	0.44	0.51	0.02	0.28
	Within	144.42	22	6.57				
9	Between	2.04	1	2.04	0.42	0.53	0.02	0.27
	Within	107.58	22	4.89				
10	Between	8.91	1	8.91	2.61	0.12	0.12	0.66
	Within	68.36	20	3.42				

Note. SS = Sum of Squares, MS = Mean Square. * = significant at $p < .05$

Table 8

ANOVAs Comparing Instructors with End-of-Lecture Quizzes over Ten Weeks

Week		SS	df	MS	*F*	*p*
1	Between	0.28	3	0.09	0.34	0.8
	Within	74.24	265	0.28		
2	Between	3.13	3	1.04	2.86	0.04
	Within	96.48	264	0.36		
3	Between	0.691	3	0.23	0.41	0.75
	Within	147.4	261	0.56		
4	Between	0.96	3	0.32	0.64	0.59
	Within	131	262	0.5		
5	Between	4.022	3	1.34	1.53	0.21
	Within	230.34	263	0.88		
6	Between	2.5	3	0.83	1.14	0.33
	Within	188.33	258	0.73		
7	Between	1.05	3	0.35	0.52	0.67
	Within	171.15	253	0.68		
8	Between	1.54	3	0.51	0.83	0.48
	Within	154.55	252	0.61		
9	Between	1.31	3	0.43	1.31	0.27
	Within	85.51	257	0.33		
10	Between	6.46	3	2.15	3.15	0.03
	Within	182.55	265	0.69		

Note. SS = Sum of Squares, MS = Mean Square. Testing the differences among instructors in relation to quiz scores. Lack of significance suggests the presence of experimental control; wherein, there were no significant differences in quiz scores between the different instructors.

Table 9

ANOVAs Comparing Instructors with Chapter Quizzes over Ten Weeks

Week		SS	df	MS	*F*	*p*
1	Between	29.36	3	9.79	1.96	0.12
	Within	1320.17	265	4.98		
2	Between	3.57	3	1.19	0.19	0.90
	Within	1662.84	264	6.3		
3	Between	88.8	3	29.6	5.93	0.00
	Within	1288.27	258	5		
4	Between	15.82	3	5.27	0.98	0.40
	Within	1403.44	260	5.4		
5	Between	25.27	3	8.42	1.25	0.29
	Within	1775.02	263	6.75		
6	Between	0.84	3	0.28	0.06	0.98
	Within	1228.98	255	4.82		
7	Between	20.75	3	6.92	1.11	0.34
	Within	1555.22	250	6.22		
8	Between	4.36	3	1.45	0.21	0.89
	Within	1731.05	250	6.92		
9	Between	19.24	3	6.41	1.34	0.26
	Within	1219.19	255	4.78		
10	Between	46.34	3	15.45	2.24	0.08
	Within	1805.92	262	6.89		

Note. SS = Sum of Squares, MS = Mean Square. Testing the differences among instructors in relation to quiz scores. Lack of significance suggests the presence of experimental control; wherein, there were no significant differences in quiz scores between the different instructors.

Table 10

Multiple Regression: Relation between Independent Variable Manipulation (PIE) and Active Responding on End-of-Lecture Quiz Scores from the Bottom 10% Performing Participants

Model		Unstandardized Coefficients		Standardized Coefficients	*t*	*p*
		B	Std. Error	Beta		
1	(Constant)	5.88	.19		30.80	< .001
	Section	-.54	.12	-.28	-4.51	< .001
2	(Constant)	5.72	.19		29.33	< .001
	Section	-.52	.12	-.27	-4.44	< .001
	AR	.22	.07	.19	3.06	.002
$R^2 = .11$, $F(2, 239) = 15.21$, $p < .001$						

Note. Dependent variable: end-of-lecture quiz scores; Data reflects quiz scores for each participant within the sample over the course of ten weeks of instruction. The negative relation for section implies that Section 1 students scored better than Section 2 students.

Table 11

Social Validity Analysis: Comparing Section 1 and 2 Responses on Course Evaluation

Q #	Question Content	Very[a], Increased[b], More[c], Yes[d]		Moderate[a], No Effect[b], Same[c]		Not[a], Decreased[b], Less[c], No[d]	
		Sect 1 (PIE)	Sect 2	Sect 1 (PIE)	Sect 2	Sect 1 (PIE)	Sect 2
2	Work required compared to other classes	15.6[c]	21.1[c]	62.6[c]	58.2[c]	21.7[c]	20.7[c]
3	Enjoyment compared to other classes	32.2[c]	29.9[c]	47.4[c]	47.9[c]	20.4[c]	22.2[c]
4	Individual attention compared to other classes	33.5[c]	31.0[c]	46.5[c]	47.9[c]	19.6[c]	21.1[c]
8	Found classes helpful, increased interest	63.9[a]	60.9[a]	29.1[a]	31.8[a]	4.8[a]	5.0[a]
10	Found classes helpful, quiz preparation	60.4[a]	56.7[a]	33.0[a]	33.3[a]	6.1[a]	8.4[a]
13	Recommend course to a friend	80.4[d]	71.3[d]			20[d]	26.4[d]
14	Learn this way again	63.8[d]	54.0[d]			35.7[d]	44.8[d]
16	Question style affect performance	48.3[b]	44.4[b]	38.3[b]	39.9[b]	12.6[b]	14.6[b]
17	Question style affect engagement	59.1[b]	49.8[b]	32.6[b]	35.6[b]	8.3[b]	14.2[b]

Note. For full list of course evaluation, see Appendix F.

Figures

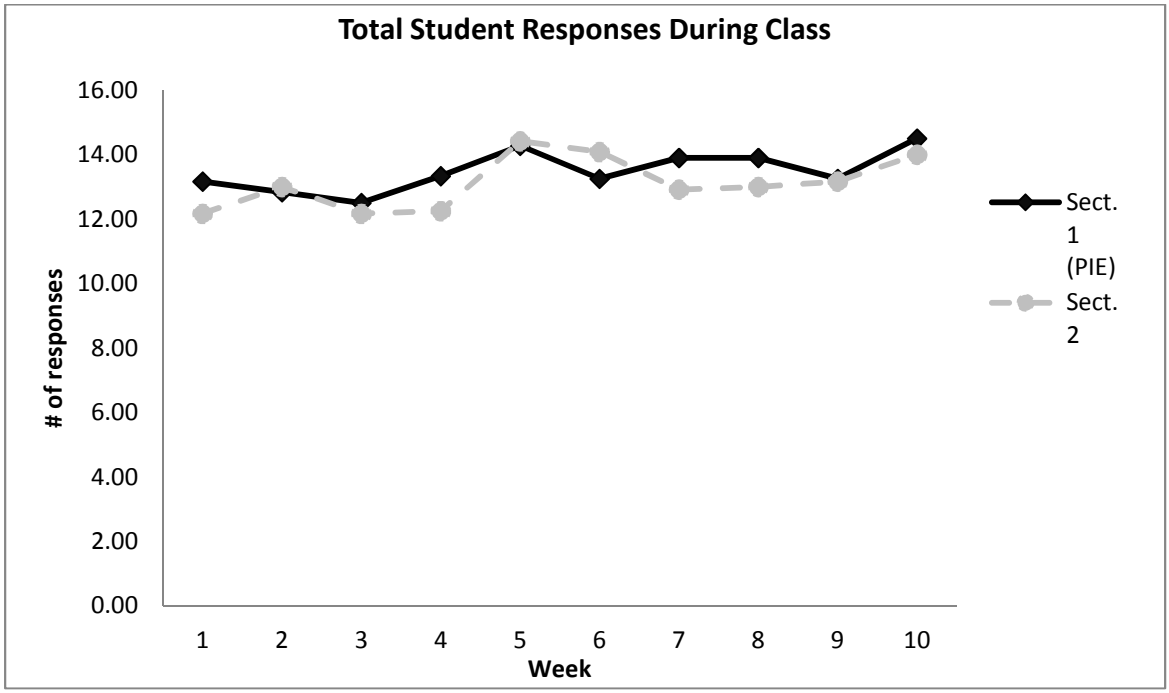


Figure 1. Average number of student responses per class for Section 1 and Section 2 across ten weeks. Twelve questions were asked by the instructor for each class for each section.

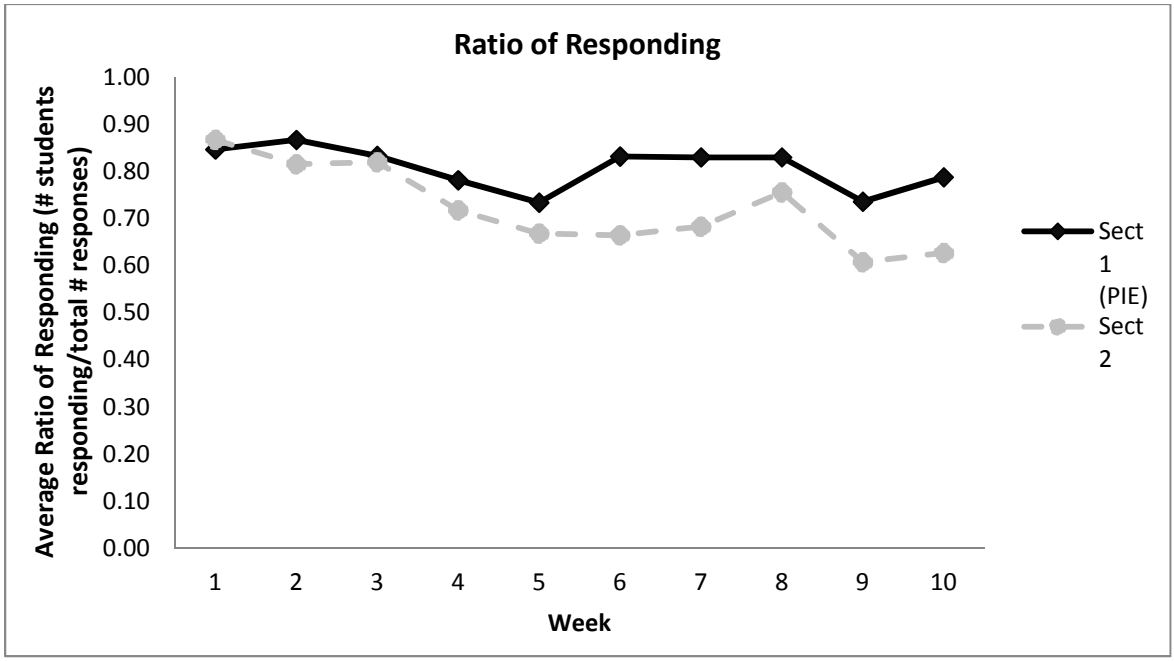


Figure 2. Average ratio (# of students responding divided by the total # of responses during class) of responding per class for Sections 1 and 2 across ten weeks.

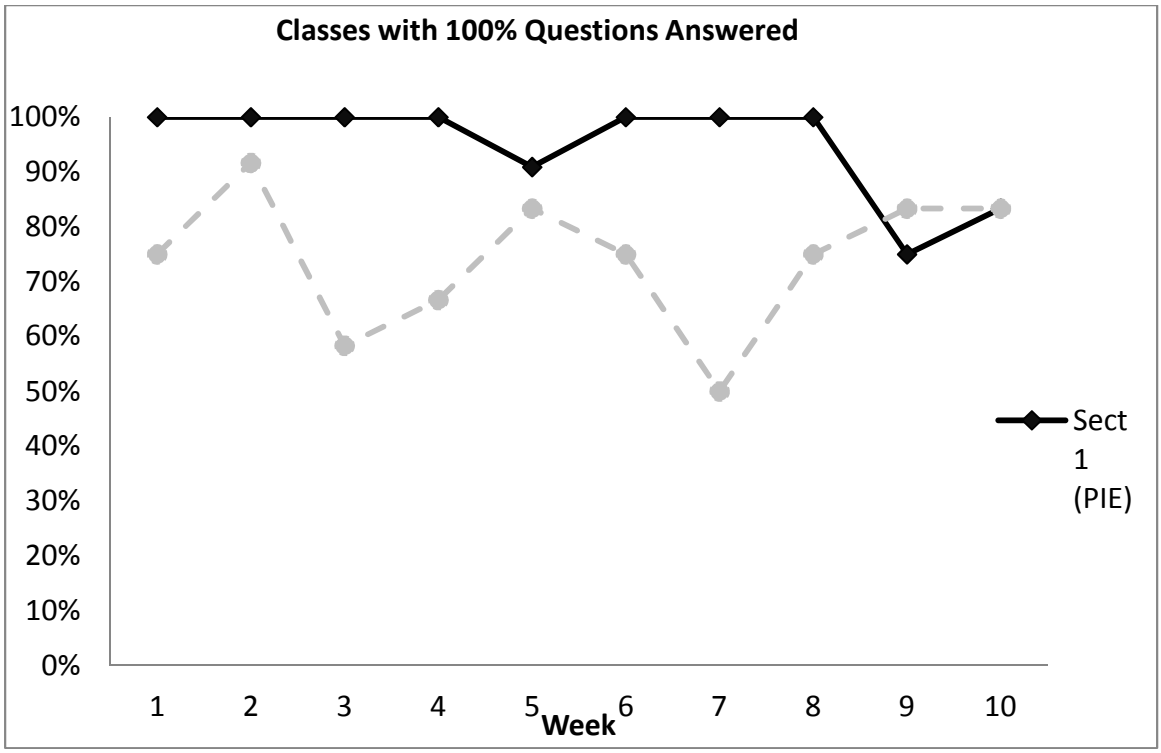


Figure 3. Percentage of Section 1 and 2 classes with 100% of instructor prompted questions answered over ten weeks of instruction.

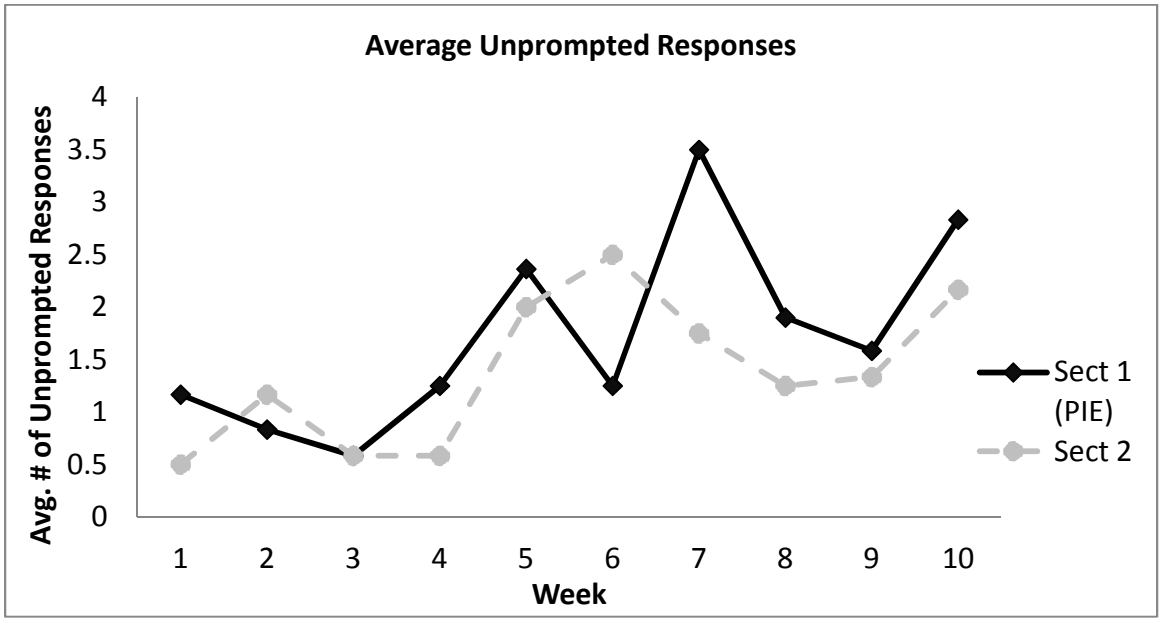


Figure 4. Average number of unprompted responses for Section 1 and 2 classes over ten weeks of instruction.

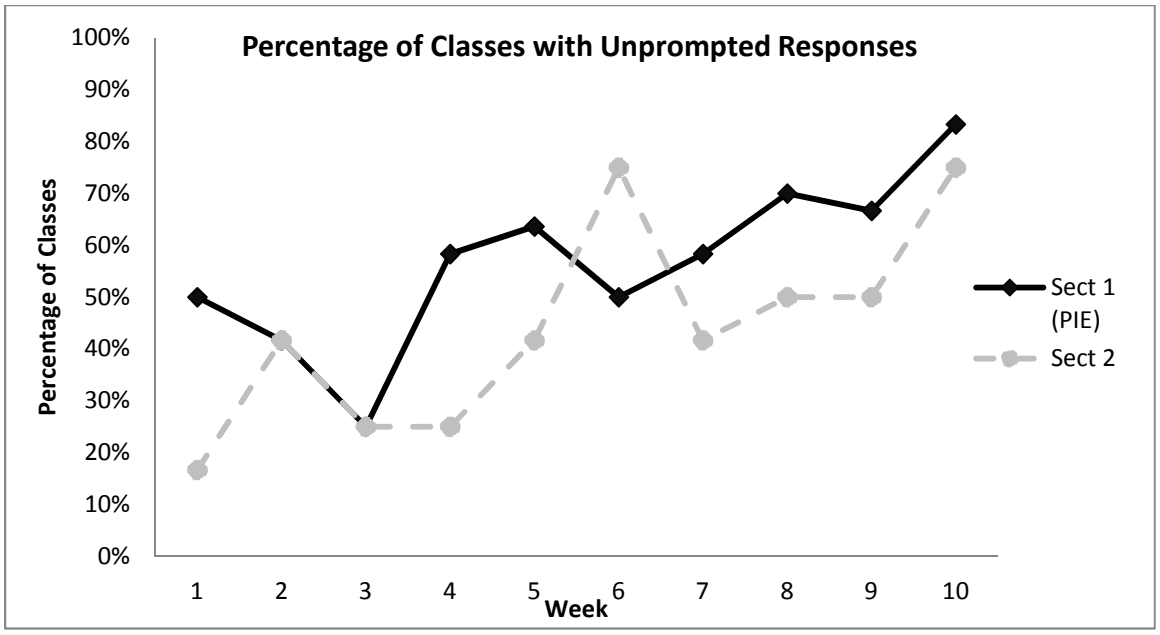


Figure 5. Percentage of classes where students responded without a prompt (e.g., asked their own question) for Sections 1 and 2 across ten weeks.

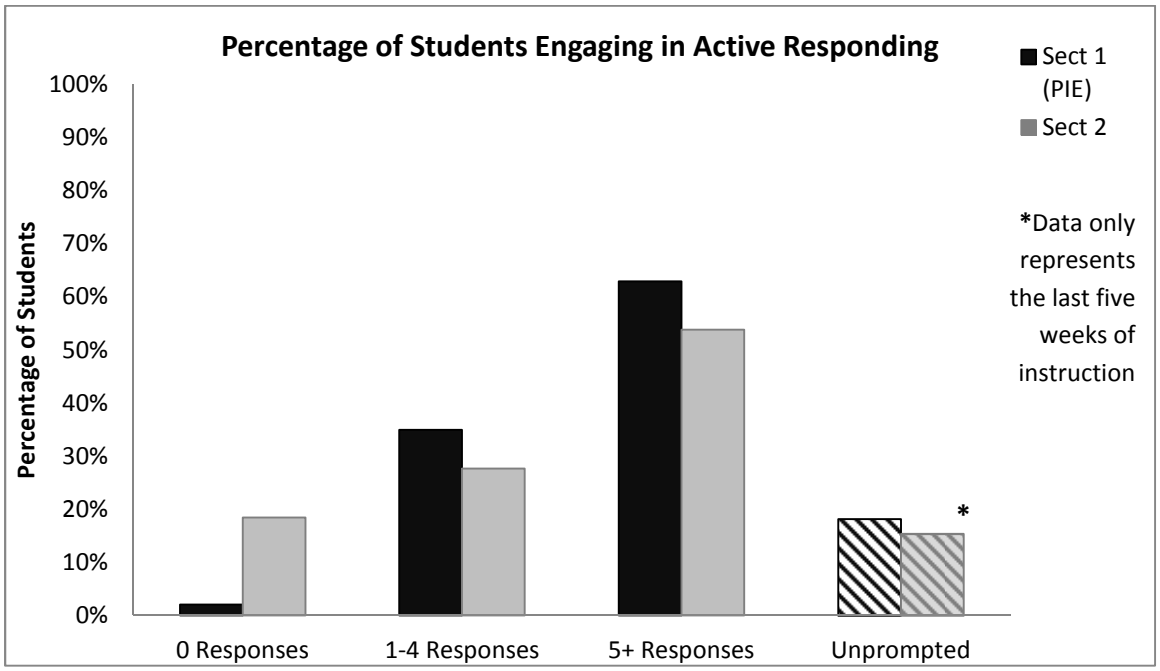


Figure 6. Percentage of students engaging in two topographies of active responding (instructor prompted and unprompted). Solid colored columns indicate the level of active responding to instructor prompts (PIE or voluntary questions) across ten weeks of instruction for all participants. The striped columns represent the percentage of students who made at least one unprompted question or comment between weeks 6 and 10.

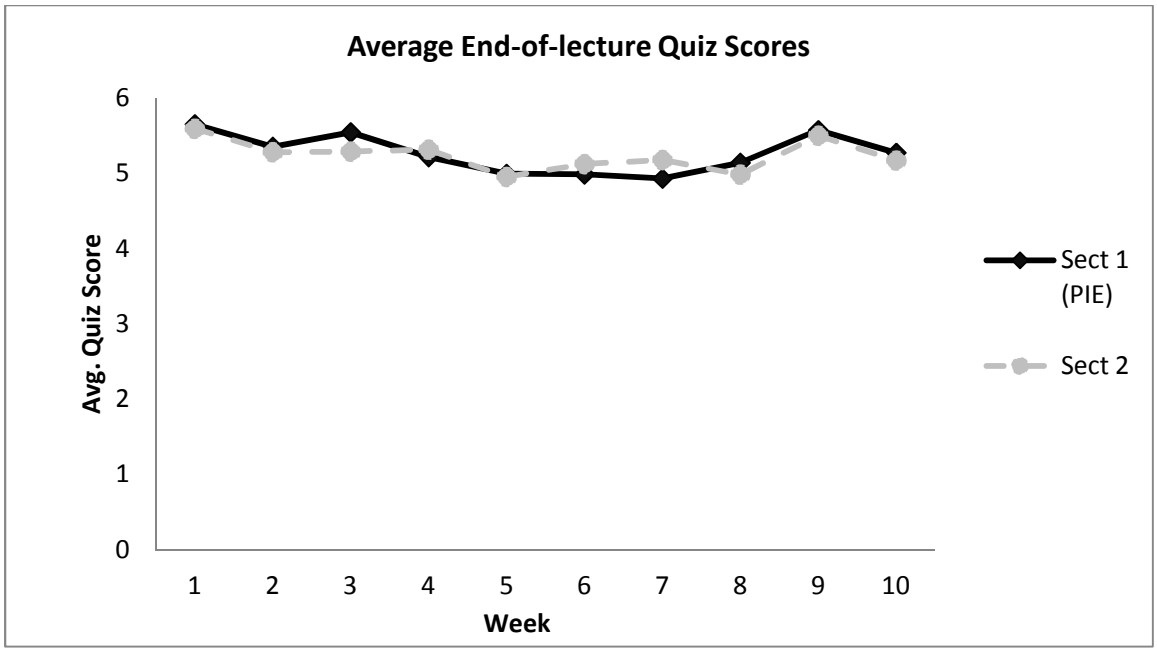


Figure 7. Average end-of-lecture quiz scores across sections over ten weeks of instruction.

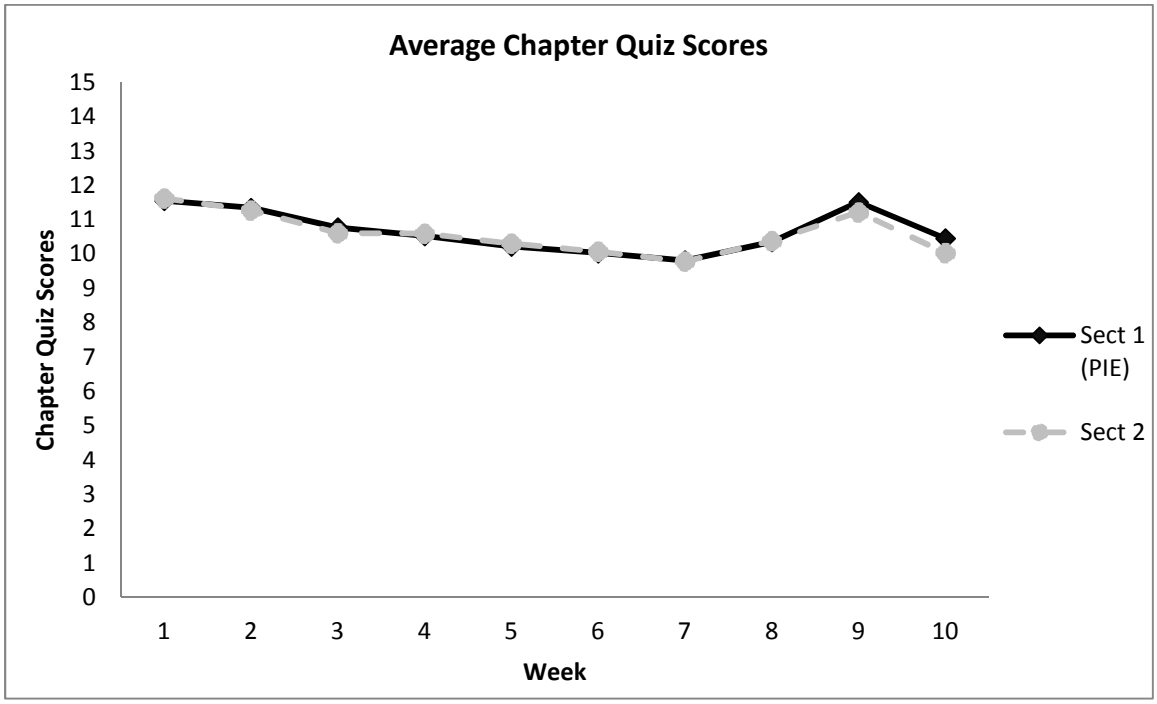


Figure 8. Average chapter quiz scores are compared across sections for ten weeks of instruction.

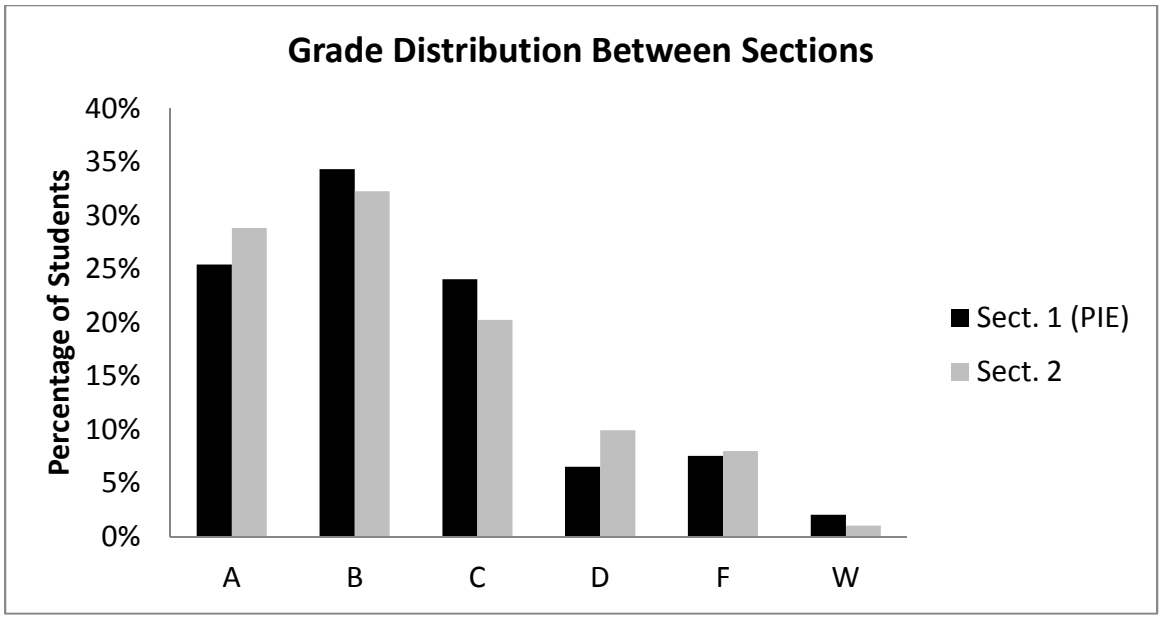


Figure 9. Overall grade distribution for all assignments and assessments between Section 1 and Section 2 students.

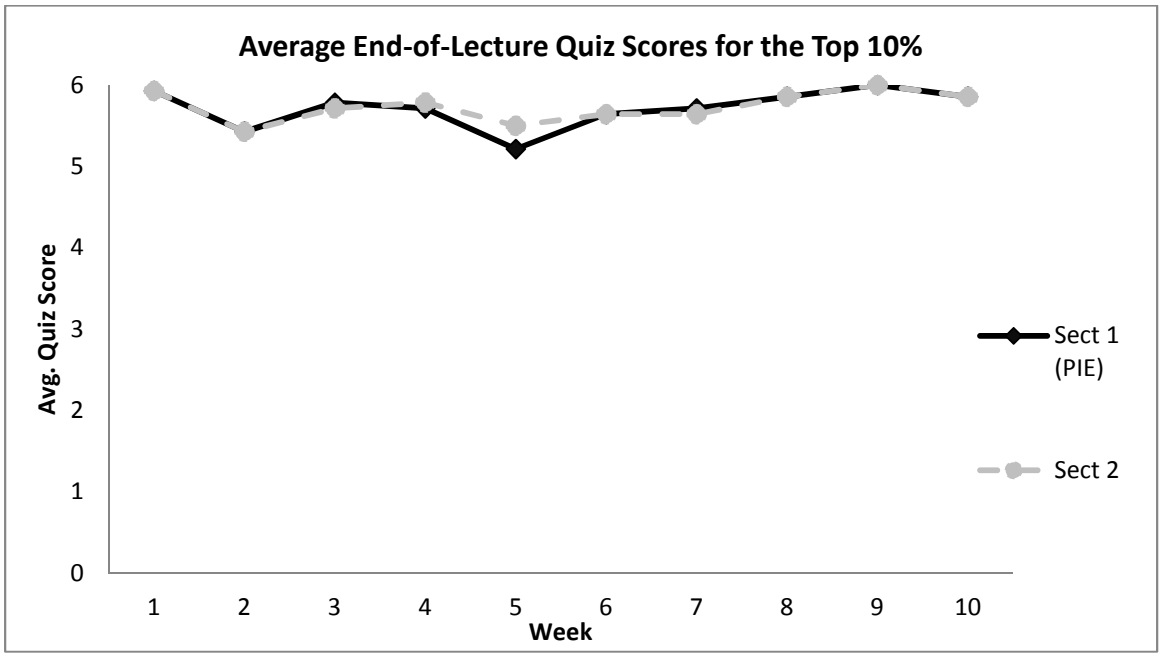


Figure 10. Average end-of-lecture quizzes among the top (10%) performers compared across sections for ten weeks of instruction.

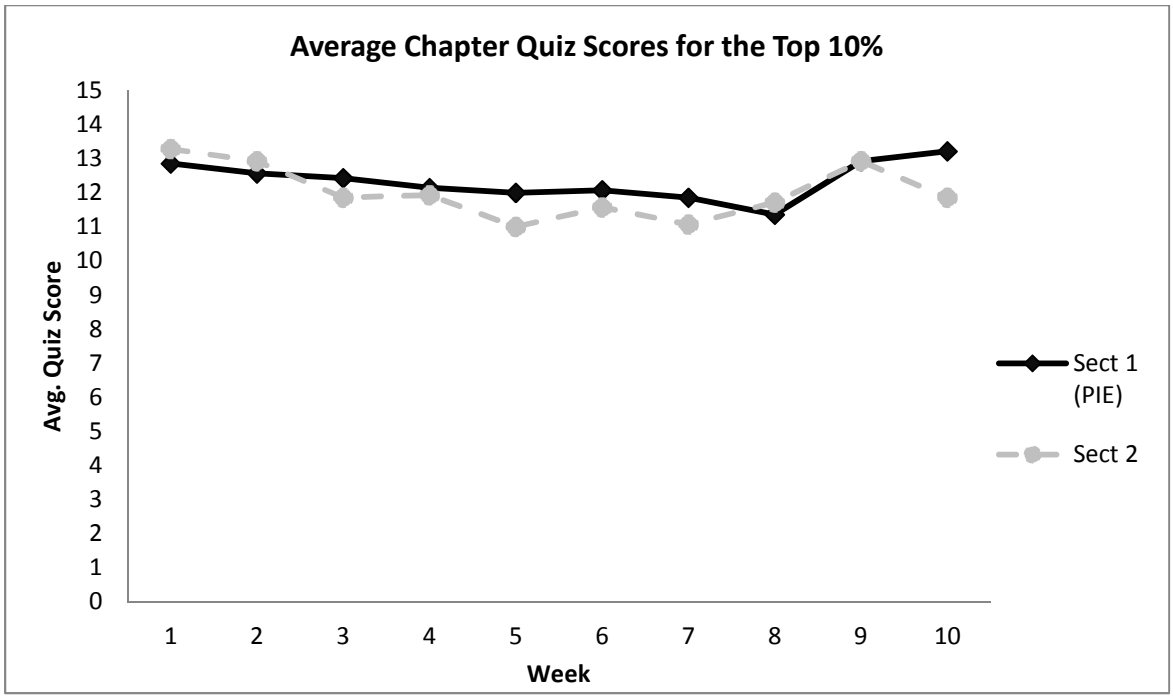


Figure 11. Average chapter quiz scores among the top (10%) performers compared across sections for ten weeks of instruction.

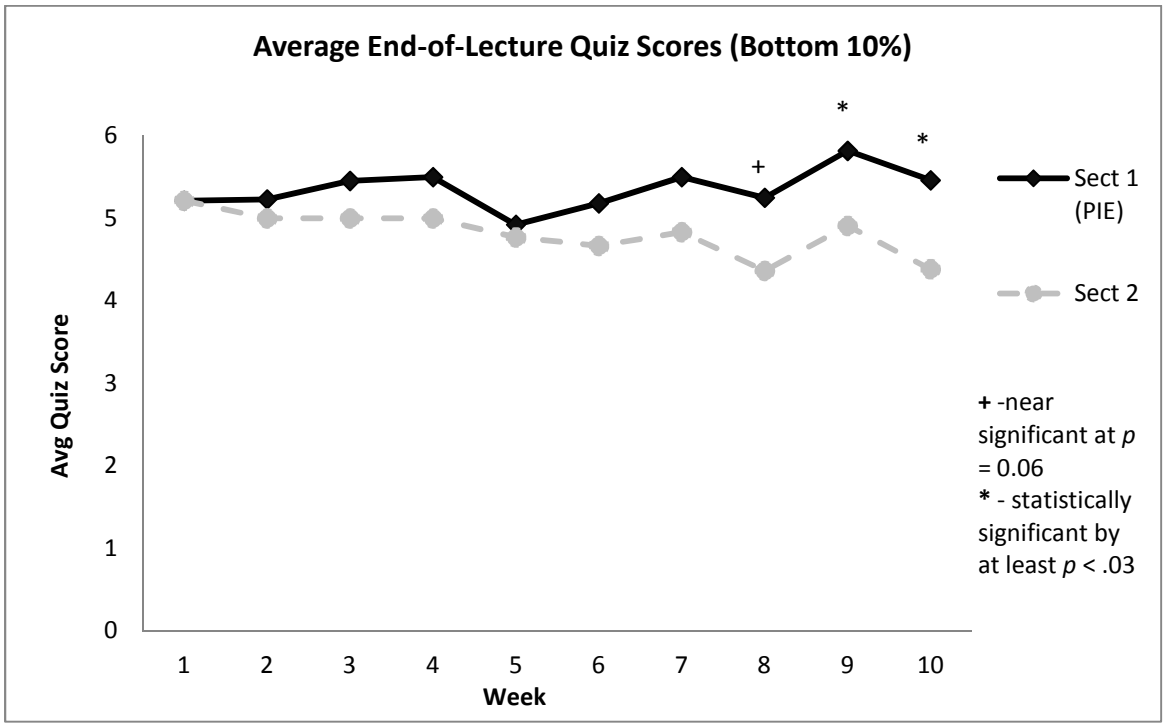


Figure 12. Average end-of-lecture quiz scores among the bottom (10%) performers compared across sections for ten weeks of instruction.

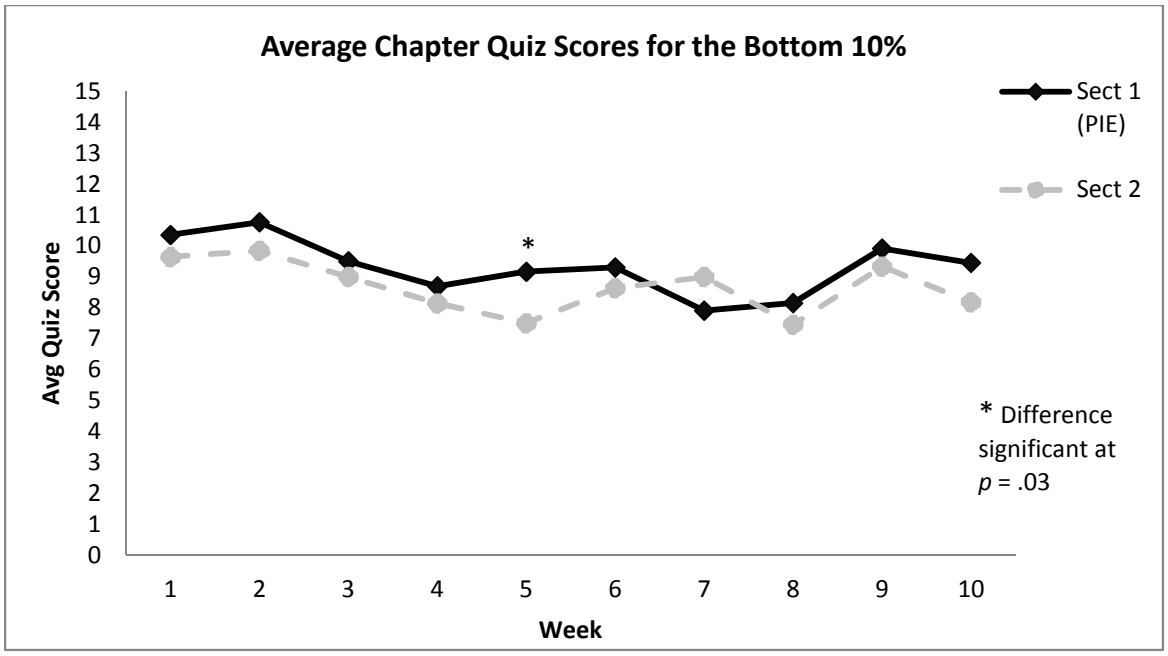


Figure 13. Average chapter quiz scores among the bottom (10%) performers compared across sections for ten weeks of instruction.

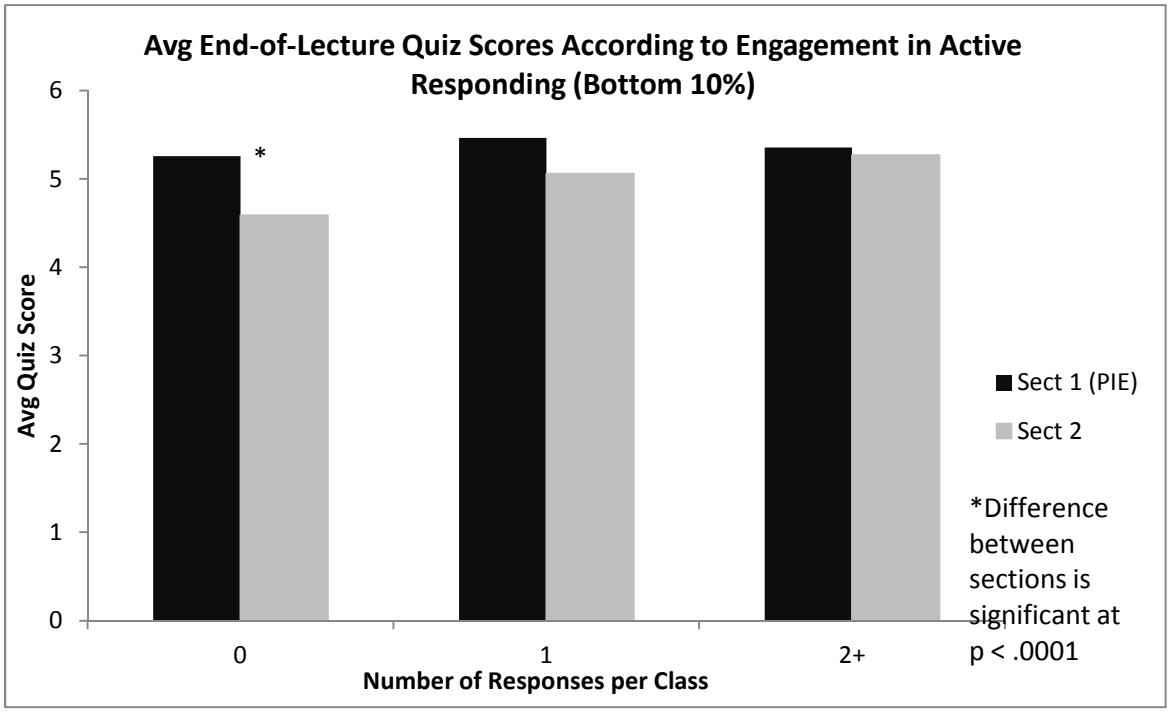


Figure 14. Average end-of-lecture quiz scores according to the degree of student responding during a single class session. Data reflects only the bottom (10%) performers. The black bars show average scores for Section 1 (PIE condition), and the grey bars show data for Section 2 students.

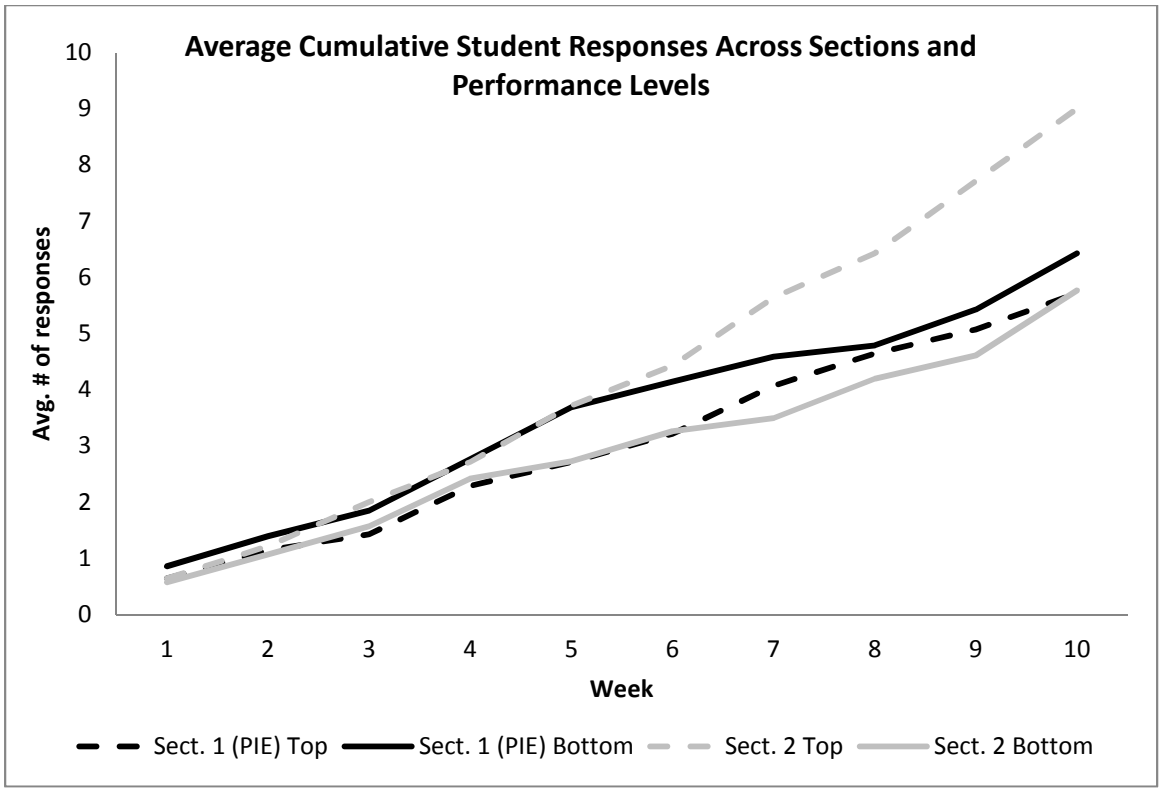


Figure 15. Average cumulative record of active responding across sections for the top and bottom 10% of performing participants.

Appendix A
Questions asked during class lectures

All 12 questions asked during lectures were held constant across both conditions. This provided the same number of opportunities for students to respond and ensured that the question difficulty and content remained the same for all classes for both sections. The difference between a PIE question and a no-PIE question was the student selection method: instructor prompted versus self-selected. Below are the questions used for the Chapter 2 lecture, based on the text *Psychology* (Gray, 2011).

For a PIE class session, eight of the questions were PIE questions where the instructor called on a student and then asked the question. For the remaining four voluntary questions, there was a box around the question and the instructor simply read the question and waited for a response. If there was no response after 10 seconds, the instructor answered the question and moved on with the lecture. For a no-PIE class session, no questions had boxes since all 12 questions were to be voluntary.

Ch. 2 PIE Questions—Boxed questions are voluntary response

1. What is the difference between a fact and a theory?

A fact is an objective statement based on direct observation; and a theory is an idea/model designed to explain existing facts and to predict new facts

2. Is the independent variable the variable that is manipulated in an experiment?

Yes

3. Does one manipulate variables in a correlational study?

No

4. True or false. Descriptive studies systematically investigate relations between specific variables.

False

5. List one advantage and one disadvantage of conducting a field study.

(Will vary) Advantages: realistic environment, less distortion of behavior; Disadvantages: low control, less certainty about causal variables

6. What can be a potential problem with using observers in a study?

The presence of observers could potentially change the participant's behavior

7. What is the function of descriptive statistics?

To summarize/describe data

8. What question are inferential statistics trying to answer?

"What is the likelihood my results are due to chance?"

9. Is error in a study random variability or non-random variability?

Random

10. What's the difference between face validity and criterion validity?

Face= common sense tells you, you're measuring the right thing; Criterion= correlating a measure with a more direct index of the trait that it is intended to measure

11. What is it called when both the observer and the participant are uninformed about the specifics of a study?

A double blind condition/study

12. True or false. Researchers must disclose to their participants that they can leave at any time without penalty.

True

Appendix B
Sample End-of-lecture Quiz Questions

Below is an example of an end-of-lecture quiz that students were required to take every week. The content was taken from Gray's textbook *Psychology* (2011), Chapter 2: Methods of Psychology. The quizzes consisted of six multiple-choice questions drawn directly from the lecture material, and consequently directly from the textbook. Each set of quiz questions contain a mixture of definitional (basic recall) and applied (using terms in a scenario) questions.

The answers to the questions are indicated in bold.

1. A(n) _____ study aims to describe the behavior of an individual or set of individuals without assessing relations between different variables.
A. correlational
B. descriptive
C. self-report
D. experimental

2. _____ is famous for his/her extensive descriptive studies on chimpanzee behaviors.
A. Robert Cialdini
B. Diana Baumrind
C. Jane Goodall
D. William James

3. Before Hannah develops an experimental study, she makes a prediction about new facts that is made from a theory, which is also called a(n):
A. fact.
B. theory.
C. guess.
D. hypothesis.

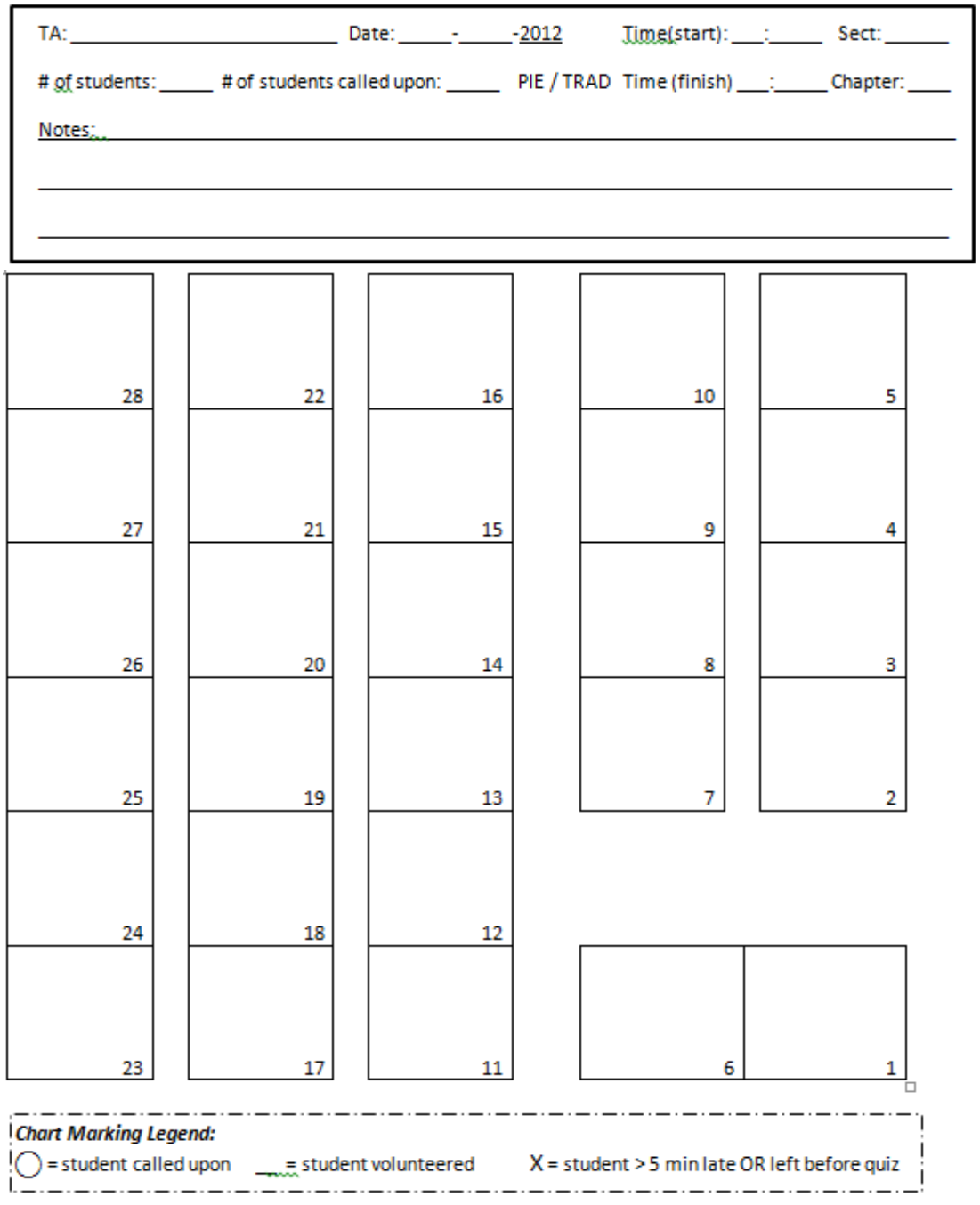
4. Jacob is running an experiment with a between-subjects group design. Before Jacob can determine whether his results arelargely due to chance or if they're statistically significant, he'll need to use:
A. inferential statistics.
B. standard deviation.
C. descriptive statistics.
D. the average group scores.

5. Gordon is about to run a study with two experimental conditions and a control (no IV manipulation). In order to avoid _____ however, he keeps his research assistants _____ to the experiment, so they never know which condition the participant is in.
A. subject-expectancy effects; committed
B. error; blind
C. observer-expectancy effects; blind
D. subject-expectancy effects; blind

6. Tahlia participated in a(n) _____ experiment where neither she nor the study facilitator knows which condition she's in. Later Tahlia is told that she was taking a(n) _____, or an inactive substance.
A. double-blind; placebo
B. unorganized; methyphenidate
C. placebo; blind
D. subject-expectancy; placebo

Appendix C
Sample Seating Chart

Below is a chart that was used for both sections of the course. Undergraduate proctors filled out the names of the students as they checked-in for their class. First and last names of the student were written in the square that represents the desk where the student sat. Instructors marked their seating chart according to how and which students responded.



Appendix D
Class and Instructor Schedule

Below is the class schedule used for the study. Classes were separated by section, with Section1 receiving the PIE condition and Section 2 receiving the no PIE condition. Classes were balanced as best possible throughout the day and week. The black boxes are Section 1 (PIE) class times, and the grey boxes are Section 2 (no PIE) class times. White boxes indicate times when the lab was open for taking quizzes, but no class was being held. White boxes with lines appear for when the lab is closed.

There were four teaching assistants (TAs) serving as instructors. After considering their availability, we attempted to schedule each instructor with roughly the same number of Section 1 classes as Section 2 classes.

	Monday	Tuesday	Wednesday	Thursday	Friday
7:30-7:59	Lab Prep (7:45)-TA 4	Lab Prep (7:45)-TA 4	Lab Prep (7:45)-TA 2	Lab Prep (7:45)-TA 3	---------
8:00-8:29	PIE	PIE		PIE	
8:30-8:59	TA 4	TA 4	TA 2	TA 3	
9:00-9:29	8-9:30	8-9:30	8-9:30	8-9:30	
9:30-9:59					TA 3
10:00-10:29		PIE	PIE		9-10:30
10:30-10:59	TA 1	TA 4	TA 4		
11:00-11:29	10-11:30	10-11:30	10-11:30		PIE
11:30-11:59				TA 1	TA 3
12:00-12:29 p	PIE			11-12:30	11-12:30
12:30-12:59 p	TA 1	TA 2	TA 4		
1:00-1:29 p	12-1:30	12-1:30	12-1:30	PIE	
1:30-1:59 p				TA 2	TA 2
2:00-2:29 p		PIE	PIE	1-2:30	1-2:30
2:30-2:59 p	TA 4	TA 2	TA 3		
3:00-3:29 p	2-3:30	2-3:30	2-3:30		PIE
3:30-3:59 p				TA 2	TA 2
4:00-4:29 p				3-4:30	3-4:30
4:30-4:59 p	TA 4		TA 3		
5:00-5:29 p	4-5:30		4-5:30	PIE	---------
5:30-5:59 p		TA 3		TA 3	---------
6:00-6:29 p	---------	5-6:30	---------	5-6:30	---------
6:30-6:59 p	---------		---------		---------

Appendix E
Instructor Protocol for Responding to Students

Regardless of the experimental condition, all instructors will follow up student responses with positive, constructive feedback. Whether the student's response is instructor prompted, self-selected, or unprompted, all instructors will use the student's name and reply to the student's comment with a note of encouragement (e.g. "That's an interesting observation, Chase. Learning that the 'young male syndrome' is partially due to brain development can help us better understand this phenomenon."). According to the protocol, instructors will avoid sarcasm or other potentially aversive comments and approach every student's question as a legitimate one. After providing feedback for the student, take a few breaths to give the students to think about the answer and ask any questions, if they have them.

The comments below are examples of how to follow up student responses with constructive feedback. Please, follow the pattern for each category as closely as possible (without being robotic) to maintain consistency across instructors. These forms of feedback will be used for *both* PIE and general questions.

For correct answers:

When the student answers a question correctly, quickly respond by saying a clear affirmative statement. Say something like, *yes*, *nice work*, *that is correct*, or *you are right* are appropriate comments. Avoid ambiguous noises like *uh-huh*, *ok*, or *mmm-hmmm*, as this vaguely positive feedback is not constructive.

For partially correct answers:

When a student answers a question partially correct, start with a positive statement along with the student's name, like *you're on the right track* or *you're almost there*, and then continue with the correct answer, demonstrating the difference between the student's answer and the correct one. For example, one could say: "Joey, you're on the right track. Iconic memory does respond to visual stimuli but it is an aspect of the sensory memory store, not working memory."

For incorrect answers:

If the student answers the question incorrectly, start with a clear (not rude, not sarcastic) statement that their response was incorrect, like *no*, or *you might be thinking of something else*, or *that's not what this term describes* and then follow up with praise for effort. For example, one could say: "You might be thinking of something else, Joey, iconic memory responds to visual stimuli and is an aspect of the sensory memory. Thank you for trying!"

Appendix F
Checklist for Observing Instructor Program Integrity

RA Instructions: Sit in the back corner of the discussion area. Avoid interacting directly with the TA's or the students. Mark X's for options on the checklist as they are observed, unless otherwise noted. Start coding after the announcements, and stop coding with the quiz transition slide. Have a reliable clock that counts seconds visible during the discussion.

Coder Name:_____TA Name: _____PIE or Trad. (circle)
Day/Date: _____ Time (start & finish): _____ to _____ Ch: ___

Discussion Protocol Checklist

Observed?	Category	Description
	Timeliness	TA started discussions on time (within 5 min of scheduled time); if not, specify in notes section.
	Seating Chart	TA called on students to confirm student placement.
Questions		TA asked 12 questions ("x" as question is asked).
PIE Questions		TA **first** called on student, **then** asked question from sheet. (n/a if not PIE).
Voluntary Questions		TA asked questions "for everyone," **then** called on student (only 4 questions for PIE condition).
	Extra Questions	Did the TA ask more than 12 questions (meaning, did they ask their own questions)? Mark Y (yes) or N (no).
Question wait time		Did the TA wait 10-s before answering own voluntary question? (only applies to questions when TA answers own "voluntary" question; y (yes), n (no) or blank)
Question Icons		Quest. # appeared right before TA asked the question (mark w/ "X", if # is showing for the whole slide, leave blank)
	Timeliness	TA finished discussion on time (after quiz, discussions should be no longer than 90 min, but no shorter than 80 min)
	General	Did you observe students using electronic devices (e.g. cell phone, laptop, tablet, etc.) *** If marked, elaborate in notes section.
NOTES:		

Appendix G
Course Evaluation: PSY 101.1001/1002 – Fall 2012

Please provide your answers on the attached Scantron sheet.

1. Please indicate your class standing:
a. Fr
b. So
c. Jr
d. Sr

2. Compared to other introductory courses I have taken, the amount of work required in this course was:
a. Much greater
b. About the same
c. Much less

3. Compared to other introductory courses I have taken, my enjoyment of this course was:
a. Much greater
b. About the same
c. Much less

4. Compared to other introductory courses I have taken, the individual attention I received in this course was:
a. Much greater
b. About the same
c. Much less

5. Compared to other introductory courses I have taken, the feeling of accomplishment in this course was:
a. Much greater
b. About the same
c. Much less

6. How helpful were the PSY 101 staff members in responding to your questions and/or concerns?
a. Very helpful
b. Moderately helpful
c. Not helpful

7. Please select your primary instructor among the following alternatives (if none, leave blank).
a. TA 1*
b. TA 4
c. TA 3
e. TA 2

8. How helpful were the group discussion sessions with the instructor in increasing your interest in the subject matter?
a. Very helpful
b. Moderately helpful
c. Not helpful

9. How helpful were the proctors in assisting you to understand the material?
a. Very helpful
b. Moderately helpful
c. Not helpful

10. How helpful were the discussions sessions in preparing you for the quizzes?
a. Very helpful
b. Moderately helpful
c. Not helpful

11. Mark the materials that you were using in studying for the course (mark all that apply):
a. Textbook
b. Chapter Outlines
c. PSY 101 Website
d. Notes from discussion sessions

* Actual names of the instructors were presented in the original document.

12. How helpful was the book in preparing you for the quizzes?
a. Very helpful
b. Moderately helpful
c. Not helpful

13. Would you recommend this course to a friend?
a. Yes
b. No

14. If given a choice, would you choose to learn this way in other courses?
a. Yes
b. No

15. What grade do you expect to receive for the course?
a. A
b. B
c. C
d. D
e. F

16. How do you think the instructor's questioning style affected your performance on discussion & chapter quizzes?
a. Their questioning style increased my quiz grade.
b. Their questioning style decreased my quiz grade.
c. Their questioning style did not affect my quiz grade.

17. How do you think the instructor's questioning style affected your overall engagement with this course?
a. Increased my engagement
b. Decreased my engagement
c. No effect on my engagement

18. What did the instructor's questioning style offer you? (you may choose more than one answer)
a. Better attention during the discussion session
b. Prepared me for possible quiz questions
b. Better learning of the course material
c. Greater interest in the course material
d. Nothing

19. Which approach would you MOST like to experience in future classes?
a. Traditional (only volunteering responses, never get called on by instructor unless you choose to participate)
b. Mandatory Participation (mix of instructor calling on you to participate and voluntary responding)
c. Strictly Lecture-based (instructor asks no question)