University of Nevada, Reno

Examining the Effects of Active Responding on Student 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:

Daniel Reimer

Dr. Ramona Houmanfar/Thesis Advisor

May, 2010

University of Nevada, Reno
Statewide • Worldwide

THE GRADUATE SCHOOL

We recommend that the thesis
prepared under our supervision by

DANIEL REIMER

entitled

Examining The Effects Of Active Responding On Student Performance In An Introductory Psychology Course

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

MASTER OF ARTS

Dr. Ramona Houmanfar, Advisor

Dr. W. Larry Williams, Committee Member

Dr. Victoria Follette, Committee Member

Dr. William Cathey, Graduate School Representative

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

May, 2010

Abstract

The purpose of this study was to examine the impact of the electronic voting devices (i.e., clickers) on immediate and delayed performance of students in an introductory psychology course. The clickers were used to promote students' active responding in small-group lectures/discussion sessions in an alternating reversal (exposure vs. lack of exposure to active responding promoted by clickers) design across two sections of students in this course. The performances of our experimental groups were also compared to the performance of a group of students from a previous semester that experienced an alternated exposure to clickers. The dependent variables included students' performance associated with discussion session quizzes, chapter quizzes, mid-term and final exams. In addition, we compared grade distribution across target semesters. The increase in student performance seen across semesters that applied the EVS, combined with the data found within the target semester provides a context for future empirical exploration of EVS as a potential source of active responding in classroom settings.

Table of Contents

ABSTRACT	i
TABLE OF CONTENTS	ii
LIST OF TABLES	iii
LIST OF FIGURES	v
INTRODUCTION	1
METHOD	18
Participants and Settings	18
Procedure	20
Dependent Variables and Measurement	23
Independent Variable	25
Experimental Design	27
Data Analysis	29
RESULTS	30
DISCUSSION	38
CONCLUSIONS	48
REFERENCES	50
TABLES	54
LIST OF FIGURES	86
FIGURES	87
APPENDICES	106

List of Tables

Table 1: ANOVA: Fall 2009 Section 1 and 2 Discussion Quiz Scores by Week

Table 2: ANOVA: Fall 2009 Section 1 and 2 First Quiz Scores by Week

Table 3: ANOVA: Fall 2009 Section 1 and 2 Final Quiz Scores by Week

Table 4: ANOVA: Fall 2009 Section 1 and 2 Number of Quizzes Attempted by Week

Table 5: ANOVA: Fall 2009 Section 1 and 2 High Achievement Group Discussion Quiz Scores by Week

Table 6: ANOVA: Fall 2009 Section 1 and 2 Low Achievement Group Discussion Quiz Scores by Week

Table 7: ANOVA: Fall 2009 Section 1 and 2 High Achievement Group First Quiz Scores by Week

Table 8: ANOVA: Fall 2009 Section 1 and 2 Low Achievement Group Discussion Quiz Scores by Week

Table 9: ANOVA: Fall 2009 Section 1 and 2 High Achievement Group Final Quiz Scores by Week

Table 10: ANOVA: Fall 2009 Section 1 and 2 Low Achievement Group Final Quiz Scores by Week

Table 11: ANOVA: Fall 2009 Section 1 and 2 High Achievement Group Number of Quizzes Attempted by Week

Table 12: ANOVA: Fall 2009 Section 1 and 2 Low Achievement Group Number of Quizzes Attempted by Week

Table 13: ANOVA: Comparison of Final Quiz Scores Between Spring 2009 with Fall 2009 by Week

Table 14: ANOVA: Comparison of Final Quiz Scores Between Fall 2008 with Spring 2009 by Week

Table 15: ANOVA: Comparison of Final Quiz Scores Between Fall 2008 with Fall 2009 by Week

Table 16: Social Validity Results Pertaining to Perceived Effectiveness of AR techniques in a Classroom setting

List of Figures

Figure 1: Average discussion quiz score for Section1 and Section 2 across ten weeks.

Figure 2: Average first quiz score after discussion for Section 1 and Section 2 across ten weeks.

Figure 3: Average final quiz score for Section1 and Section 2 across ten weeks.

Figure 4: Average number of quiz opportunities utilized by Section 1 and Section 2 across ten weeks.

Figure 5: Average final score across both sections, grouped according to number of quiz opportunities utilized across ten weeks.

Figure 6: Average discussion quiz score for Section 1 and Section 2, separated by High and Low Achievement Groups, across ten weeks.

Figure 7: Average first quiz score after discussion for Section 1 and Section 2, separated by High and Low Achievement Groups, across ten weeks.

Figure 8: Average final score across both sections, separated by High and Low Achievement Groups, across ten weeks.

Figure 9: Average number of quiz opportunities utilized by Section 1 and Section 2, separated by High and Low Achievement Groups, across ten weeks.

Figure 10: Average discussion quiz score for Instructor 1 for Section 1 and Section 2 across ten weeks.

Figure 11: Average discussion quiz score for Instructor 2 for Section 1 and Section 2 across ten weeks.

Figure 12: Average first quiz score after discussion for Instructor 1 for Section 1 and Section 2 across ten weeks.

Figure 13: Average first quiz score after discussion for Instructor 2 for Section 1 and Section 2 across ten weeks.

Figure 14: Average final quiz score after discussion score for Instructor 1 for Section 1 and Section 2 across ten weeks.

Figure 15: Average final quiz score after discussion score for Instructor 2 for Section 1 and Section 2 across ten weeks.

Figure 16: Average final quiz score for Fall 2008, Spring 2009 and Fall 2009 semesters across ten weeks.

Figure 17: Comparison between the average Final Weekly Quiz Score between Fall 2008, Spring 2009 and Section 1 and Section 2 of Fall 2009.

Figure 18: Average number of retakes students in each section completed during the Spring 2009 semester.

Examining the Effects of Active Responding on Student Performance in
an Introductory Psychology Course

B.F. Skinner was one of the first behavioral scientists to suggest that education could be analyzed using behavioral principles (Skinner, 1954). He strongly emphasized the need to create a more effective educational system that increased the number of students who excelled. His solution to increasing the effectiveness of an education system promoted data driven methods, used positive reinforcement, and encourage empirically validated and scientifically sound practices (Skinner, 1958). His emphasis on these three important aspects of education led to many teaching strategies, such as the utilization of teaching machines (Benjamin, 1988), Programmed Instruction (Vargas and Vargas, 1991), Precision Teaching (White, 1986) and Personalized System of Instruction (Keller, 1968).

The teaching machine was first developed by Sidney Pressey (1926). However, the teaching machines developed by Pressey were not based on behavioral principles. Rather, the machines developed by Pressey and his colleagues required the student to make numerous errors to facilitate learning, while Skinner's machines attempted to diminish the number of errors (Benjamin, 1988).

Programmed instruction was facilitated by the advent of the teaching machine. Programmed instruction is a technique that is intended to slowly shape a student's verbal repertoire (Vargas & Vargas, 1991). The shaping takes place using frames of information. The frames break down concepts into manageable pieces, allowing the teacher to slowly introduce the student to basic ideas and then reinforce the formation of those basic ideas into overarching concepts. In order for programmed instruction to be effective, it needs to include important, behaviorally based aspects. These aspects include defining a clear learning objective and determining small steps to reach the objective, making sure the small steps are in a logical and sequential order. In addition, the student must be required to respond actively, have access to immediate feedback, have sufficient practice time and be slowly faded from discriminative stimuli (Vargas & Vargas, 1991).

Precision teaching is an instructional evaluation tool to measure student's educational progression (White, 1986). It focuses on five major tenets rooted in the behavioral sciences: the idea that the learner knows best, the exclusive focus on observable behavior, the use of standardized charts to aid in data interpretation, the use of frequency counts as the measure of analysis, and a systematic analysis of the environmental influences of behavior (Lindsley, 1971).

Personalized system of instruction was first developed by Keller (1968). His system, also referred to as PSI, attempted to alter the traditional teaching method in a way that allowed all students to succeed at their own pace. He incorporated behavioral techniques in order to exemplify the learning experience. These techniques include requiring each student to achieve a mastery of the material while allowing the student to work at their own pace, treating the traditional lecture format as a supplement to personal exploration, encourage written communication between students and teachers and using proctors or peer tutors to facilitate test taking, provide immediate feedback on performance and tutoring if necessary (Keller, 1968).

Active responding (AR) is one of the defining characteristics of behaviorally based teaching techniques. All four of the techniques discussed require some form of active student responding in order to be effective. Teaching machines inherently require an active response, as it is the student that is required to behave in order for the machine to work. Programmed instruction specifically requires active responding to frames of information in order to continue on in the lesson. Precision teaching measures observable interaction with the course material, which is an active response. The Personalized System of Instruction utilizes proctors to provide feedback and discuss questions that the student has. This one-on-one interaction with an individual more experienced in the material promotes active responding in terms of discussion.

John Dewey was the first to publish support for active responding (1916). He did not utilize behavioral language in his discussion of effective teaching methods, or data to support his claims in his book *Democracy and Education,* but his theories on interactive learning and engaging students in the material are supported by behavioral research. He discusses the importance of experience in learning, and claims that learning in "not primarily cognitive." His stress on students as part of the learning environment, rather than what he terms "theoretical spectators" implies the importance of active responding in learning.

The concept of active responding was further developed by Skinner (1954). Other scientists have explored AR under a variety of different names, from a variety of different fields (Cue, 1998; Draper & Brown, 2004; Dugresne, R. J., Gerace, W. J., Leonard, W. J., Mestre, J. P. & Wenk, L.,1996; Hake, 1997; Kennedy & Cutts, 2005; Wit, 2003), including behavior analysis (Cavanaugh, R. A., Heward, W. L. & Donelson, F., 1996; Gardner, R., III, Heward, W., & Grossi, T. A., 1994; Malanga, P. R. & Sweeney, W. J., 2008; Miller, M. L. & Mallot, R. W., 1997). However, active responding has not been objectively defined in the literature, perhaps because there are many numerous and diverse ways to respond actively.

Active responding in the context of a classroom adds significant connotations to active responding that sufficiently narrows down what is and is not considered active responding. Fernald and Jordan (1991) state that active responding occurs when the student is required to construct an observable response to the material, instead of only recognizing a correct response as seen in traditional lecture formats. This concept appears slightly different than the notion of overt responding as defined by Miller and Malott (1997) and Haggas and Hantula (2002). According to these two articles, an overt response is simply a response that is observable. It is only assumed that the overt response will be in reference to the target material. Active responding is not just an overt response, such as talking, moving or pressing computer buttons. It is a response that is targeted toward the material being taught, and implies engagement with the material. For example: asking a question about the material, not talking with a fellow student; raising a hand to ask a question about the material, not stretching; and clicking a computer mouse to respond to a question for an on-line quiz, not playing a computer game while listening to a lecture.

Students who interact with the material actively, rather than passively attend, will maximize their learning (Austin, 2000; Miller & Malott, 1997). Results have been mixed (Malanga & Sweeney, 2008; Shabani & Carr, 2004) but most studies have shown that active responding increases student's success (Fernald & Jordan, 1991; Hake, 1997; Miller & Malott, 1997; Tudor, 1995; Tudor & Bostow, 1991).

The traditional method to increase active responding is to promote more class discussion. This is ineffective because very few students participate in class discussions, and discussions take up time that the instructor needs to communicate important information (Austin, 2000). Typically only a very small percentage of students will interact with the instructor in this format. Many students find speaking up in class aversive, decreasing the probability that they will participate in discussions unless forced. Consequently, the discussions are brief and infrequent. The majority of students do not benefit from active responding in this way because they are not actively engaged in the instructional material, but rather passively attending to the discussion in the same way as they would to a lecture format. To combat this, some instructors break the class into groups to encourage more students to participate in discussion. Small group discussions not only require preparation time on the part of the instructor, but they also consume a sufficient amount of class time and will substantially diminish the class time that is utilized to impart important information to the students (Austin, 2000). These detrimental side-effects become even more pronounced as class size increases, making discussion an impractical solution for large enrollment classes.

Behavioral scientists have typically studied three alternative methods to class discussion, specifically choral responding, guided notes and response cards (Austin, 2000; Malanga & Sweeney, 2008). For choral responding, the instructor poses a question to the class. The instructor then requires that all students in a class to orally respond to the question simultaneously (Heward, Courson & Narayan, 1989). This technique attempts to ensure that all students are actively responding to the material. However, a number of flaws make this technique untenable. First, it is impossible to distinguish between individual student's answers. The instructor is expected to gauge the student's mass response to determine if every student adequately understood the material that was presented. Not only is this a subjective measurement of student performance because the instructor has to conduct a "best guess" type analysis, it makes data collection impossible. The instructor can, at best, determine one or two student's responses, which is far from ideal in terms of data collection. Also, additional effort is needed to record the perceived amount correct or the estimated percentage of the class that answered correctly. The teacher would have to generate a data collection system and input the subjective data that was collected. Secondly, the instructor has limited ability to ensure that all students are in fact responding. Some students may simply refuse to respond in any overt way, or they might respond by simply moving their mouth. Their response can in no way be contingent on consequences delivered by the teacher, because it is impossible to differentiate answers when they are given as a group response. When grades are not contingent on responding, the number of students that participate will decrease. In addition to these disadvantages, choral responding has never been studied in a collegiate setting. The utilization of choral response in the university classroom has probably never been studied because, in addition to all the disadvantages listed above, the students would think the process was juvenile and refuse to participate.

Guided notes require each student to answer independently. The notes are created by the instructor beforehand and are modified so that there is important information missing (Austin, Lee & Carr, 2004). In order for students to get the complete set of notes, they must attend to the lecture and determine where the missing information is presented. Once they deduce what is missing, they fill in the appropriate information in the areas that were removed by the instructor. While this technique has shown a substantial increase in student's note-taking abilities and increased quiz and test scores (Neef, McCord & Ferreri, 2006; Austin, Lee & Carr, 2004), there are substantial disadvantages that must be taken into account. First, it is difficult to collect data on guided notes. The notes are intended for the student's use, so the instructor cannot collect them to verify that students are filling in the missing information correctly. The only verification data that the instructor can collect on the utilization of active responding in terms of note-taking is the movement of the student's pens or pencils on the paper. The teacher cannot possibly verify that the students are accurately taking notes without watching each individual student fill out the guided notes. It becomes impossible to verify that students are utilizing the notes correctly as the class size increases. It is possible that the students are not filling in the blanks correctly, or they are merely doodling on the note sheet and not actually taking notes. Data cannot be collected on correct or incorrect note-taking without collecting the notes that are taken, which defeats the purpose of requiring students to take notes.

Second, guided notes only allow active responding in the form of note-taking. As the instructor lectures, the students must identify what information is missing from the notes. This sets up contingencies that compete with other forms of active responding. If a student attempts to respond to the material verbally, in the form of a question or comment, he or she might not attend to the information that is missing from the guided notes. It is also possible that students will not verbally interact with the instructor as much as usual while they are writing down the notes from the lecture. This can be attributed to the length of time it takes to write down the correct information. The focus becomes filling in the blanks on the page of notes, and not overtly interacting with the material in other ways.

Lastly, guided notes can easily be filled out while not attending to the material being presented. A student can copy the needed information of another student in the class, or ask to see another student's notes after the class is over. This emphasizes the fact that it is note the actual note-taking that increases participation and scores for students. Rather, guided notes are another way to increase the probability that students will respond to the material presented by the instructor and structure the response so that all students answer uniformly.

Response cards are an attempt to control for the disadvantages of the other two types of active responding. Response cards attempt to increase student active responding by assigning each individual student cards. These cards can come in multiple forms. Some utilize individual whiteboards or dry erase boards that allow each student to respond to multiple-choice, true/false, or short answer questions (Cavanaugh, Heward & Donelson, 1996; Christle & Schuster, 2003; Gardner, Heward & Grossi, 1994; Narayan, Heward, Gardner, Courson, & Omness, 1990). The instructor poses a question to the class and gives the students time to write out what they think the correct answer on their white board. The instructor then asks the class to reveal their answers at the same time, and all students display their written answer to the instructor.

Response cards can also be pre-printed. Pre-printed response cards have the possible answers printed on both sides, so that the student can show either side to the instructor as a response. These cards are most often used to clarify important principles (Malanga & Sweeney, 2008). When utilizing pre-printed response cards, an instructor passes out cards pertaining to an important concept, such as dependent and independent variables with “IV” on one side and “DV” on the other. The instructor would then describe an experiment in detail and then specify one variable in the experiment. The instructor would ask the students to identify whether the variable was dependent or independent by holding the response card with the correct answer was facing the instructor. Another version of pre-printed cards allows for more flexible questions. Instead of having concepts pre-printed, letters are printed on the cards so that the instructor can ask multiple-choice questions with two possible answers, such as true and false, yes and no, A and B (Kellum, Carr, & Dozier, 2001; Shabani & Carr, 2004), or four possible answers such as A, B, C, and D (Marmolejo, Wilder, & Bradley, 2004) to which the students hold up the appropriate card.

Lastly, response cards can be color coded. An instructor can raise a question and request that students respond to what they believe the correct answer is with a corresponding color. This is very similar to a multiple choice or true/false question, except that instead of four letters or a “T” or an “F”, the student has the option to raise a yellow, blue, purple or orange card for multiple choice or a red or green card for true and false. This is supposed to provide the instructor with an easier way to determine student’s answers because it is easier to distinguish a colored card than a card with writing on it. In a majority of cases studied (Gardner, Heward & Grossi, 1994; Kellum, Carr, & Dozier, 2001; Malanga & Sweeney, 2008; Marmolejo, Wilder, & Bradley, 2004) a combination of the three variations were used.

Though response cards have been used to effectively increase test scores (Cavanaugh et al., 1996; Malanga & Sweeney, 2008; Marmolejo et al., 2004, Shabani &

Carr, 2004), as well as increase other forms of class participation (Gardner et. al., 1994; Kellum et al., 2001), there are still a number of disadvantages that are inherent in their use. First, using response cards does nothing to diminish the aversive effects of speaking up in class. In fact, it may be even more aversive if the student is trying to avoid emitting an incorrect response. When the instructor asks each student to hold up their response card, it becomes visible to the entire class not just the instructor. If a majority answered correctly, the minority's behavior might be punished for getting the answer incorrect when all their peers could see. The students who got the question incorrect might refuse to answer the question or begin to cheat to ensure they do not get another answer wrong.

Second, cheating behavior is easy to engage in and difficult to control. If a student does not know the answer or is unsure and afraid of getting an answer wrong, they only have to wait until the students in front of them raise their cards. In the few seconds it takes for the majority to raise their cards, a cheater can select the card that the majority is holding and raise it themselves. This eliminates the social punishment of incorrect answers, but it does not accurately reflect the students understanding of the topic. It is difficult for the instructor to control for this, even if they request that all students raise their cards at the same time, because the speed at which students raise their cards is variable.

A third disadvantage to response cards is that the probability of guessing the correct answer decreases proportionally to the number of options the student has to answer. If the response cards only give the option of true and false, the student has a fifty percent chance of guessing the correct answer. The more available options there are to answer, the less likely it is for the student to guess. However, this does not control for cheating, as was discussed earlier. This confound is eliminated in the white board response card method, because it allows the instructor to ask questions that require the student to fill in the blank of complete a sentence. However, the instructor must then evaluate each response and determine if it is close to the correct answer, as student's responses may vary.

Fourth, trying to increase active responding through response cards is difficult because of the inability of the instructor to create contingencies for answering, and answering correctly. The teacher cannot ensure that all students are answering without paying very close attention to each student systematically. Most likely the teacher will only make a cursory glance and move on. Contingencies for correct answers are just as difficult to administer. Usually, if a majority of the class holds up the right card, the teacher delivers social praise and moves on. This social praise is immediate, but may have very little reinforcing effect on performance. If additional reinforcers were implemented, it would be very easy for students who are unsure about the answer to cheat in order to access the reinforcer, as mentioned above.

Similarly, the fifth disadvantage of response cards is the collection of the student's responses. When teachers implement response cards they usually do not take data on student responses. It is used as a tool to approximately gauge the general classes understanding, not each individual answer. The research published on response cards have not analyzed the effect that the cards have on individual students, with the exception of Gardner et al. in 1994. Rather, the designs look at group performance on quizzes, tests and exams. This is most likely because of the difficulty of collecting data on each individual member in a class. Gardner et al. did a study with five individual students, so data collection was easier than is usual in a classroom setting. Their procedures cannot be generalized to a full class because of the labor intensive data collection process. Collecting individual data is required to determine if active responding is beneficial to all students. If data is not collected for individuals it is impossible to make reinforcement contingent on individual responses.

All the disadvantages of response cards are exacerbated as the number of students in the class increase. Data collection, contingent reinforcement, and allowing the instructor to observe the progression of a class using response cards becomes more difficult to successfully implement as the class size increases, while the negative effects, such as cheating, refusing to participate and forcing uncomfortable students to publicly respond become more prevalent, making response cards an effective solution for small to medium sized classes.

Though studies have shown that the abovementioned techniques improve student success, all come with disadvantages to both the student and the instructor. The main disadvantages are the inability to prevent cheating, guessing or refusing to respond, as well as difficulties in effectively assessing each student's individual response and the accurate collection of response data. These disadvantages are eliminated in the technology referred to as an Electronic Voting System (EVS).

An EVS in the classroom consists of individual voting devices colloquially called "clickers." The instructor can present a question verbally or visually and have each student respond using their clicker. The devices allow each student to submit their answer electronically using a radio frequency that is collected by a receiver attached to the instructor's computer. The result is a recorded response count. The results of the response count can be posted publically, usually as a histogram that represents how many students chose which answer, kept for the instructor's personal use or both. The software typically does not show each individual's answer, but rather the results in terms of the class as a whole. An EVS allows the instructor to access how many students understand a concept, but in a much more accurate and informational way than other, traditional, techniques. This technology has many of the same benefits of the typical behavioral methods of increasing active responding but also controls many of the disadvantages.

The technology available for clickers is advanced enough to allow instructors to ask true/false, multiple choice, fill in the blank or short answer questions. This ability allows much more flexibility in integrating active responding into a typical lecture, because the instructor can require whatever technology best suits the information being presented.

The class response is depicted in a way that the instructor can determine if the majority of the students understand the material. The responses that are collected by the receiver are instantly tabulated and displayed so that the class response is depicted according to each student's response. For example, four students though the answer was option A, zero thought it was option B, 23 thought it was option C and zero thought it was option D. If the correct answer was option C, the instructor could move on with reasonable confidence that the class understood the concepts that they were exposed to. This is even more accurate than response cards, because the instructor can access the exact number that answered correctly, as well as those that answered incorrectly.

The display of the students answers are done in a way that greatly minimizes the aversive effects of response cards or discussion. In response cards, everyone in the room can access each student's answer, making it just as aversive as speaking up in class. The software that the clickers deliver the response to does not reveal individual answers. Even if only one student answers a question incorrectly it is impossible to determine which student it was, as the topography of responding is the same regardless of answer.

However, the instructor has the ability to access each individual answer to each question. This ability allows the instructor to make reinforcement contingent on response. The instructor can give participation points to any student answer, or can give points for correct answers. This flexibility not only encourages all students to respond, it can also provides points contingent on correct answers to encourage students to attend to information being presented, which is impossible to implement with response cards. If responses are tied to a grade, the use of the clicker to respond might reinforce studying behavior so as to answer the questions correctly.

Clickers also dramatically diminish the possibility of cheating. The correct answer is not displayed until all individual answers have been received, and each individual answer is sent in an almost imperceptible way so a student cannot merely wait until another student visibly responds or wait for the correct answer before responding. The student is forced to answer without assistance, more accurately reflecting their own learning on a concept.

The instructor has access to historical data, as well as the data from the current class. This is beneficial for many reasons. Analyses can be conducted to see if some students are consistently struggling with the material, so the instructor can precisely identify who needs help and what they need help on. They can also pinpoint who is excelling. This is impossible to do with response cards because no historical data can be taken accurately.

An instructor can continually analyze their effectiveness and develop techniques that accurately and effectively present material so that students understand. The feedback that the instructor receives after each question is almost instantaneous, so techniques can be changed rapidly during actual instruction. The instructor can also systematically vary instruction and use the historical data that is collected to determine what improves student scores most effectively.

The size of the class has little effect on the use of clicker software. While the answers using the response cards get increasingly difficult to distinguish as the class's size increases, the technology of the voting system still instantly collects and analyses the data so that the instructor can see the exact amount of answers. The instructor will have very little difficulty in deciding whether or not a majority of the students understand the material.

Despite the assumed superiority over the three traditional active responding techniques, EVS have very little empirical evidence to support their utilization. Numerous studies have been conducted to evaluate the effect EVSs have on students (Beekes, 2006; Cue, 1998; Draper & Brown, 2004; Draper, Cargill & Cutts, 2002; Haggas & Hantula, 2002; Kennedy & Cuts, 2005; Wit, 2003; Wood, 2004). Seven of these articles only collected data on student's perceived usefulness or satisfaction of the EVS technology. For example, student was asked to estimate if the clickers that were utilized in the class were beneficial to them, if they think it improved their grade, if they were satisfied with the clicker system, and if they would recommend that the teacher used clickers again. No data were collected in relation to any of the student's academic performance. These results are little better than anecdotal evidence for the successful use of clickers in a classroom setting.

Kennedy & Cutts (2005) was the only article that analyzed the use of clickers in relation to academic performance. Clicker use was alternated with a traditional lecture format for approximately half of the class periods. The class periods that were exposed to clicker technology were scheduled for their convenience, not for scientific integrity. The students were asked between two to six questions during class periods that utilized clickers. The study examined the frequency of usage of the clicker, the number of clicker questions answered correctly and scores on important exams. A cluster analysis was used to relate frequency of clicker use and clicker questions answered correctly. The analysis revealed four types of usage; moderate users that answered a moderate number of questions correct, high users that usually got responses incorrect, low users that usually got responses incorrect and frequent users that got a moderate number of questions correct. The experimenters then compared exam scores across these four groups. The results showed a significant difference between the two groups that used clickers at a higher rate, regardless of how many questions they got correct. This suggests that just using the clickers during class, regardless of if the answers are correct, will improve performance on exams.

However, the results from this analysis cannot experimentally support the claim that clickers positively affect student's academic performance. Experimental control was not demonstrated in this study. The class periods that allowed students to access clickers were determined by their convenience. The majority were conducted during the middle of the semester, so that they did not interfere with orientation at the beginning and testing at the end. There was no control group used to measure change against, and no experimental manipulation to demonstrate cause or effect. The results are correlational at best. Numerous other factors can account for the differences seen in exam scores seen between each of the four clusters. Those that used the clickers frequently may also have been accessing the material outside of class at a higher rate than those who did not access the clickers as frequently. The extra work they put in would result in higher test scores.

Since there have been very few studies assessing the effectiveness of EVSs, and even fewer demonstrating the effects these systems have on actual student performance in terms course work scores, the results of the present study could change the way that EVSs and AR are used in the classroom. It would be extremely beneficial to the educational field if this teaching technique were to be scientifically validated, in addition to the satisfaction surveys and anecdotal evidence that are the only studies that currently support its use.

Recently a program evaluation was conducted to examine the impact of the EVS on immediate and delayed performance of students in an introductory psychology course. The clickers were used to promote students' active responding in small-group lectures/discussion sessions during the target semester. The performance of this group was compared to the performance of a group of students from a previous semester that experienced the traditional style of instruction. In addition, an analysis was conducted of students' performance in accordance to their participation in discussion sessions (exposure vs. lack of exposure to active responding promoted by clickers) during the target semester. The dependent variables included students' performance associated with discussion session quizzes, chapter quizzes, mid-term and final exams. Grade distribution across target semesters was also compared. In summary, little differentiation was observed across alternating conditions during the target semester. Our results suggested the potential sequence effect. In addition, rapid alternation of conditions may have contributed to the lack of differentiation. Moreover, students were only able to access the clicker software once before they were exposed to a discussion session without access to a clicker. Accordingly, the more stable (longer with higher frequency of exposure to independent variable) counterbalanced reversal design utilized in the proposed experiment is intended to allow a more frequent access to the independent variable.

The purpose of the current study was to analyze the effects that EVSs have on student performance. The study looked at student performance in terms of answering multiple choice questions correctly when exposed to a class session that integrates EVS technology into the material, and compare it to student performance when exposed to only traditional teaching methods. The practical applications of the results could result in an economic shift in classroom structure, as an EVS will allow teachers to instruct large groups of students more effectively, as well as additional technology that can be developed to increase the effectiveness of an EVS in the classroom.

Method

Participants and Settings

Psychology 101 is a large enrollment course aimed at teaching incoming University of Nevada, Reno students the basic principles of the scientific field of psychology. Typical enrollment for this course is approximately 600 students. Psychology 101 is a lower division requirement for all majors, so the class regularly consists of freshman and sophomore university students and small number of junior and senior students. During the orientation period of the class, each student must fill out a demographic survey detailing their age, gender, year in school (i.e. freshman, sophomore), and major.

The students who were enrolled in two sections of Psychology 101 with Dr. Ramona Houmanfar were used as participants in this study. All student participants were given the option to not have their data collected without fear of retribution, but they were informed that all protocols must be completed, regardless of participation in the study since the protocols were integrated into the class curriculum. If a student declined to participate, their data was not included in the analysis. In Section 1 of the course, 199 students agreed to participate. A total of 177 students agreed to participate from Section 2. Five graduate students who worked as teaching assistants were also involved in this study, as primary instructors. Each graduate student was enrolled in the University of Nevada, Reno's Psychology Program at the time of study, and were teaching assistants as part of their practicum or stipend placements.

Participants were only included in the data analysis if they met certain criteria. First, those who agreed to participate had to attend at least eight out of ten discussion sessions during the semester. Only one of these missed discussions could be for a discussion that received the AR treatment. This was to insure that participants had adequate exposure to the experimental variables. Student participants who were late to a discussion were considered as absent for the purposes of this experiment. Student participants were counted as late if they arrived to discussion after the first AR or T slide was presented. In addition, student participants who decided to move at a faster pace were also not included, since they were not exposed to the same variables as participants attending regular discussion sessions. In total, 29 participants in Section 1 and 26 participants in Section 2 did not meet the inclusion criteria, so their data set was removed from analysis.

All available data were still collected for student participants who did not meet the inclusion criteria (i.e. those who decided to move ahead at a faster pace, those who were absent for more than two sessions, and those who forget their clickers) because it was required to calculate their grade in the class, but their data was not included in the final analysis.

Each discussion was held in the Psychology 101 Learning Laboratory, which is located in Cain Hall, room 253 on the campus of the University of Nevada, Reno. The room consists of an entry-way, a discussion room resembling a classroom for approximately 25 students, a quiz room equipped with 15 computers available for students to take weekly quizzes, mid-terms and finals, and an office used by the teaching assistants and proctors. Each discussion session took place in the discussion room. All discussion quizzes took place in the discussion session room. All other quizzes, mid-terms and finals were conducted in the quiz room.

Procedure

Each week participants were required to attend a discussion session on one chapter. Attendance for weekly discussions was compulsory, unless the student participant is moving ahead. Discussion was held five times a day, except Friday which offered four discussion times. The discussions lasted approximately 90 minutes long, and focused on one chapter from Peter Gray's book on introductory psychology (2007). The discussion was led by one of five Teaching Assistants participants, also called instructors, who taught from a standardized Powerpoint presentation. Each instructor followed a teaching protocol to attempt to control for style differences between instructors (Appendix A and B). Directly following the discussion session student participants took a discussion quiz. Each discussion quiz was taken with the same EVS that is being used in the present study. Six questions were presented to student participants one at a time, on a Powerpoint slide. The student participants were given a maximum of one minute to answer the question. The answer was then displayed after all student participants have answered the question. Once the student participants saw the correct answer, the slide was changed and another question was asked.

Discussion sessions were held separately for each section, so that only participants from Section 1 were able to attend discussion sessions for their section. Participants from one section were only able to schedule a discussion time that was available for their section. For example, a participant from Section 1 may have scheduled a discussion time on Monday at 3:30 PM. If a participant from Section 2 tried to schedule a discussion at the same time, the appointment would be denied. This was implemented in order to isolate the variables being applied so that some participants were being asked not to respond actively in a class with others who were asked to respond.

Student participants were also required to take a weekly quiz on the material from discussion. The weekly quiz did not have to occur in a specific relation to the discussion, but its completion was required sometime during the week that the discussion on that material was offered. In other words, if a weekly chapter focused on behavioral psychology, the student participant was required to attend a discussion session and complete a discussion quiz on behavioral psychology as well as a weekly quiz on behavioral psychology. Weekly quizzes held in the quiz room consisted of 15 multiple choice questions, randomly selected from a test bank of about 150 questions. Student participants were allowed two opportunities to take the quiz Monday through Thursday, and once on Friday. Sections taking the weekly quizzes, mid-terms and finals were not isolated, so that either of the sections were allowed to take weekly quizzes during the same times. The mid-term and final tests were administered during the weeks designated by the University as testing weeks while the pre-test was administered during orientation week, or the first week of the semester.

Student participants who progressed through the material faster than the minimum pace did not attend discussion sessions. Instead they participated in a one-on-one discussion session with a proctor or tutor. The one-on-one sessions were conducted similarly to standard discussion sessions, but were lead by an undergraduate proctor that demonstrated proficient knowledge and ability to present the material, and received college credit. Student participants moving ahead were not included in the general data analysis, because the general analysis is meant to investigate the affect clickers have in normal classroom settings. Since the participants who moved ahead at a faster pace received different instructional methods, namely a one-on-one session with a qualified proctor, they were not considered to be representative of normal classroom interaction. However, their data was still collected and analyzed if they gave permission, to determine if those receiving advanced instruction were affected by clicker technology (See section on data analysis).

In order to ensure that student participants had full access to the variables in this study, only student participants who were in the discussion room when the discussion began were counted as present for the purposes of this experiment. Any participants that arrived after the first active responding slide for discussion sessions that are exposed to AR, or after the first question prompt slide for traditional discussion sessions, were counted as absent for the purposes of this study. This was intended to eliminate data of student participants who did not access the experimental variables to the same extent as other student participants.

Dependent Variables and Measurement System

Dependent variables included scores on the discussion quizzes, the weekly chapter quizzes, the mid-term and the final. The actual response class was the behavior of selecting one of the multiple choice answers (A, B, C or D) during the discussion quiz if they were exposed to the clicker, the discussion quiz, the weekly chapter quiz, the mid-term and the final. A key was already in place to determine which responses were labeled as correct, and which as incorrect. The system was set up so that each response for each participant was counted as correct if they picked the correct answer, to ensure consistency in data collection. The electronic system that was already in place as part of the design of the course was capable of collecting the AR discussion session quizzes, discussion quizzes, weekly quiz scores as well as the scores on the pre-test, the mid-term and the final. All student participants' scores from these quizzes and test were stored on a computerized classroom grade-book tool, called WebCT. Discussion quizzes were conducted with the EVS for all sections, regardless of if they came in contact with AR that session or not. The scores for the discussion quizzes were collected using the same technology that collects the scores for the AR discussion questions. The technology allows for exportation to WebCT by formatting the quiz information into an Excel file.

There were processes already in place in order to minimize cheating. Each student participant was required to present photographic identification before they were allowed to access the weekly quiz material, discussion sessions, the mid-term material and the final material. Undergraduate proctors were always present in the quiz room to prevent cheating. Student participants were required to leave all personal belongings, including electronics, by the door before they could gain access to the test material. Proctors also periodically checked the student participant's computer screens to ensure that they were only accessing the quizzing material and not acquiring answers from the internet. All instances of cheating were reported to instructors and to Dr. Houmanfar. Any student participant who was reported as having cheated had their data removed from analysis for the week they were caught cheating. If there were additional instances of cheating reported, the student participant's entire data set was removed from analysis.

Cheating on AR questions was minimized by requiring student participants to produce photographic identification in order to enter the discussion room. Though this did not prevent student participants from entering the discussion room with an alternate individual EVS device, it was difficult for student participants to get credit for answers given for an absent individual. Each clicker has a unique radio frequency ID, as well as a programmable ID used to identify individuals. The program utilized was able to detect a clicker that has sent an answer twice, even if the answer is sent in on a different day.

In addition to producing photo-identification and technological systems to reduce cheating, each discussion group were told to wait outside the Learning Lab while another discussion group was in session. A break of fifteen minutes was scheduled to give instructors adequate time to transition between sessions, but occasionally a student arrived more than fifteen minutes early for discussion. Having all students wait outside until the prior discussion was over was intended to prevent early arrivals from auditory or visual exposure to the discussion quiz questions.

Collection of correct student participant responses to the AR questions was facilitated by a similar system that collects each individual student participant's answer and exports it to an Excel file. That file was uploaded to WebCT and integrated into the student grade book. All procedures were implemented for both students who agreed to be participants and for those that did not. This is because the procedures described for those who agree to participate are the same procedures that are conducted as part of the course requirements for Dr. Houmanfar's Psychology 101 course.

Independent Variable

The primary independent variable of the study was the type of instruction that the student participant received each week. The two variations possible include whether they had access to AR technology or if they were only exposed to the traditional lecture method. In a traditional lecture session, participants were exposed to a lecture format that consisted of a standard Powerpoint presentation. The instructor read the information on the Powerpoint and elaborated on the concept presented. After the instructor covered the entire slide, he or she would move on to the next. Once the instructor finished a section (usually 1-4 slides) a prompt slide was exposed asking the student participants if they had any questions. The instructor would also verbally ask if anyone had questions. If there were questions, the instructor answered them as accurately as possible and then moved on to the next slide. If there were no questions, the instructor would move on. In order to ensure treatment integrity, slides with a prompt for questions appeared after each section to remind instructors to prompt the student participants for questions. The reminder slide also contained a brief outline of the section previously discussed. If a participant had a question, the instructor answered it to the best of his or her ability. If there were no questions the instructor would move on. (Appendix A)

An AR session began with the instructor announcing that the session would require the student participant's clickers. The same procedure was used, in terms of reading the slides and elaborating on the material, and the same Powerpoint presentation was used but an additional task was included. The student participants would not be asked if they had any questions on the material. Instead, they were tolded to answer a multiple choice question with four possible answers by submitting their answer using their clicker. The question directly related to a topic that was discussed during the last section of the Powerpoint slides. Once the student participants have answered the review question the instructor would post the answer distribution over the Powerpoint presentation.   The distribution depicted the number of students who answered each possibility. If more than 80% answered the question correctly, the instructor praised the class and explained why the answer was correct. If fewer than 80% of the class got it correct, the instructor identified the correct answer, explained why it was correct as well as redisplay the slide the correct answer was found. Once the instructor had followed these steps, they would ask if the student participants had any questions on the material. If a participant had a question, the instructor would answer it to the best of his or her ability. If there were no questions, the instructor would move on. (Appendix B)

The focus of the AR sessions was the overt responses to the review questions presented. It is possible that each individual participant could actively respond to the material in a different way, such as taking notes or highlighting important parts of the book, but these are all possible AR techniques seen in traditional classrooms. The variable in AR sessions that contrasts with the traditional teaching sessions is the individual electronic responses from each student participant, which can be instantly recorded.

After both types of session, an in-class quiz was conducted on the material covered in the chapter. Both types of sessions took this in-class quiz using the EVS, regardless of whether they were exposed to AR or not. This was to ensure that all student participants got their money's worth out of the clicker. In addition, the EVS is a much more accurate technology to identify correct answers and record them correctly and efficiently.

A schedule was posted designating which section had access to AR for each discussion week. The instructor would confirm if a section should be exposed to AR before every discussion session. In addition, the primary researcher confirmed that each instructor administered the variable correctly, and removed any participant data if the variable was administered incorrectly. The instructor's response to the percentage of student participants that answered an AR question correct was also a variable that depended on student participant performance. The participants were not exposed to identical questions at any point. Similar questions could be asked, but the wording of the question and the order of the answer was different.

Experimental Design

There were two levels of the independent variable, exposure to AR in addition to traditional methods and only exposure to traditional methods. A quasi-experimental counterbalanced reversal between group design (Appendix C) was used to isolate the educative value of the EVS between the first half and second half of a semester long class. Simply stated, Section 1 was exposed to AR during the first half of the semester and Section 2 was not. During the second half of the semester the application of AR rotated and Section 2 was exposed to the AR and Section 1 was not. It is important to note, however, that both sections used the clickers at the end of the discussion session.

It was crucial to ensure that the correct section was exposed to the correct independent manipulation during the correct week. A sheet was posted, informing proctors and instructors which discussion session would be exposed to which variable. Discussions were isolated according to section, so that members of Section 1 were only able to attend discussion with other members of Section 1, which prevented interaction between sections. Student participants were only allowed to take the discussion quiz once, so they would not be repeatedly exposed to AR.

Technical difficulties were also accounted for. Historically, there have been incidences were the clicker software does not function correctly. This can result in lost data, or failure to collect data in the first place. A protocol was put in place to ensure that data were not lost, and that the failure was reported. Typically, the equipment malfunction results in manual collection and analysis of AR sessions with EVS. The protocol stated that each student participant was handed a slip of paper and was be instructed to answer the AR question on the paper, using a pen or pencil. The papers were then collected after class, scored by a proctor and entered into WebCT manually.

The instructor would indicate the sessions that did not use the technology, which were reported when discussing results.

The curriculum in Psychology 101 is non-linear in the sense that demonstrating comprehension in one chapter is not required in order to understand concepts in another chapter. The chapters are tied together by the use of psychological methodology and an underlying focus on evolutionary explanations for human phenomena, but the order in which the chapters are taught will not affect a student's ability to learn information from the next chapter. This is beneficial in the sense that understanding concepts in one chapter will not influence the student participant's ability to understand concepts from another chapter.

The experiment took ten weeks to complete, not including the three weeks allocated for orientation (which includes the pre-test), and the week allocated for both the mid-term and final. The time frame was largely dependent on fitting the experiment into the semester, so the time allotted to each condition was predetermined and inflexible.

At the end of the semester, each student was asked to respond to a University survey regarding their experience in the Psychology 101 course. This survey is used to gather information on student satisfaction of the course, the content and the instructors. In addition, student participants were asked to rate how much they enjoyed using clickers, if they thought the clickers were useful and if they thought that clickers helped them interact with the material and the instructor more effectively. They were asked if they would like to use clickers again and if they would recommend clickers to other course instructors.

Data Analysis

Data analysis was facilitated by the WebCT computer system, which was used to record and compile each student participant's individual score. The system can analyze each student participant's individual grade, as well as their scores grouped into sections.

Two main analyses were conducted. The first consisted of descriptive and inferential statistics describing the effect the independent variable had on each section. The descriptive data was graphed for visual analysis. In addition, an ANOVA was conducted to determine if the differences between each section were statistically significant.

The second analysis consisted of visual analyses of 40% of participants from each section. The participants were chosen for analysis retroactively based on their average weekly quiz scores for the entire semester. Participants who scored in the top 20% of their section were analyzed, as well as the bottom 20%. This analysis is beneficial for application, because it examined the effects that AR with an EVS has in terms of level of student ability. An advanced placement course instructor may decide not to use an EVS if the results of this experiment show that it does not increase performance in participants who are more advanced in terms of knowledge of the material. It is important to statistically validate the use of clickers, but it is equally important to analyze the data to observe the minute changes that statistical analyses will not depict.

Results

Section Analysis

The average scores for the Discussion Quiz for each section across the semester are depicted in Figure 1 (see also Appendix D). The discussion score, out of six possible points, depicts a basic trend which both sections follow; decreasing from the first half of the semester to the middle of the semester and then increasing from the middle of the semester to the end of the semester. There were slight differentiations between the sections across different weeks. Section 1 appeared to do marginally better during weeks one, two, and three which is when the students in that section were exposed to the EVS during discussion. However, Section 1 was equivalent to Section 2 during week 4 and was lower than Section 2 during week five. The last half of the semester depicts Section 1 scoring slightly below or equivalent to Section 2, which is again expected since Section 2 was exposed to the EVS during discussion for those two weeks. A one-way ANOVA was conducted on the scores for each section per week to determine if the variance between the two groups was statistically significant. Week three's scores were significant, $F\ (1, 208) = 4.184$, $p = .042$, while all other discussion quiz scores that were analyzed were not significant as seen in Table 1.

The average scores for the First Weekly Chapter Quiz for each section across the semester are depicted in Figure 2 (see also Appendix D). The scores were collected after each participant attended their discussion, to ensure each participant was exposed to the material. The chapter quiz score, out of fifteen possible points, demonstrates a trend similar to that of the discussion quiz, with a decrease from the beginning to the middle of the semester and an increase from the middle to the end of semester. However, the increase from the middle to the end of the semester was not as pronounced as that seen with the discussion quiz scores. There was limited differentiation seen across sections during the first half of the semester, with a slight drop in the average score of Section 1 during the fifth week. After the middle of the semester, Section 2 scores slightly higher for week six, seven and eight. A one-way ANOVA was conducted on the scores for each section per week to determine if the variance between the two groups was statistically significant. Week four's scores were significant, F (1, 307) = 4.711, p = .031, while all other chapter quiz scores analyzed were not significant as seen in Table 2.

The average scores for the Final Weekly Chapter Quiz for each section across the semester are depicted in Figure 3 (see also Appendix D). The final scores were collected from the participants' final grade for each quiz, which was determined by the participants' highest score on each respective quiz. The final chapter quiz score, also out of fifteen points, mirror the averages for the first chapter quiz but the differences between the two sections appear more exaggerated. This was especially true during weeks four, six, seven and eight. Week Four shows that Section 2 dropped far below Section 2. They increased slightly afterwards, catching up to Section 2 in week five. Weeks Six and Seven depict a substantial differentiation, with Section 2 scoring much higher than Section 1. Week Eight also shows a differentiation, though it was not as pronounce and both Sections demonstrate a decrease compared to the previous week. A one-way ANOVA was conducted on the scores for each section per week to determine if the variance between the two groups was statistically significant. No statistical significance was found for any of the weeks (see Table 3).

The average number of times participants in each section took a quiz can be seen in Figure 4 (see also Appendix D). There was a total of nine possible retakes per week. A slight increase can be seen during the first half of the semester, appearing to peak at Week Five. There then was a slight decrease at the beginning of the second half of the semester with the number of quizzes remaining about the same. A one-way ANOVA was conducted on the scores for each section per week to determine if the variance between the two groups was statistically significant. No statistical significance was found for any of the weeks (see Table 4).

The results of the final scores were analyzed in relation to the number of quizzes student participants took. Participants were divided into two groups; those who took the weekly chapter quiz once and those who took the weekly chapter quiz more than once. These results were compared across both sections in Figure 5 (see also Appendix D). Both groups (those who took the quiz once and those who took it more than once) depict the downward trend seen the first half of the semester in previous graphs. There appears to be a distinct variation between the groups, with the exception of quiz six where participants in Section 2 who took the quiz once scored the highest compared to the other groups.

Group Analysis

A total of 40% of the 321 student participants whose data was used in the Section Analysis were analyzed in smaller groups of 20%. Participants that scored within the top 20% on their average chapter score were added to the High Performance Group, and the participants that scored in the bottom 20% on their average chapter quiz score were added to the Low Performance Group. A total of 68 student participants were included for Section 1's groups (34 participants in High Achievement, 34 participantss in Low Achievement) and 58 participants were included for Section 2's groups (29 participants in High Achievement, 29 participants in Low Achievement).

Figure 6 depicts the average discussion quiz for each High and Low Achievement Group in both sections (see also Appendix D). The variability of the High Achievement

Group across sections is visibly less than the variability of the Low Achievement Group across sections. The High Achievement Groups for both sections did not demonstrate the same decrease towards the middle of the semester, as is seen in the Section Analysis. There was a very slight downward trend, but it is much less noticeable for the High Achievement Groups. The decrease in performance towards the middle of the semester was also less noticeable for the Low Achievement Groups, especially for Section 2, which shows the greatest amount of variability. All four groups appear to improve on their quiz scores from the middle of the semester to the end of the semester. A one-way ANOVA was conducted on the scores for each section's achievement group (High Achievement and Low Achievement were analyzed in relation to the same group for the opposite section) per week to determine if the variance between the achievement groups were statistically significant. No statistical significance was found for any of the weeks (see Table 5 and 6).

Figure 7 depicts the average first chapter quiz score for each High and Low Achievement Group in both sections (see also Appendix D). The variability seemed to be comparable across the two achievement groups. The performance of the two groups also appeared consistent across the semester, without the typical mid-semester decline in scores. The High Achievement Groups' scores remain around 13 points out of 15, and the Low Achievement Groups' scores remain around eight points out of 15. No substantial differentiation was observed when comparing the two levels of achievement group across sections. A one-way ANOVA was conducted on the scores for each achievement group across sections per week to determine if the variance between the two groups was statistically significant. No statistical significance was found for any of the weeks (see Table 7 and 8).

Figure 8 shows the average final chapter quiz score for each High and Low Achievement Group in both sections (see also Appendix D). The variability was similar to that seen in the first chapter quiz, with the High Achievement Groups showing less variability in their scores than the Low Achievement Groups. Additionally, the High Achievement Groups showed even less variability than they did in their scores for the first chapter quiz, while the Low Achievement Groups show approximately the same variability across those two scores. The scores of both groups did not depict the typical drop towards the middle of the semester, but remained fairly consistent, with the High Achievement Group remaining above an average of 14 points throughout the semester, and the Low Achievement Group remaining between 11 and 7 points. A one-way ANOVA was conducted on the scores for each achievement group across sections per week to determine if the variance between the two groups was statistically significant. No statistical significance was found for any of the weeks (see Table 9 and 10).

The average number of times participants in each section took a quiz can be seen in Figure 9 (see also Appendix D). The Low Achievement Groups' quiz averages were maintained at between two and three quizzes a week. The High Achievement Groups' averages show more variability. Section 1's High Achievement Group appears to have a slight downward trend, with a number of increases and subsequent decreases as the semester progresses. Section 2's High Achievement Group has a sharp increase from Week 1 to Week 3, and then shows a gradual decrease for the rest of the semester. A one-way ANOVA was conducted on the scores for each section per week to determine if the variance between the two groups was statistically significant. No statistical significance was found for any of the weeks (see Table 11 and 12).

Additional Analyses

In addition to the analyses discussed, supplementary analyses were conducted to account for other possible extraneous interactive variables that might affect the results. Data from student participants who attended individual instructors were evaluated. Two instructors were chosen based on their similarity in teaching schedules. In addition, Instructor 1 had taught the material for two semesters prior to the current experiment, while Instructor 2 had not taught the material before. The comparison between experienced and new instructor was examined. Student participants were included in the data set if they attended the instructor's discussion sessions for eight out of ten weeks during the semester. Instructor 1 had three participants included for Section 1 and 28 participants included for Section 2. Instructor 2 had eighteen participants included for Section 1 and eleven participants included in Section 2.

The differences between the two instructors' participant's scores for the discussion quiz can be seen in Figure 10 and 11 (see also Appendix D). Instructor 2 has more variability in the results, with a sharp decrease in scores halfway through the semester. After discussion quiz 5 there appears to be a substantial differentiation between the two sections. Instructor 1's scores has a moderate amount of variability but the participants scored, on average, higher than those from Instructor 2's sessions. The results of the first weekly quiz attempt by participants from the two instructors can be seen in Figure 12 and 13 (see also Appendix D). Comparison across instructors depicts Section 2 participants as more stable in their scores than those from Section 1, which depicts more variability. The final scores compared across both instructors demonstrate a decrease in variability within each group, as well as between groups for the instructors, as can be seen in Figure 14 and 15 (see also Appendix D).

The two previous semesters were selected for comparison, as seen in Figures 16 (see also Appendix D). The final scores were compared across these three semesters. The class in Fall 2008 received clickers after discussion for their discussion quiz, but not during discussion. The class of Spring 2009 received alternating clickers similar to the current study, but instead of changing once a semester, the alternation occurred every week. The comparison demonstrates moderately equal scores the first three weeks. After those three weeks, the scores collected for the current experiment begin to differentiate from the previous two semesters, consistently scoring higher. The downward trend from the beginning of the semester to the middle is not as steep with the current experimental data, and the jump after the first half of the semester is much more substantial.

Statistical analyses were conducted on the comparison data to determine the significance of these differences (Table 13). There was moderate significance during weeks 2 and 3, $F(1,835) = 4.242$, $p = .04$ and $F(1,834) = 4.667$, $p = .031$, when comparing the differences of the Spring 2009 and Fall 2009 semester which were the two semesters in which student participants received an AR treatment during discussion. In addition, these two semesters had discussion quizzes scheduled directly following discussion.

A statistical comparison of the Spring 2009 and Fall 2009 semesters was also conducted (Table 14). The first week was not statistically significant. The second and third week was significant at $p < .05$ and every week after the third week was significant at the $p < .01$ (see Table 15). The Spring 2009 semester was conducted directly after the Fall 2008 semester and was the first semester to introduce the EVS software into the classroom.

Lastly, a statistical comparison was conducted to determine significant differences between Fall 2008 and Fall 2009 semesters (Table 15). The first semester, Fall 2008 did not receive an experimental EVS treatment, while the AR was required during discussion for Fall 2009. The first three weeks were not statistically significant, while the fourth week was significant at $p < .05$. Significance was found for the five week through the tenth week at $p < .01$ (see Table 16).

Social validity was collected in conjunction with the course evaluation that students are required to fill out. The responses were anonymous, so the results presented are reflective of the entire class, not only those who agreed to participate in the experiment and not separated by section. The results are depicted in Table 16. A total of 86% of students believed that the AR slides were very or moderately helpful when preparing for quizzes. Only 4% believed that clickers were less helpful, and 10% believed that clickers made no difference. A self-report preference assessment revealed that 83% of students preferred the AR slides embedded in the discussion presentation, while 17% did not. Sixty one percent of students believed that the EVS affected their overall grade, 64% said it helped them with their discussion quiz score and 71% said it helped them with their weekly quiz score. Over three fourths of the students surveyed said that they would recommend EVS in future courses.

Discussion

The utilization of EVSs as an educational tool could have implications in college classroom settings. The ability to require student's active responses and measure those responses in an accurate way is a relatively new technology to the field of education. The current study offered an initial step in analyzing the effect of a relatively new instructional technology on students' academic performance. In short, our findings provided a starting point for the examination of EVSs as a source for promotion of active responding, with an emphasis on objective, measurable and functional outcomes.

There appeared to be some differentiation between Section 1 and Section 2 quiz scores during the target semester, but the differentiation that was detected was not statistically significant. This differentiation, even though statistically insignificant, should not be overlooked. More specifically, the visual differentiation (associated with quiz performance of the AR group compared to the Traditional group) during the second half of the semester hints at a potential interactive effect of EVS on performance. On average, the largest differentiation seen when applying the clickers across the two sections was approximately one-half point. If this differentiation was consistent across the entire semester, there was potential of increasing a student's score by five points. It is important to note that an average five point increase across two large classroom sections could possibly increase a large number of students' letter grade, changing a lower letter grade to a higher one. More specifically, the small cumulative impact for students whose scores are below the cut off could result in dramatic effect in terms of final grade.

The three week increase after the first half of the semester warrants further attention, as it is the most noticeable differentiation between groups and it provides the greatest potential for the increase in student grades, as mentioned above. It should be noticed that Group 1, who received the AR treatment during the first half of the semester, did not display an increase in quiz scores when they were first exposed to the EVS. Group 2, however, did display an increase lasting for approximately three weeks after their first exposure to the EVS. This change might be attributable to a few different phenomena. First, Group 1's exposure to the EVS was implemented during a time that many novel stimuli were also presented. These included the start of class, and the new system of learning in which they had to become accustom. Group 2, on the other hand, was first exposed to the AR Treatment following their adjusted exposure to other class related contingencies. In short, the timing of exposure to EVS might have occasioned an additive effect on student participants' performance in Section 2.

Another possible explanation for the difference in average quiz scores between sections could pertain to concurrent exposure to EVS during and after discussion sessions for Section 1 (during the first half of the semester) compared to additive exposure to EVS for Section 2 (during the second half of the semester). More specifically, Group 1 received the AR treatment embedded in the Powerpoint while also taking their discussion quiz with the EVS. Group 2 did not receive the AR treatment embedded in the Powerpoint until after five weeks, accessing the EVS when taking the discussion quiz at the end of the class. In that regard, Section 2 may have acquired more experience interacting with the EVS than Group 1 before exposure to the embedded content.

The change in performance seen during the target semester should also be examined in the context of the historical data collected from previous semesters (Figure 16 & Figure 17, see also Appendix D). There appears to be an increase in overall student performance when the EVS was applied in the Fall 2009 semester compared to previous semesters when the EVS wasn't applied, as seen in Spring 2009, as well as the semester in which the current study took place, Fall 2009. This increase is more apparent when each semester of Fall 2009 is separated and examined in relation to the other two semesters. There is a noticeable differentiation during the last half of the semester, where Section 2 of the Fall 2009 semester increases substantially. This increasing trend itself doesn't support the use of an EVS, but it should still be considered a possible contributing factor.

These previous semesters encountered AR in different ways. The Fall 2009 semester has already been discussed extensively. During Spring 2009, a program evaluation was conducted on the EVS. The program evaluation utilized a design similar to Fall 2009. During the evaluation two class sections were alternately exposed to the same AR treatment with the EVS as Fall 2009. However, the program evaluation utilized a counterbalance design so that on the odd weeks, one section received an EVS treatment and on even weeks, the other section received EVS. Both sections received a discussion quiz consisting of six questions at the end of every discussion, similar to the discussion quiz participants received during Fall 2009. In short, the program evaluation provided the same access to AR through the EVS as Fall 2009, but the exposure was alternated every week opposed to every five weeks.

During Fall 2008, the EVS was utilized exclusively for the discussion quiz across all sections, without additional slides for AR. The two sections from this semester were exposed to the quiz questions during the presentation, in a sequence similar to the slides used for the AR treatment, but these questions were graded and counted for discussion quiz points. No questions were presented at the end of the discussion.

The difference seen between the results of three semesters is indicative of the amount of exposure to AR using the EVS. In Fall 2008, all students were required to respond to a total of 60 questions during the semester. These interactions with the EVS were graded and students knew that their responses would count towards their final grade in the class. In Spring 2009, all students were required to respond to a total number of 90 questions; 60 for ten weeks of discussion quizzes and 6 each week in alternation (one section received 6 questions on odd weeks, the other section received 6 questions on even weeks) for a total of 30 questions. Participants of the current study in Fall 2009 also received 90 questions; 60 for ten weeks of discussion quizzes and 6 questions each week for a five week period (one section received 6 questions for the first five weeks, the other section received 6 questions for the last five weeks) for a total of 30 questions. The two semesters with the highest number of opportunities to respond actively also tend to have the highest final scores, which is supported by statistical significance (Tables 13, 14 & 15).

The differences between the two semesters that have the highest number of opportunities to respond can be discussed multiple ways. Most important is the rapidly alternating AR via EVS manipulation, which contrasts with target semester's (Fall 2009) much slower alternation. This alternation provided only one opportunity to access the additional AR slides before the treatment is removed. In addition, it is important to note that there was a difference between the two sections concerning the number of quizzes students took during Spring 2009 (see Figure18). The average number of quizzes taken by Section 1 during Spring 2009 was consistently higher than Section 2. This differentiation of quizzes taken was not observed in Fall 2009 (see Figure 4).

The option to retake quizzes up to nine times has historically been one of the most powerful institutional variables in the target instructional setting. The effect of multiple retakes (Figure 5) is evident with increases in scores in the case of participants who took more than one quiz opportunity. However, even with the differentiation between the two groups of quiz takers, similar patterns of quiz score over the semester are seen. Section 2 achieved higher scores compared to Section 1 beginning the second half of the semester. This shows that Section 2 achieved higher scores compared to Section 1 for their respective groups (1 quiz taken and more than 1 quiz taken). This difference is seen specifically in the three weeks that depict a differentiation in the analysis across class sections (see Figure 3).

Another possible interacting factor is the differences between instructors. Though a protocol was implemented to standardize the lecture as much as possible and control for these differences, they are not likely to be completely controlled. The variations seen in Figures 10 through 15 should be considered while analyzing the performance graphs, because they depict the variability seen between the two instructors. However, it is important to note the number of student participants included in that analysis for each instructor and each section. The population difference could have an effect on the variability seen in the results.

All the data discussed so far cannot be corroborated with previous research since the current study is the only analysis that includes student grades over the semester. A majority of studies on AR with an EVS only reported social validity data (Draper & Brown, 2004; Wit, 2003). The results of these studies show that typically, three quarters of students think that clickers help them in the classroom. The results of this study substantiate the previous findings (Table 13) by providing data associated with student satisfaction. In addition, our findings indicated participants' identification of EVS as a useful tool for studying and active responding.

Potential Confounds

Traditional classes would allow for more control of interactive variables that were not able to be accounted for in the current study. For instance, an important variable with a potential effect on the outcome measures was instructor differences. There were five different instructors that presented the material during the semester, so participants could have been exposed to five different instructors over the semester. In a traditional large enrollment class, one professor would present a lecture with all the students in the same room.

A second variable was the additional supplemental materials that were offered to students. These materials, such as self-tests and outlines of the lecture, proctoring and retake opportunities are not typically offered in traditional courses. Participants' exposure to these items was not controlled or measured in this study, and therefore had an unseen effect on the results.

Student participants were quizzed on the content of discussion sessions frequently. They were required to take a discussion quiz and a chapter quiz every week. This frequent and relatively rapid exposure to the course content encourages active responding in itself, without the help of the EVS. The exposure also provided rapid feedback on student performance. In traditional classrooms the instructor tests student knowledge infrequently. Typically, large enrollment courses only provide two opportunities to demonstrate what they have learned, a mid-term and a final. This not only encourages procrastination phenomena, but also provides little opportunity for students to obtain feedback on their performance in the class. In essence, the course utilized for this research offers numerous opportunities for students to respond to the course material and receive immediate and contingent performance feedback. These inherently established instructional variables most likely masked the potential effect of EVS on student performance. Considering the current structure of the class in which this research was conducted, eliminating these potentially confounding variables would be detrimental to the course, as well as to the students. The course was constructed in such a way as to give students multiple opportunity and resources to succeed. Restricting these resources could not have been done without severely disadvantaging the students who were enrolled in the course.

In short, academic performance is multiply controlled in an instructional system such as Psychology 101 at UNR. Therefore, separating the effect of each instructional variable on performance in this type of setting is methodologically challenging. In that regard, basic analog studies that would offer a controlled process account of EVS' promotion of active responding and improved academic performance seem warranted.

Future Research

The previous research pertaining to active responding with other techniques (e.g. choral responding, guided notes and response cards) has been conducted in traditional classroom settings (Austin, Lee & Carr, 2004; Beekes, 2006; Gardner, Heward & Grossi, 1994; Malanga & Sweeney, 2008; Narayan, Heward, Gardner, Courson & Omness, 1990). In that regards, it would be beneficial to examine the effect of EVS and AR in a more controlled, traditional class since the current study was conducted in a nontraditional classroom environment. It is feasible that the small effects demonstrated in the current study may be amplified when using the same methodology in a traditional setting where the multi effect of other interactive instructional variables (i.e., immediate feedback, retake opportunity, etc) are not present.

Implementing the methodology utilized in the current research to a more traditional setting could provide a more accurate comparison with past analyses, since the number of variables that have a potential interactive effect would most likely be decreased in a traditional classroom. The instructor's student interaction would be controlled since all students would be exposed to a single instructor. The number of interactions students could have with the instructor, both in and out of the classroom, would be minimal due to time constraints and the large number of students. Moreover, the amount of delivered feedback from testing or quizzing would also be minimal in a traditional classroom setting, and more reflective of settings traditionally used to study active responding techniques.

Most importantly, the ability to actively respond during a large enrollment class that does not typically allow for such interactions to occur could provide a context for demonstrating the utility of EVS in terms of student performance. In fact, the traditional classroom configuration is the very environment that the EVS was designed for. It is also important to continue the empirical analyses that would determine the additive or interactive effects of AR and EVS on student performance in courses that offer instructional enhancement technology (such as the setting associated with this study). In that regard, follow-up analyses associated with the current study has continued (Spring 2010). The current design is the same as the previous design, but instead of the students being prompted for questions during the Traditional treatment, the same question that appears during the AR treatment also included. The students are then asked to answer the question in a more traditional way, such as one student raising their hand and being called on to answer. Since all students are exposed to the same questions, only the application of the EVS is different.

In addition, Psychology 101 will not be implementing the use of an EVS at all next semester (Fall 2010). This group will be used as a non-equivalent comparison group whose differences will be analyzed in relation to the previous three semesters. The primary purpose of the post-hoc analysis between quiz performance in Fall 2010 and Fall 2009 will be to determine the effect of EVS' usage at the class level. This is especially important since student participants from the current study were exposed to the EVS throughout the semester since all discussion quizzes were conducted with EVS technology. In short, the lack of exposure to EVS technology during Fall 2010 could possibly demonstrate any differences seen by exposing students to EVS technology in any capacity or not.

Future research should consider including additional experimental groups to test EVSs. The current study only analyzed the effects that an EVS had on two groups. Additional groups to analyze during the same semester could provide excellent insight to the effect of EVSs have on AR and class grades. Ideally, four sections would be run in one semester. One session could act as a control group, while another could receive the EVS treatment the entire semester. The other experimental groups could be set up as they were during Fall 2009, with alternating treatments. This design would control for many different variables that cannot be controlled when comparing different semesters, such as demographics, historical events and changes in weather. The comparison between the groups that alternate AR and traditional classes could depict a clearer distinction when contrasted to the group that received AR the entire semester, and the group that didn't receive AR at all.

As mentioned earlier, it would be pertinent to examine the use of clickers more molecularly. The current study examines the outcome of AR with EVS at a very general level, concerned with only large scale aggregated data. These data present the outcome of a very complex set of behaviors so that it is difficult to determine if AR is occurring. Single subject research on the effect of an EVS would provide a process account by which AR is prompted by usage of EVS technology. This molecular analysis might provide additional understanding of students' responding during the usage of EVS, and demonstrate important variables essential to the process of responding actively with an EVS. Further research on this subject could utilize the methodology seen in studies of the other techniques of AR, such as response cards and guided notes (Cavanaugh, Heward, & Donelson, 1996; Heward, Courson, & Narayan, 1989; Malanga, & Sweeney, 2008), to provide a comparison across technologies associated with active responding (e.g., systematic comparison between response card, guided notes and EVS utilization) as well as continue to analyze processes involved in actively responding with an EVS.

Conclusion

The current results provide an initial evaluation of the effect of EVSs in a self-paced, individualized classroom setting. The addition of an EVS to a setting with a multitude of other tools to facilitate learning might increase students' scores, but results are inconclusive since it is difficult to separate the effects of contributing variables.

However, the implemented methodology can be replicated in other classroom settings and clickers can continue to be analyzed.

In summary the results of this study have empirical and applied implications when examined in conjunction with previous semesters. The increase in student performance seen across semesters that applied the EVS, combined with the data found within the Fall 2009 semester provide a context for future empirical exploration of EVS as a potential source of active responding in classroom settings (particularly the ones with large enrollment). As mentioned earlier, the above findings have prompted follow up analyses that would offer further clarification regarding the impact of AR via EVS on student performance in the target instructional setting. Further experimental analyses in controlled laboratory environments as well as concurrent between group comparisons (in one academic semester) in classroom settings would provide additional examinations of EVS as a potential source of active responding.

References

Austin, J. L. (2000). Behavioral approaches to college teaching. In J. Austin & J. E. Carr (Eds.), Handbook of Applied Behavior Analysis (449-473). Reno, NV: Context Press.Benjamin, L. T. (1988). A history of teaching machines. *American Psychologist, 43*(2), 703-712.

Austin, J. L., Lee, M. & Carr, J. P. (2004). The effects of guided notes on undergraduate students' recording of lecture content. *Journal of Instructional Psychology, 31*(4), 314-320.

Beekes, W. (2006). The 'millionaire' method for encouraging participation. *Active Learning in Higher Education, 7,* 25-36.

Cavanaugh, R. A., Heward, W. L. & Donelson, F. (1996). Effects of response cards during lesson closure on the academic performance of secondary students in an earth science course. *Journal of Applied Behavior Analysis, 29*(3), 403-406.

Christle, C. A., & Schuster, J. W. (2003). The effects of using response cards on student participation, academic achievement, and on task behavior during whole-class, math instruction. *Journal of Behavioral Education 12*(3), 147-165.

Cue, N. (December, 1998) *A Universal Learning Tool for Classrooms?* Proceedings of the First Quality in Teaching and Learning Conference, Hong Kong SAR, China. Retrieved February 15, 2009, from http://celt.ust.hk/ideas/prs/pdf/Nelsoncue.pdf.

Dewey, J. (1916). Democracy and education. New York: Macmillan.

Draper, S. W., Cargill, J. & Cutts, Q. (2002). Electronically enhanced classroom interaction. *Australian Journal of Educational Technology,18*(1), 12-23.

Draper, S. W. & Brown, M. I. (2004). Increasing interactivity in lectures using an electronic voting system. *Journal of Computer Assisted Learning, 20,* 81-94.

Dugresne, R. J., Gerace, W. J., Leonard, W. J., Mestre, J. P. & Wenk, L. (1996). Classtalk: A classroom communication system for active learning. *Journal of Computing in Higher Education, 7,* 3-47.

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

Gardner, R., III, Heward, W., & Grossi, T. A. (1994). Effects of response cards on student participation and academic achievement: A systematic replication with inner-city students during whole-class science instruction. *Journal of Applied Behavior Analysis, 27*(1), 63-71.

Gray, P. (2007) *Psychology, Fifth Edition*. 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*, 165-172.

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 Pyschics*, *66*(1), 64-74.

Heward, W. L., Courson, F. H., & Narayan, J. S. (1989). Using choral responding to increase active student response during group instruction. *Teaching Exceptional Children, 21*(3), 72-75.

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

Kennedy, G. E. & Cutts, Q. I. (2005). The association between students' use of an electronic voting system and their learning outcomes. *Journal of Computer Assisted Learning, 21* 260-268.

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. (2008). 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.

Marmolejo, E. K., Wilder, D. A., & Bradley, L. (2004). A preliminary analysis of the effects of response cards on student performance and participation in an upper division university course. *Journal of Applied Behavior Analysis, 37*(3), 405-410.

Miller, M. L. & Mallot, R. W. (1997). The importance of overt responding in programmed instruction even with added incentives for learning. *Journal of Behavioral Education, 7*(4), 497-503.

Narayan, J. S., Heward, W. L., Gardner, R., III, Courson, F. H., & Omness, C. (1990). Using response cards to increase student participation in an elementary classroom. *Journal of Applied Behavior Analysis 23*, 483-490.

Pressey, S. L. (1926). A simple device which gives tests and scores—and teaches. School and Society, 23, 373–376.

Shabani, D. B. & Carr, J. E. (2004). An evaluation of response cards as an adjunct to standard instruction in university classrooms: A systematic replication and extension. *North American Journal of Psychology, 6*(1), 85-100.

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.

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.

Tudor, R. M. (1995). Isolating the effects of active responding in computer-based instruction. *Journal of Applied Behavior Analysis*, *28*(3), 343-344.

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

White, O. R. (1986) Precision teaching: Precision learning. *Exceptional Children, Special Issue. 52*(6), 522-534.

Wit, E. (2003). Who wants to be… the use of a personal response system in statistics training. *MSOR Connections, 3*(2), 14- 20.

Table 1

ANOVA: Fall 2009 Section 1 and 2 Discussion Quiz Scores by Week

		Sum of Squares *SS*	df	Mean Square *MS*	*F*	Sig./Effect
W1_DQ_Score	Between Groups	1.348	1	1.348	2.438	.119
	Within Groups	160.931	291	.553		
	Total	162.280	292			
W2_DQ_Score	Between Groups	1.881	1	1.881	2.399	.123
	Within Groups	228.140	291	.784		
	Total	230.020	292			
W3_DQ_Score	Between Groups	3.057	1	3.057	4.184	.042 .015
	Within Groups	204.578	280	.731		
	Total	207.635	281			
W4_DQ_Score	Between Groups	.311	1	.311	.318	.573
	Within Groups	289.934	296	.980		
	Total	290.245	297			
W5_DQ_Score	Between Groups	3.892	1	3.892	2.935	.088
	Within Groups	395.228	298	1.326		
	Total	399.120	299			
W6_DQ_Score	Between Groups	.268	1	.268	.210	.647
	Within Groups	338.211	265	1.276		
	Total	338.479	266			
W7_DQ_Score	Between Groups	.001	1	.001	.001	.971
	Within Groups	268.076	307	.873		
	Total	268.078	308			
W8_DQ_Score	Between Groups	1.284	1	1.284	1.534	.216
	Within Groups	242.729	290	.837		
	Total	244.014	291			

W9_DQ_Score	Between Groups	.004	1	.004	.006	.939
	Within Groups	176.066	285	.618		
	Total	176.070	286			
W10_DQ_Score	Between Groups	.016	1	.016	.020	.887
	Within Groups	216.793	270	.803		
	Total	216.809	271			

Table 2

ANOVA: Fall 2009 Section 1 and 2 First Quiz Scores by Week

		Sum of Squares *SS*	*df*	Mean Square *MS*	*F*	Sig./Effect
W1_First_WQ	Between Groups	.004	1	.004	.001	.982
	Within Groups	2150.026	304	7.072		
	Total	2150.029	305			
W2_First_WQ	Between Groups	.277	1	.277	.042	.838
	Within Groups	2018.458	304	6.640		
	Total	2018.735	305			
W3_First_WQ	Between Groups	.457	1	.457	.070	.792
	Within Groups	2002.611	307	6.523		
	Total	2003.068	308			
W4_First_WQ	Between Groups	4.352	1	4.352	.700	.403
	Within Groups	1926.866	310	6.216		
	Total	1931.218	311			
W5_First_WQ	Between Groups	32.869	1	32.869	4.711	.031 .015
	Within Groups	2142.069	307	6.977		
	Total	2174.939	308			
W6_First_WQ	Between Groups	9.675	1	9.675	1.834	.177
	Within Groups	1450.282	275	5.274		
	Total	1459.957	276			
W7_First_WQ	Between Groups	8.045	1	8.045	1.539	.216
	Within Groups	1599.266	306	5.226		
	Total	1607.312	307			
W8_First_WQ	Between Groups	3.105	1	3.105	.480	.489
	Within Groups	1945.271	301	6.463		
	Total	1948.376	302			

W9_First_WQ	Between Groups	.211	1	.211	.033	.855
	Within Groups	1878.438	297	6.325		
	Total	1878.649	298			
W10_First_WQ	Between Groups	.208	1	.208	.038	.846
	Within Groups	1669.600	301	5.547		
	Total	1669.809	302			

Table 3

ANOVA: Fall 2009 Section 1 and 2 Final Quiz Scores by Week

		Sum of Squares *SS*	*df*	Mean Square *MS*	*F*	Sig./Effect
W1_Final_WQ	Between Groups	.169	1	.169	.046	.830
	Within Groups	1118.072	305	3.666		
	Total	1118.241	306			
W2_Final_WQ	Between Groups	1.862	1	1.862	.471	.493
	Within Groups	1201.801	304	3.953		
	Total	1203.663	305			
W3_Final_WQ	Between Groups	.030	1	.030	.008	.930
	Within Groups	1204.300	307	3.923		
	Total	1204.330	308			
W4_Final_WQ	Between Groups	9.819	1	9.819	2.263	.134
	Within Groups	1345.152	310	4.339		
	Total	1354.971	311			
W5_Final_WQ	Between Groups	.602	1	.602	.146	.703
	Within Groups	1270.608	307	4.139		
	Total	1271.210	308			
W6_Final_WQ	Between Groups	8.952	1	8.952	3.061	.081
	Within Groups	804.290	275	2.925		
	Total	813.242	276			
W7_Final_WQ	Between Groups	6.568	1	6.568	1.968	.162
	Within Groups	1027.770	308	3.337		
	Total	1034.339	309			
W8_Final_WQ	Between Groups	2.508	1	2.508	.565	.453
	Within Groups	1337.142	301	4.442		
	Total	1339.650	302			

W9_Final_WQ	Between Groups	.775	1	.775	.197	.657
	Within Groups	1163.309	296	3.930		
	Total	1164.084	297			
W10_Final_WQ	Between Groups	2.065	1	2.065	.545	.461
	Within Groups	1141.446	301	3.792		
	Total	1143.512	302			

Table 4

ANOVA: Fall 2009 Section 1 and 2 Number of Quizzes Attempted by Week

		Sum of Squares *SS*	*df*	Mean Square *MS*	*F*	Sig./Effect
W1_Quiz Attempts	Between Groups	.007	1	.007	.004	.948
	Within Groups	495.758	304	1.631		
	Total	495.765	305			
W2_Quiz Attempts	Between Groups	.468	1	.468	.196	.658
	Within Groups	726.029	304	2.388		
	Total	726.497	305			
W3_Quiz Attempts	Between Groups	2.916	1	2.916	1.072	.301
	Within Groups	835.324	307	2.721		
	Total	838.239	308			
W4_Quiz Attempts	Between Groups	1.268	1	1.268	.448	.504
	Within Groups	877.575	310	2.831		
	Total	878.843	311			
W5_Quiz Attempts	Between Groups	2.993	1	2.993	.889	.347
	Within Groups	1034.023	307	3.368		
	Total	1037.016	308			
W6_Quiz Attempts	Between Groups	.081	1	.081	.042	.838
	Within Groups	529.695	275	1.926		
	Total	529.776	276			
W7_Quiz Attempts	Between Groups	2.330	1	2.330	.887	.347
	Within Groups	809.412	308	2.628		
	Total	811.742	309			
W8_Quiz Attempts	Between Groups	1.703	1	1.703	.923	.338
	Within Groups	555.584	301	1.846		
	Total	557.287	302			

W9_Quiz Attempts	Between Groups	.625	1	.625	.287	.593
	Within Groups	648.344	297	2.183		
	Total	648.970	298			
W10_Quiz Attempts	Between Groups	.175	1	.175	.070	.791
	Within Groups	748.663	301	2.487		
	Total	748.838	302			

Table 5

ANOVA: Fall 2009 Section 1 and 2 High Achievement Group Discussion Quiz Scores by Week

		Sum of Squares *SS*	*df*	Mean Square *MS*	*F*	Sig./Effect
W1_DQ_Score	Between Groups	.036	1	.036	.094	.761
	Within Groups	22.653	59	.384		
	Total	22.689	60			
W2_DQ_Score	Between Groups	.471	1	.471	1.996	.163
	Within Groups	14.386	61	.236		
	Total	14.857	62			
W3_DQ_Score	Between Groups	.972	1	.972	3.365	.072
	Within Groups	16.761	58	.289		
	Total	17.733	59			
W4_DQ_Score	Between Groups	.304	1	.304	.352	.555
	Within Groups	51.905	60	.865		
	Total	52.210	61			
W5_DQ_Score	Between Groups	.332	1	.332	.354	.554
	Within Groups	56.265	60	.938		
	Total	56.597	61			
W6_DQ_Score	Between Groups	.175	1	.175	.173	.679
	Within Groups	56.721	56	1.013		
	Total	56.897	57			
W7_DQ_Score	Between Groups	.079	1	.079	.128	.721
	Within Groups	37.477	61	.614		
	Total	37.556	62			
W8_DQ_Score	Between Groups	1.392	1	1.392	2.576	.114
	Within Groups	31.341	58	.540		
	Total	32.733	59			

W9_DQ_Score	Between Groups	.487	1	.487	1.146	.289
	Within Groups	24.225	57	.425		
	Total	24.712	58			
W10_DQ_Score	Between Groups	.089	1	.089	.159	.691
	Within Groups	32.960	59	.559		
	Total	33.049	60			

Table 6

ANOVA: Fall 2009 Section 1 and 2 Low Achievement Group Discussion Quiz Scores by Week

		Sum of Squares *SS*	*df*	Mean Square *MS*	*F*	Sig./Effect
W1_DQ_Score	Between Groups	.582	1	.582	.789	.378
	Within Groups	39.127	53	.738		
	Total	39.709	54			
W2_DQ_Score	Between Groups	.030	1	.030	.026	.872
	Within Groups	57.517	51	1.128		
	Total	57.547	52			
W3_DQ_Score	Between Groups	.533	1	.533	.551	.461
	Within Groups	50.300	52	.967		
	Total	50.833	53			
W4_DQ_Score	Between Groups	1.926	1	1.926	1.725	.195
	Within Groups	60.289	54	1.116		
	Total	62.214	55			
W5_DQ_Score	Between Groups	1.875	1	1.875	1.622	.209
	Within Groups	60.125	52	1.156		
	Total	62.000	53			
W6_DQ_Score	Between Groups	2.946	1	2.946	1.833	.182
	Within Groups	78.741	49	1.607		
	Total	81.686	50			
W7_DQ_Score	Between Groups	3.561	1	3.561	3.873	.054
	Within Groups	54.242	59	.919		
	Total	57.803	60			
W8_DQ_Score	Between Groups	.814	1	.814	.626	.432
	Within Groups	68.931	53	1.301		
	Total	69.745	54			

W9_DQ_Score	Between Groups	.170	1	.170	.289	.593
	Within Groups	31.175	53	.588		
	Total	31.345	54			
W10_DQ_Score	Between Groups	.022	1	.022	.015	.903
	Within Groups	63.217	44	1.437		
	Total	63.239	45			

Table 7

ANOVA: Fall 2009 Section 1 and 2 High Achievement Group First Quiz Scores by Week

		Sum of Squares *SS*	*df*	Mean Square *MS*	*F*	Sig./Effect
W1_First_WQ	Between Groups	.001	1	.001	.000	.993
	Within Groups	499.935	60	8.332		
	Total	499.935	61			
W2_First_WQ	Between Groups	.677	1	.677	.125	.725
	Within Groups	330.593	61	5.420		
	Total	331.270	62			
W3_First_WQ	Between Groups	1.937	1	1.937	.364	.548
	Within Groups	319.047	60	5.317		
	Total	320.984	61			
W4_First_WQ	Between Groups	.504	1	.504	.098	.756
	Within Groups	315.210	61	5.167		
	Total	315.714	62			
W5_First_WQ	Between Groups	6.471	1	6.471	.825	.367
	Within Groups	478.609	61	7.846		
	Total	485.079	62			
W6_First_WQ	Between Groups	1.488	1	1.488	.468	.497
	Within Groups	178.029	56	3.179		
	Total	179.517	57			
W7_First_WQ	Between Groups	1.332	1	1.332	.307	.582
	Within Groups	251.602	58	4.338		
	Total	252.933	59			
W8_First_WQ	Between Groups	.665	1	.665	.127	.723
	Within Groups	309.565	59	5.247		
	Total	310.230	60			

W9_First_WQ	Between Groups	8.113	1	8.113	2.289	.136
	Within Groups	205.620	58	3.545		
	Total	213.733	59			
W10_First_WQ	Between Groups	.677	1	.677	.147	.703
	Within Groups	280.593	61	4.600		
	Total	281.270	62			

Table 8

ANOVA: Fall 2009 Section 1 and 2 Low Achievement Group Discussion Quiz Scores by Week

		Sum of Squares *SS*	*df*	Mean Square *MS*	*F*	Sig./Effect
W1_First_WQ	Between Groups	.059	1	.059	.011	.917
	Within Groups	319.875	59	5.422		
	Total	319.934	60			
W2_First_WQ	Between Groups	8.411	1	8.411	1.237	.271
	Within Groups	380.710	56	6.798		
	Total	389.121	57			
W3_First_WQ	Between Groups	1.071	1	1.071	.149	.701
	Within Groups	415.929	58	7.171		
	Total	417.000	59			
W4_First_WQ	Between Groups	10.789	1	10.789	2.595	.112
	Within Groups	249.485	60	4.158		
	Total	260.274	61			
W5_First_WQ	Between Groups	9.011	1	9.011	1.437	.235
	Within Groups	363.589	58	6.269		
	Total	372.600	59			
W6_First_WQ	Between Groups	.381	1	.381	.084	.772
	Within Groups	243.458	54	4.508		
	Total	243.839	55			
W7_First_WQ	Between Groups	.157	1	.157	.037	.848
	Within Groups	244.826	58	4.221		
	Total	244.983	59			
W8_First_WQ	Between Groups	16.346	1	16.346	2.570	.114
	Within Groups	362.501	57	6.360		
	Total	378.847	58			

W9_First_WQ	Between Groups	3.721	1	3.721	.520	.474
	Within Groups	408.075	57	7.159		
	Total	411.797	58			
W10_First_WQ	Between Groups	2.840	1	2.840	.671	.416
	Within Groups	237.229	56	4.236		
	Total	240.069	57			

Table 9

ANOVA: Fall 2009 Section 1 and 2 High Achievement Group Final Quiz Scores by Week

		Sum of Squares *SS*	*df*	Mean Square *MS*	*F*	Sig./Effect
W1_Final_WQ	Between Groups	.169	1	.169	.193	.662
	Within Groups	52.798	60	.880		
	Total	52.968	61			
W2_Final_WQ	Between Groups	.114	1	.114	.204	.653
	Within Groups	33.886	61	.556		
	Total	34.000	62			
W3_Final_WQ	Between Groups	1.497	1	1.497	1.874	.176
	Within Groups	47.923	60	.799		
	Total	49.419	61			
W4_Final_WQ	Between Groups	.009	1	.009	.010	.921
	Within Groups	51.928	61	.851		
	Total	51.937	62			
W5_Final_WQ	Between Groups	3.046	1	3.046	3.417	.069
	Within Groups	54.382	61	.892		
	Total	57.429	62			
W6_Final_WQ	Between Groups	.224	1	.224	.244	.623
	Within Groups	51.500	56	.920		
	Total	51.724	57			
W7_Final_WQ	Between Groups	.223	1	.223	.298	.587
	Within Groups	44.825	60	.747		
	Total	45.048	61			
W8_Final_WQ	Between Groups	.053	1	.053	.066	.798
	Within Groups	47.684	59	.808		
	Total	47.738	60			

W9_Final_WQ	Between Groups	.913	1	.913	.862	.357
	Within Groups	61.421	58	1.059		
	Total	62.333	59			
W10_Final_WQ	Between Groups	.086	1	.086	.098	.756
	Within Groups	53.660	61	.880		
	Total	53.746	62			

Table 10

ANOVA: Fall 2009 Section 1 and 2 Low Achievement Group Final Quiz Scores by Week

		Sum of Squares *SS*	*df*	Mean Square *MS*	*F*	Sig./Effect
W1_Final_WQ	Between Groups	.198	1	.198	.052	.821
	Within Groups	227.047	59	3.848		
	Total	227.246	60			
W2_Final_WQ	Between Groups	4.628	1	4.628	.969	.329
	Within Groups	267.441	56	4.776		
	Total	272.069	57			
W3_Final_WQ	Between Groups	4.430	1	4.430	1.176	.283
	Within Groups	218.554	58	3.768		
	Total	222.983	59			
W4_Final_WQ	Between Groups	3.492	1	3.492	.785	.379
	Within Groups	266.847	60	4.447		
	Total	270.339	61			
W5_Final_WQ	Between Groups	.030	1	.030	.008	.930
	Within Groups	219.304	58	3.781		
	Total	219.333	59			
W6_Final_WQ	Between Groups	2.751	1	2.751	1.033	.314
	Within Groups	143.802	54	2.663		
	Total	146.554	55			
W7_Final_WQ	Between Groups	6.696	1	6.696	1.732	.193
	Within Groups	224.304	58	3.867		
	Total	231.000	59			
W8_Final_WQ	Between Groups	16.239	1	16.239	2.980	.090
	Within Groups	310.608	57	5.449		
	Total	326.847	58			

W9_Final_WQ	Between Groups	3.017	1	3.017	.481	.491
	Within Groups	357.389	57	6.270		
	Total	360.407	58			
W10_Final_WQ	Between Groups	1.283	1	1.283	.343	.561
	Within Groups	209.631	56	3.743		
	Total	210.914	57			

Table 11

ANOVA: Fall 2009 Section 1 and 2 High Achievement Group Number of Quizzes Attempted by Week

		Sum of Squares *SS*	*df*	Mean Square *MS*	*F*	Sig./Effect
W1_Quiz Attempts	Between Groups	.020	1	.020	.012	.914
	Within Groups	99.464	60	1.658		
	Total	99.484	61			
W2_Quiz Attempts	Between Groups	1.280	1	1.280	.305	.583
	Within Groups	256.434	61	4.204		
	Total	257.714	62			
W3_Quiz Attempts	Between Groups	4.788	1	4.788	1.368	.247
	Within Groups	210.067	60	3.501		
	Total	214.855	61			
W4_Quiz Attempts	Between Groups	.422	1	.422	.095	.759
	Within Groups	271.006	61	4.443		
	Total	271.429	62			
W5_Quiz Attempts	Between Groups	1.618	1	1.618	.343	.560
	Within Groups	287.652	61	4.716		
	Total	289.270	62			
W6_Quiz Attempts	Between Groups	.008	1	.008	.002	.961
	Within Groups	190.837	56	3.408		
	Total	190.845	57			
W7_Quiz Attempts	Between Groups	.510	1	.510	.087	.769
	Within Groups	352.909	60	5.882		
	Total	353.419	61			
W8_Quiz Attempts	Between Groups	.000	1	.000	.000	.994
	Within Groups	167.934	59	2.846		
	Total	167.934	60			

W9_Quiz Attempts	Between Groups	4.676	1	4.676	1.489	.227
	Within Groups	182.174	58	3.141		
	Total	186.850	59			
W10_Quiz Attempts	Between Groups	.056	1	.056	.015	.904
	Within Groups	233.023	61	3.820		
	Total	233.079	62			

Table 12

ANOVA: Fall 2009 Section 1 and 2 Low Achievement Group Number of Quizzes Attempted by Week

		Sum of Squares *SS*	*df*	Mean Square *MS*	*F*	Sig./Effect
W1_Quiz Attempts	Between Groups	.840	1	.840	.604	.440
	Within Groups	82.013	59	1.390		
	Total	82.852	60			
W2_Quiz Attempts	Between Groups	1.034	1	1.034	.923	.341
	Within Groups	62.691	56	1.119		
	Total	63.724	57			
W3_Quiz Attempts	Between Groups	.107	1	.107	.068	.796
	Within Groups	92.076	58	1.588		
	Total	92.183	59			
W4_Quiz Attempts	Between Groups	1.059	1	1.059	.548	.462
	Within Groups	115.989	60	1.933		
	Total	117.048	61			
W5_Quiz Attempts	Between Groups	.086	1	.086	.039	.844
	Within Groups	126.897	58	2.188		
	Total	126.983	59			
W6_Quiz Attempts	Between Groups	.292	1	.292	.378	.541
	Within Groups	41.708	54	.772		
	Total	42.000	55			
W7_Quiz Attempts	Between Groups	.576	1	.576	.509	.478
	Within Groups	65.607	58	1.131		
	Total	66.183	59			
W8_Quiz Attempts	Between Groups	.176	1	.176	.209	.649
	Within Groups	47.960	57	.841		
	Total	48.136	58			

W9_Quiz Attempts	Between Groups	.026	1	.026	.016	.899
	Within Groups	90.889	57	1.595		
	Total	90.915	58			
W10_Quiz Attempts	Between Groups	.002	1	.002	.002	.964
	Within Groups	56.774	56	1.014		
	Total	56.776	57			

Table 13

ANOVA: Comparison of Final Quiz Scores Between Spring 2009 with Fall 2009 by Week

		Sum of Squares *SS*	*df*	Mean Square *MS*	*F*	Sig./Effect
1 Final WQ	Between Groups	11.563	1	11.563	2.869	.091
	Within Groups	3373.316	837	4.030		
	Total	3384.880	838			
2 Final WQ	Between Groups	19.262	1	19.262	4.242	.040 .005
	Within Groups	3791.765	835	4.541		
	Total	3811.027	836			
3 Final WQ	Between Groups	20.982	1	20.982	4.667	.031 .005
	Within Groups	3749.328	834	4.496		
	Total	3770.310	835			
4 Final WQ	Between Groups	12.159	1	12.159	2.540	.111
	Within Groups	3954.173	826	4.787		
	Total	3966.332	827			
5 Final WQ	Between Groups	16.528	1	16.528	3.744	.053
	Within Groups	3593.546	814	4.415		
	Total	3610.074	815			
6 Final WQ	Between Groups	1.644	1	1.644	.494	.482
	Within Groups	2543.442	764	3.329		
	Total	2545.086	765			
7 Final WQ	Between Groups	8.094	1	8.094	2.045	.153
	Within Groups	3217.654	813	3.958		
	Total	3225.747	814			
8 Final WQ	Between Groups	9.025	1	9.025	1.855	.174
	Within Groups	3864.064	794	4.867		
	Total	3873.089	795			

9 Final WQ	Between Groups	.003	1	.003	.001	.979
	Within Groups	3027.802	796	3.804		
	Total	3027.805	797			
10 Final WQ	Between Groups	.217	1	.217	.059	.808
	Within Groups	2872.144	783	3.668		
	Total	2872.362	784			

Table 14

ANOVA: Comparison of Final Quiz Scores Between Fall 2008 with Spring 2009 by Week

		Sum of Squares *SS*	*df*	Mean Square *MS*	*F*	Sig./Effect
1 Final WQ	Between Groups	.000	1	.000	.000	.997
	Within Groups	3323.225	752	4.419		
	Total	3323.225	753			
2 Final WQ	Between Groups	30.176	1	30.176	6.180	.013 .008
	Within Groups	3652.438	748	4.883		
	Total	3682.615	749			
3 Final WQ	Between Groups	31.487	1	31.487	6.550	.011 .008
	Within Groups	3518.932	732	4.807		
	Total	3550.420	733			
4 Final WQ	Between Groups	72.023	1	72.023	11.984	.001 .016
	Within Groups	4429.149	737	6.010		
	Total	4501.172	738			
5 Final WQ	Between Groups	94.029	1	94.029	17.521	.000 .023
	Within Groups	3965.930	739	5.367		
	Total	4059.960	740			
6 Final WQ	Between Groups	110.112	1	110.112	23.663	.000 .032
	Within Groups	3238.715	696	4.653		
	Total	3348.827	697			
7 Final WQ	Between Groups	376.985	1	376.985	78.353	.000 .096
	Within Groups	3526.748	733	4.811		
	Total	3903.733	734			
8 Final WQ	Between Groups	192.154	1	192.154	37.830	.000 .05
	Within Groups	3636.821	716	5.079		
	Total	3828.975	717			

9 Final WQ	Between Groups	457.676	1	457.676	94.014	.000 .118
	Within Groups	3461.255	711	4.868		
	Total	3918.931	712			
10 Final WQ	Between Groups	165.916	1	165.916	35.267	.000 .047
	Within Groups	3335.502	709	4.705		
	Total	3501.418	710			

Table 15

ANOVA: Comparison of Final Quiz Scores Between Fall 2008 with Fall 2009by Week

		Sum of Squares *SS*	*df*	Mean Square *MS*	*F*	Sig./Effect
1 Final WQ	Between Groups	15.175	1	15.175	3.317	.069
	Within Groups	4456.153	974	4.575		
	Total	4471.328	975			
2 Final WQ	Between Groups	2.153	1	2.153	.415	.519
	Within Groups	5028.707	970	5.184		
	Total	5030.860	971			
3 Final WQ	Between Groups	1.957	1	1.957	.382	.536
	Within Groups	4845.600	947	5.117		
	Total	4847.557	948			
4 Final WQ	Between Groups	35.131	1	35.131	5.829	.016 .006
	Within Groups	5653.235	938	6.027		
	Total	5688.366	939			
5 Final WQ	Between Groups	42.265	1	42.265	7.882	.005 .008
	Within Groups	5008.385	934	5.362		
	Total	5050.650	935			
6 Final WQ	Between Groups	115.608	1	115.608	25.241	.000 .027
	Within Groups	4144.989	905	4.580		
	Total	4260.598	906			
7 Final WQ	Between Groups	361.597	1	361.597	71.721	.000 .072
	Within Groups	4663.607	925	5.042		
	Total	5025.204	926			
8 Final WQ	Between Groups	154.032	1	154.032	28.860	.000 .031
	Within Groups	4819.581	903	5.337		
	Total	4973.613	904			

9 Final WQ	Between Groups	597.361	1	597.361	130.762	.000 .126
	Within Groups	4157.155	910	4.568		
	Total	4754.516	911			
10 Final WQ	Between Groups	196.927	1	196.927	44.527	.000 .048
	Within Groups	3914.044	885	4.423		
	Total	4110.972	886			

Table 16

Social Validity Results Pertaining to Perceived Effectiveness of AR techniques in a Classroom setting

Compared to the standard sessions without power points questions throughout the lecture, how helpful were the discussion sessions that included power point questions (that required your answering with clickers) in preparing you for the quizzes?			
Very	Moderately	Less	No Difference
59.01%	27.48%	3.83%	9.68%

Which type of discussion sessions did you prefer?	
Clicker	Non-clicker
82.9%	17.1%

How do you think the use of clickers during your discussion sessions affected your overall grade in this course?		
Increase	Decrease	No Difference
61.8%	13.0%	25.2%

How do you think the use of clickers during your discussion sessions affected your discussion quiz score in this course?		
Increase	Decrease	No Difference
64.2%	8.1%	27.7%

How do you think the use of clickers during your discussion sessions affected your weekly quiz score in this course?		
Increase	Decrease	No Difference
71.2%	18.7%	10.1%

Would you like to see the use of clickers in your future courses at UNR?	
Yes	No
77.2%	22.8%

What did the use of clickers in answering power point questions offer you? Select all that apply.				
Attention	Prepared	Learning	Interest	None
75.4%	57.6%	33.0%	13.7%	9.0%

List of Figures

Figure 1. Average discussion quiz score for Section1 and Section 2 across ten weeks.

Figure 2. Average first quiz score after discussion for Section 1 and Section 2 across ten weeks.

Figure 3. Average final quiz score for Section1 and Section 2 across ten weeks.

Figure 4. Average number of quiz opportunities utilized by Section 1 and Section 2 across ten weeks.

Figure 5. Average final score across both sections, grouped according to number of quiz opportunities utilized across ten weeks.

Figure 6. Average discussion quiz score for Section 1 and Section 2, separated by High and Low Achievement Groups, across ten weeks.

Figure 7. Average first quiz score after discussion for Section 1 and Section 2, separated by High and Low Achievement Groups, across ten weeks.

Figure 8. Average final score across both sections, separated by High and Low Achievement Groups, across ten weeks.

Figure 9. Average number of quiz opportunities utilized by Section 1 and Section 2, separated by High and Low Achievement Groups, across ten weeks.

Figure 10. Average discussion quiz score for Instructor 1 for Section 1 and Section 2 across ten weeks.

Figure 11. Average discussion quiz score for Instructor 2 for Section 1 and Section 2 across ten weeks.

Figure 12. Average first quiz score after discussion for Instructor 1 for Section 1 and Section 2 across ten weeks.

Figure 13. Average first quiz score after discussion for Instructor 2 for Section 1 and Section 2 across ten weeks.

Figure 14. Average final quiz score after discussion score for Instructor 1 for Section 1 and Section 2 across ten weeks.

Figure 15. Average final quiz score after discussion score for Instructor 2 for Section 1 and Section 2 across ten weeks.

Figure 16. Average final quiz score for Fall 2008, Spring 2009 and Fall 2009 semesters across ten weeks.

Figure 17. Comparison between the average Final Weekly Quiz Score between Fall 2008, Spring 2009 and Section 1 and Section 2 of Fall 2009.

Figure 18. Average number of retakes students in each section completed during the Spring 2009 semester.

Figure 1. Average discussion quiz score for Section1 and Section 2 across ten weeks.

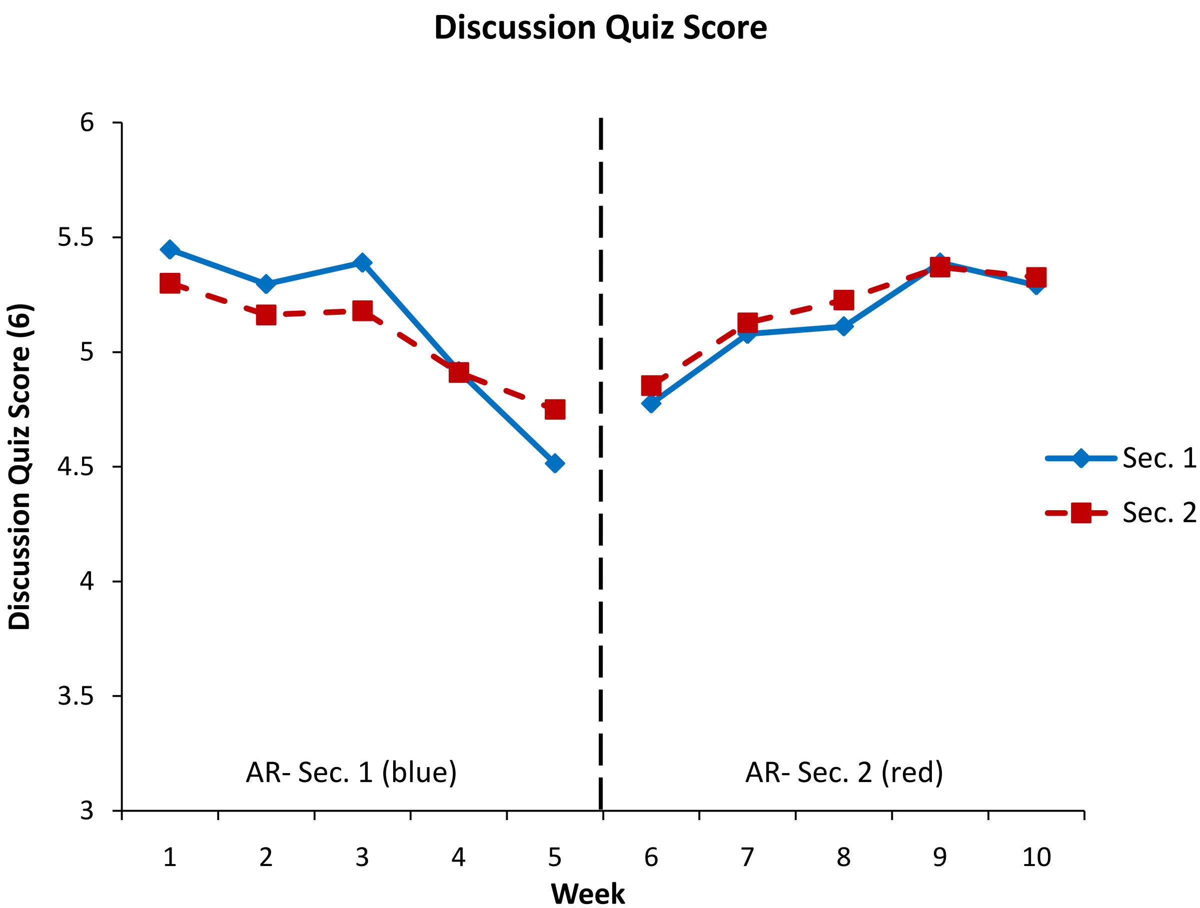


Figure 2. Average first quiz score after discussion for Section 1 and Section 2 across ten weeks.

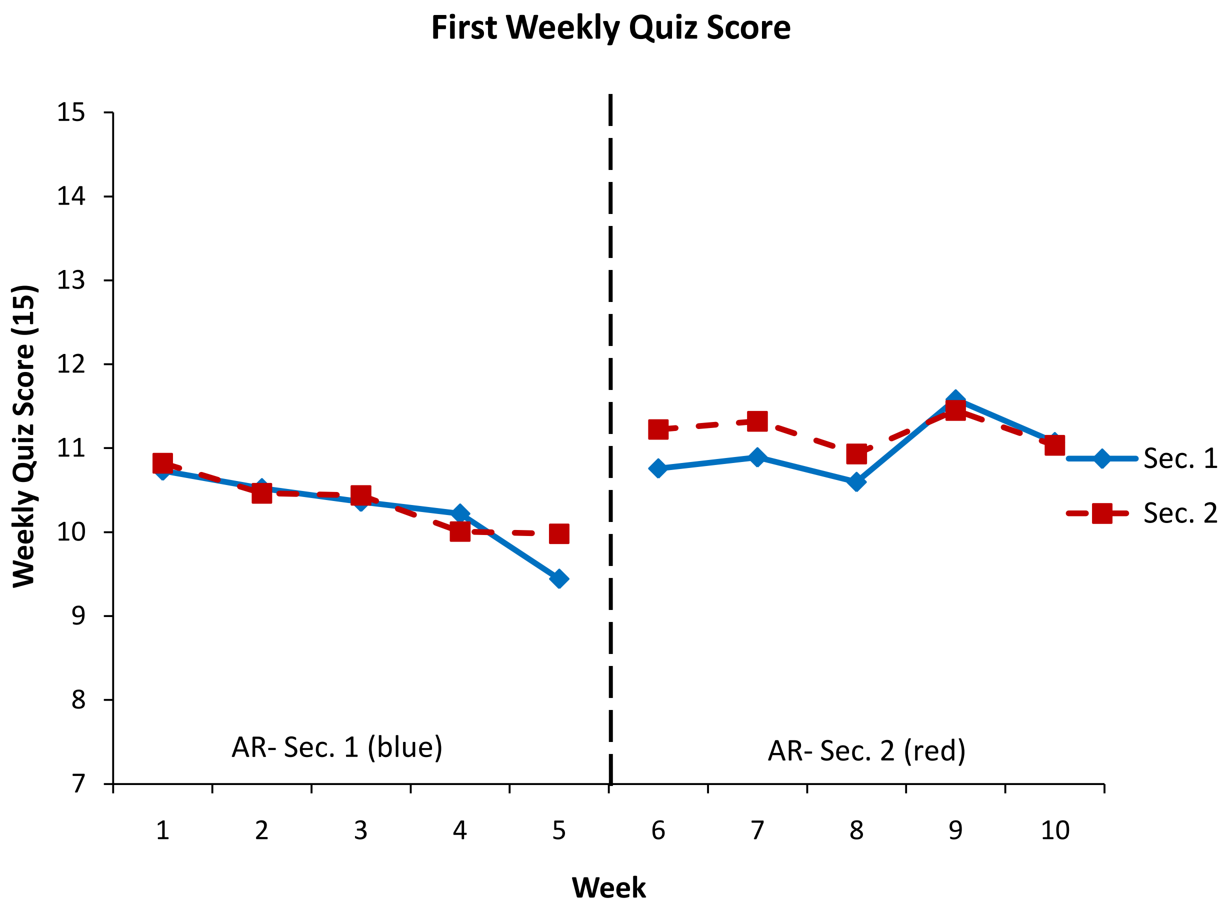


Figure 3. Average final quiz score for Section1 and Section 2 across ten weeks.

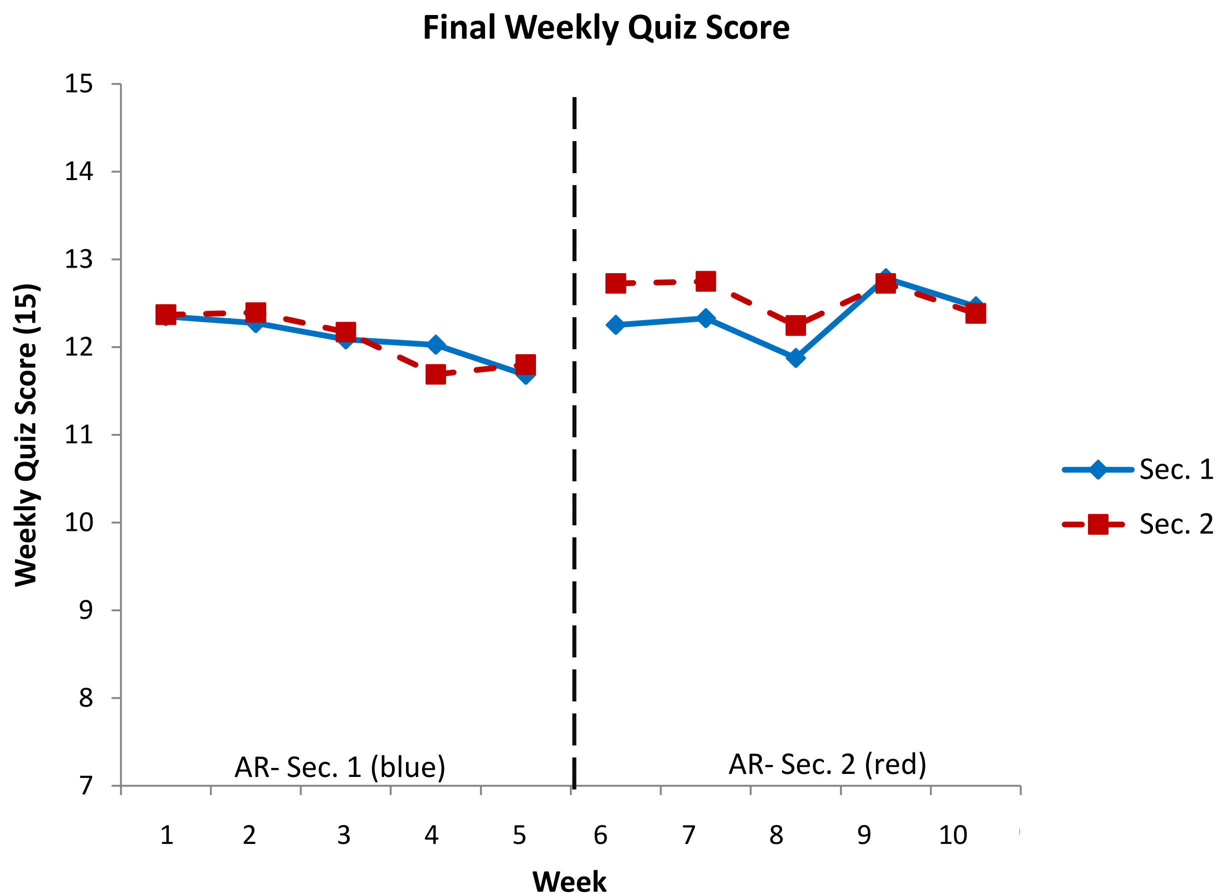


Figure 4. Average number of quiz opportunities utilized by Section 1 and Section 2 across ten weeks.

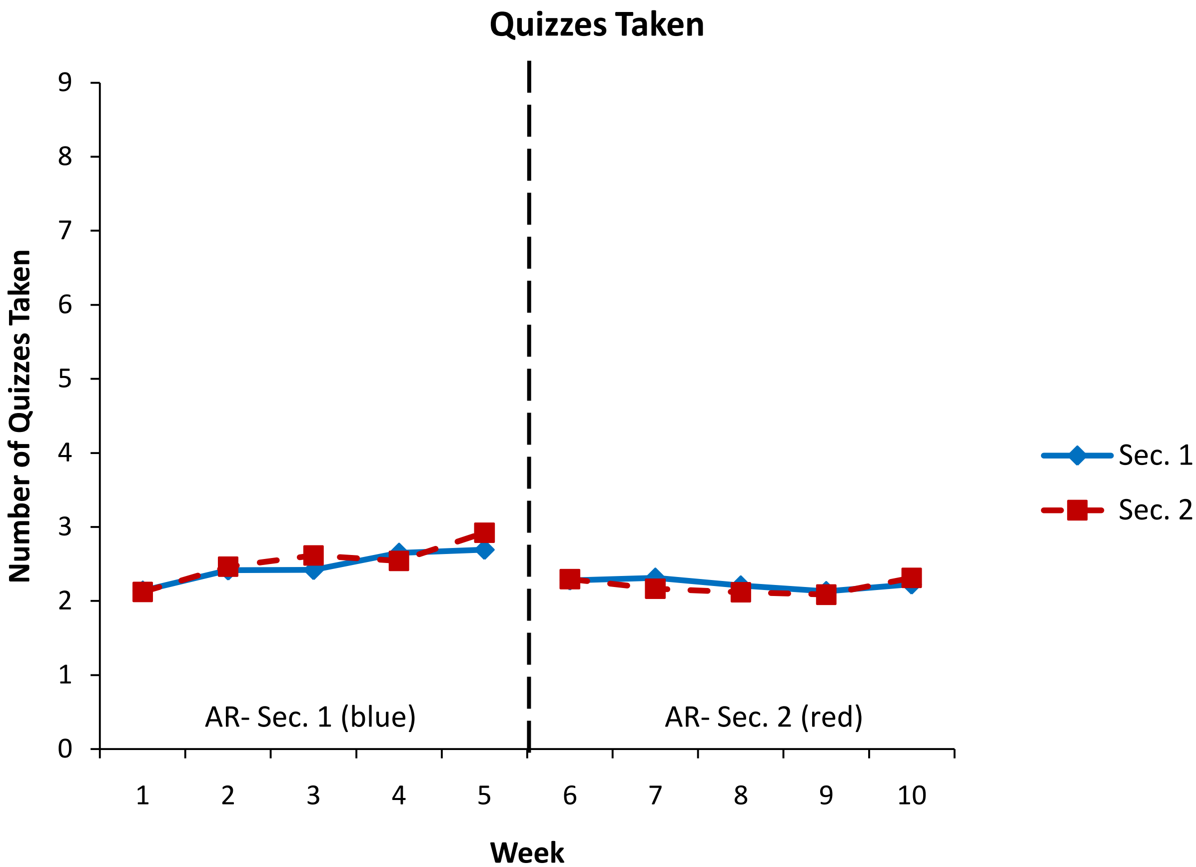


Figure 5. Average final score across both sections, grouped according to number of quiz opportunities utilized across ten weeks.

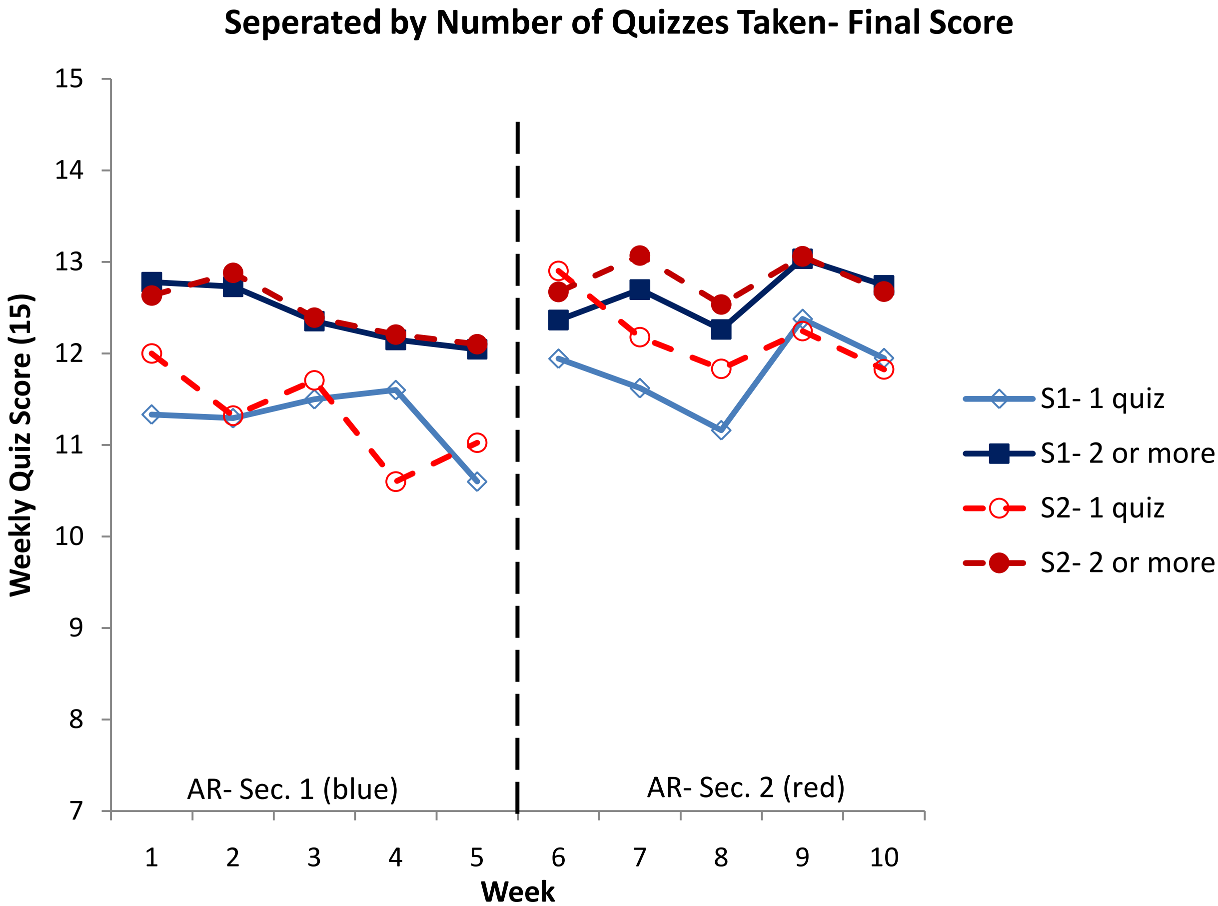


Figure 6. Average discussion quiz score for Section 1 and Section 2, separated by High and Low Achievement Groups, across ten weeks.

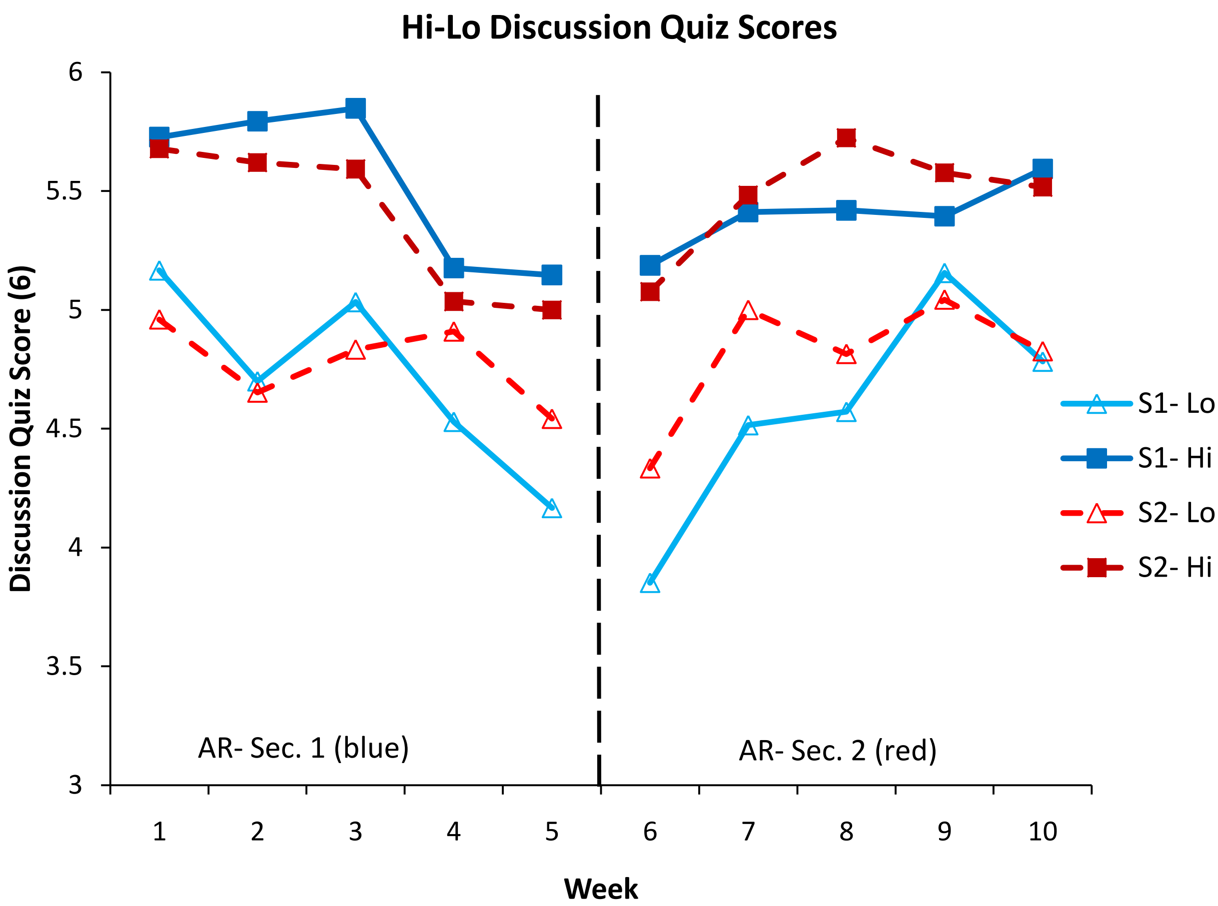


Figure 7. Average first quiz score after discussion for Section 1 and Section 2, separated by High and Low Achievement Groups, across ten weeks.

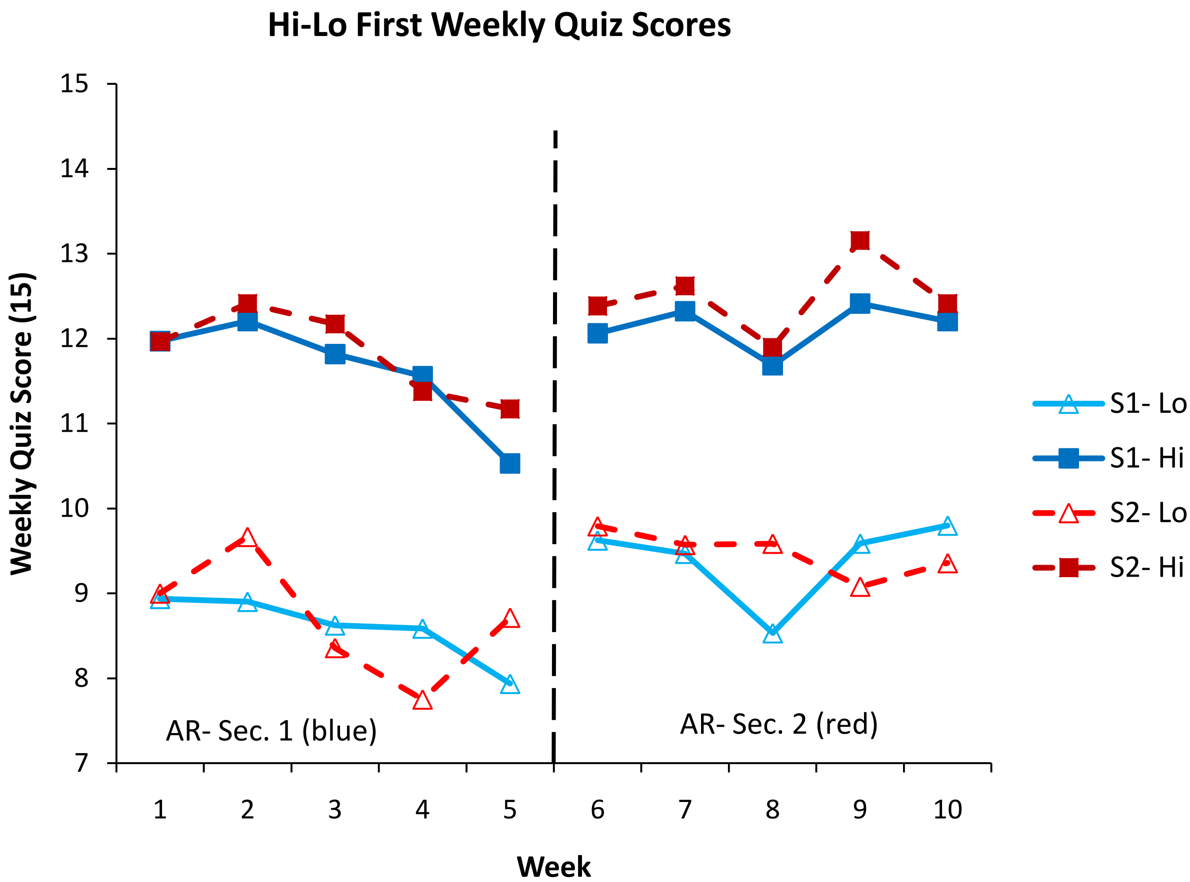


Figure 8. Average final score across both sections, separated by High and Low Achievement Groups, across ten weeks.

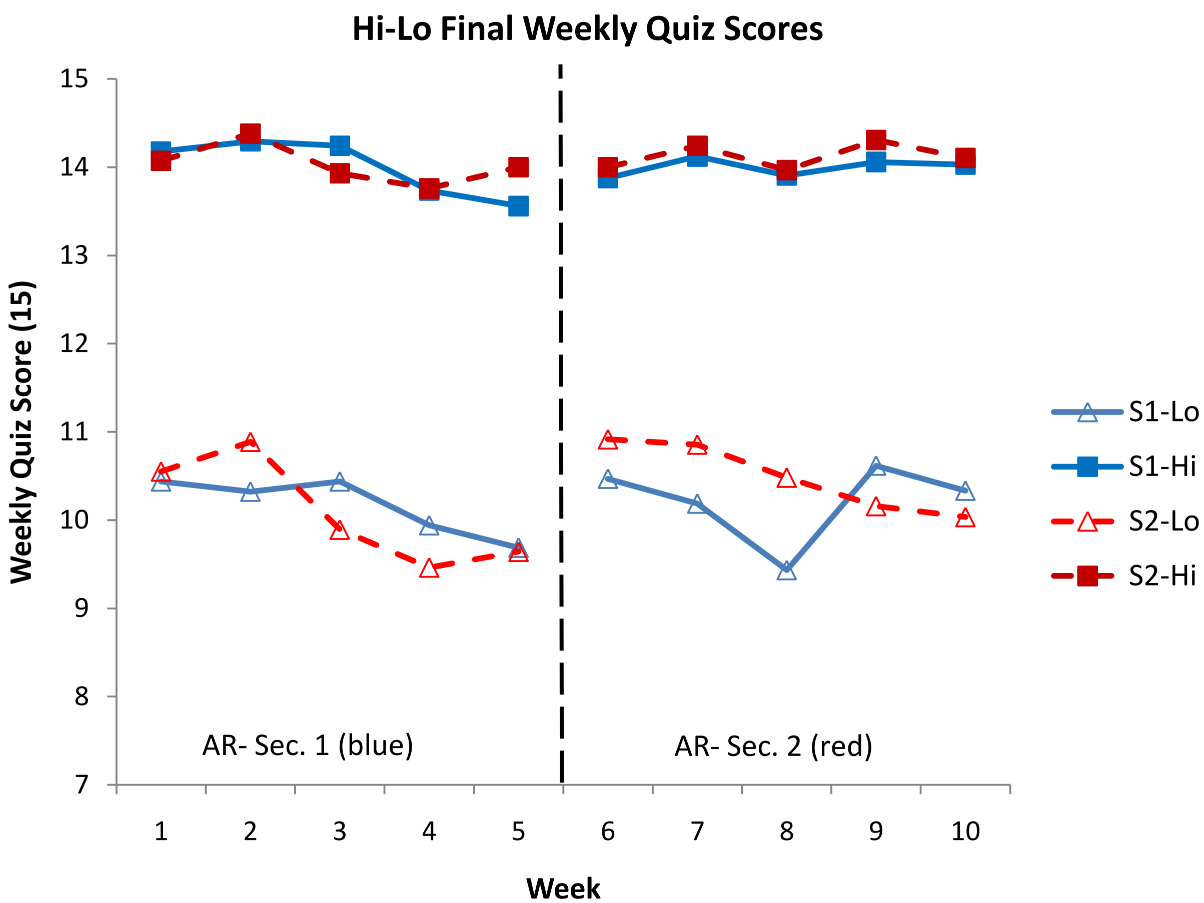


Figure 9. Average number of quiz opportunities utilized by Section 1 and Section 2, separated by High and Low Achievement Groups, across ten weeks.

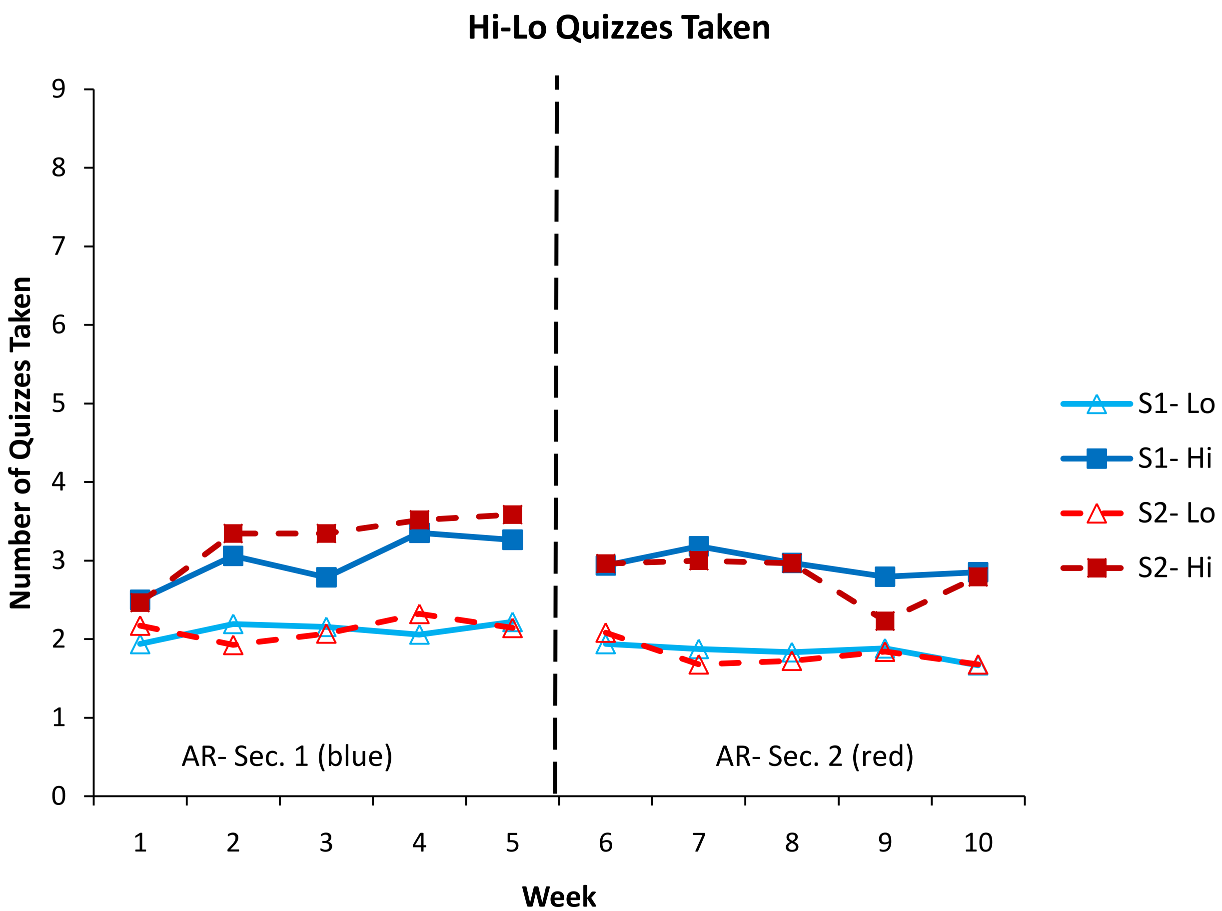


Figure 10. Average discussion quiz score for Instructor 1 for Section 1 and Section 2 across ten weeks.

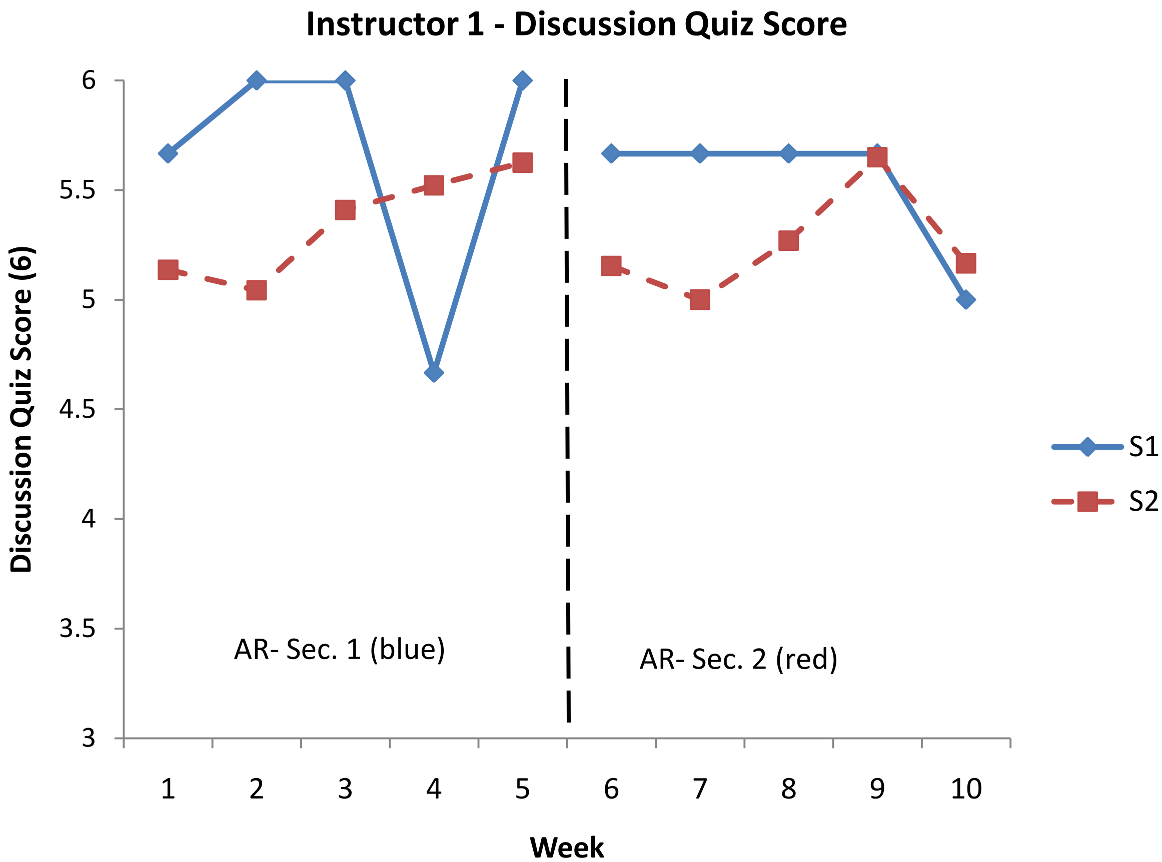


Figure 11. Average discussion quiz score for Instructor 2 for Section 1 and Section 2 across ten weeks.

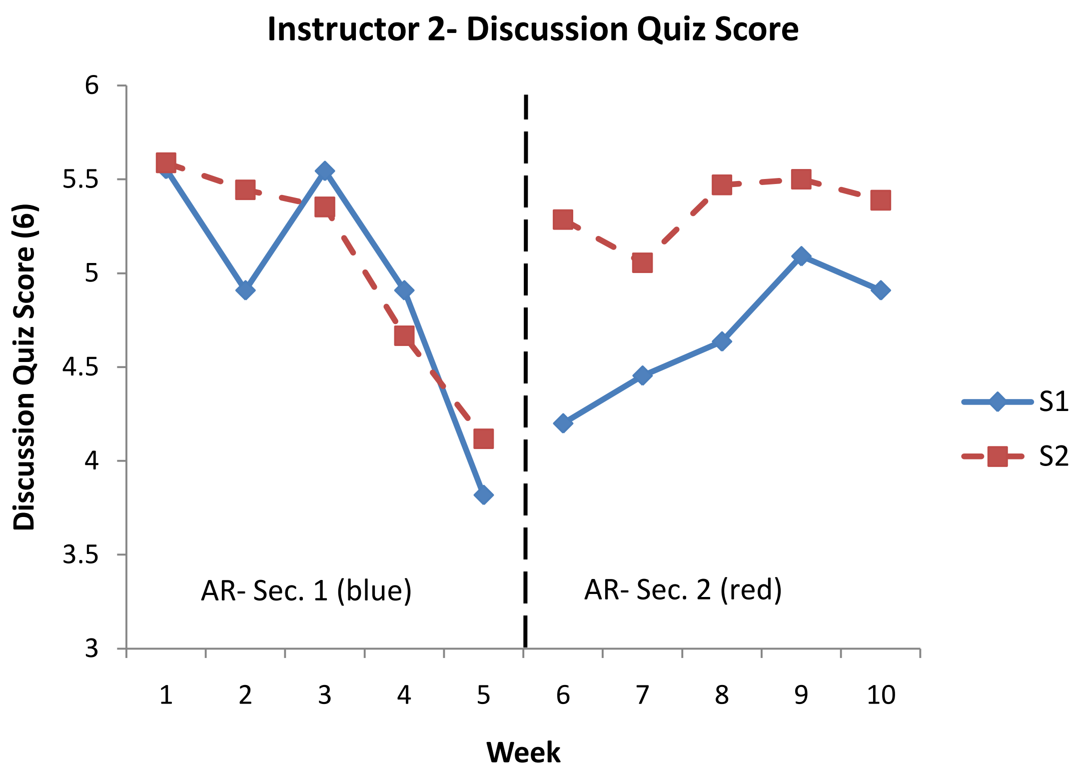


Figure 12. Average first quiz score after discussion for Instructor 1 for Section 1 and Section 2 across ten weeks.

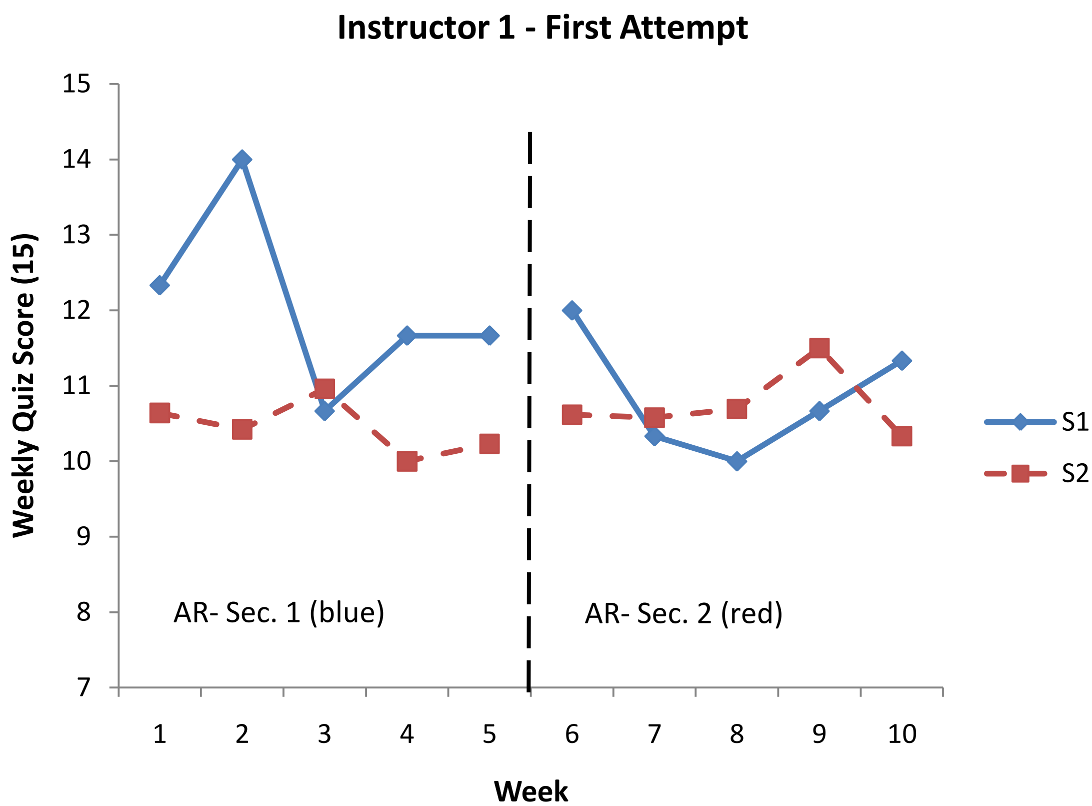


Figure 13. Average first quiz score after discussion for Instructor 2 for Section 1 and Section 2 across ten weeks.

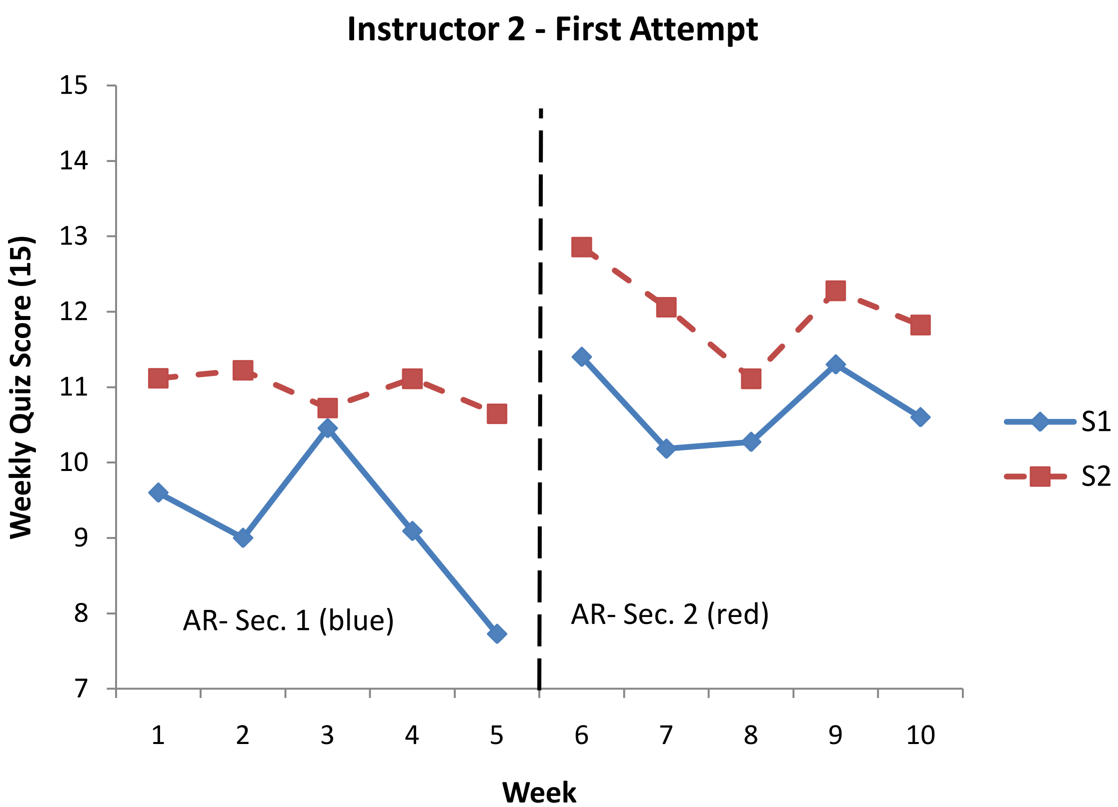


Figure 14. Average final quiz score after discussion score for Instructor 1 for Section 1 and Section 2 across ten weeks.

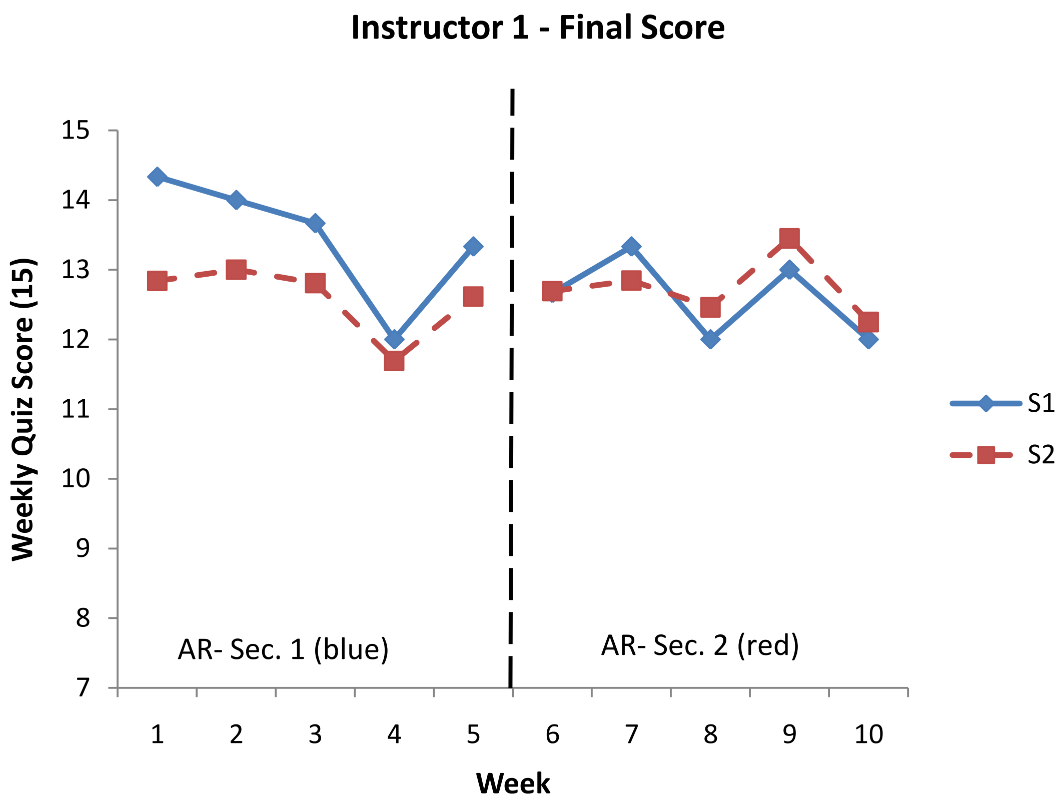


Figure 15. Average final quiz score after discussion score for Instructor 2 for Section 1 and Section 2 across ten weeks.

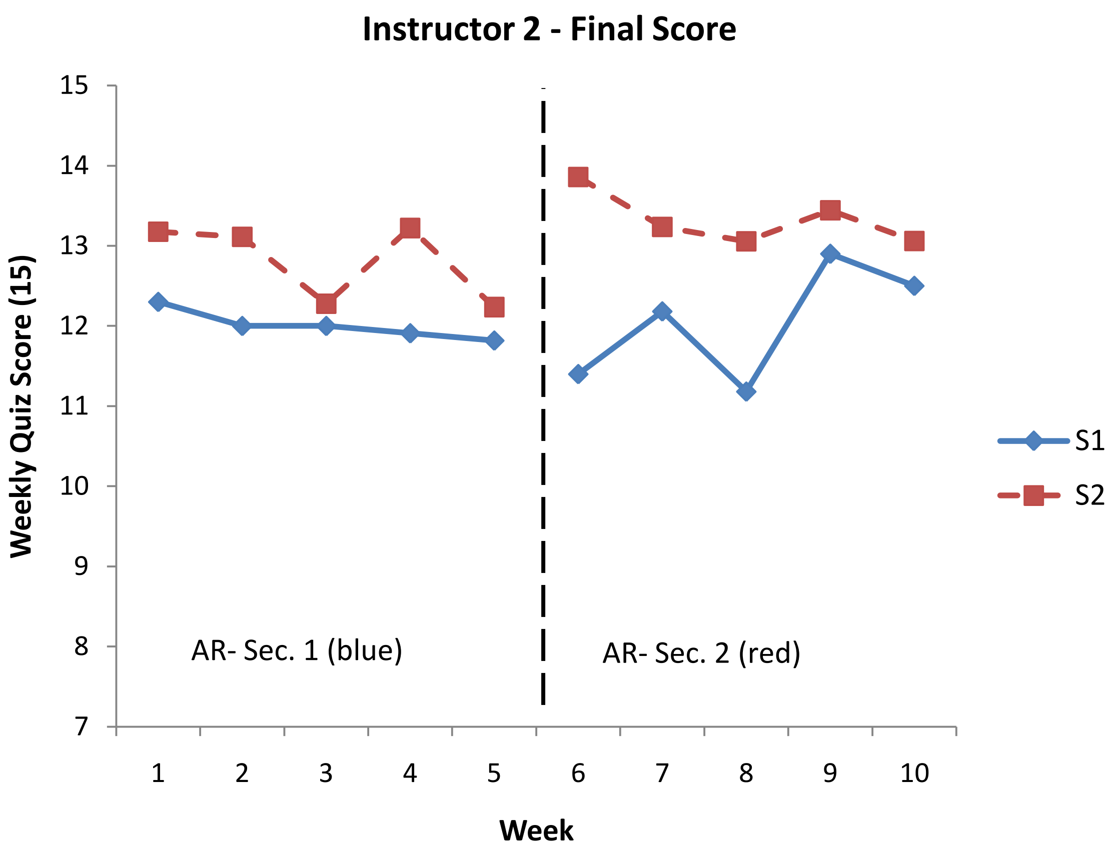


Figure 16. Average final quiz score for Fall 2008, Spring 2009 and Fall 2009 semesters across ten weeks.

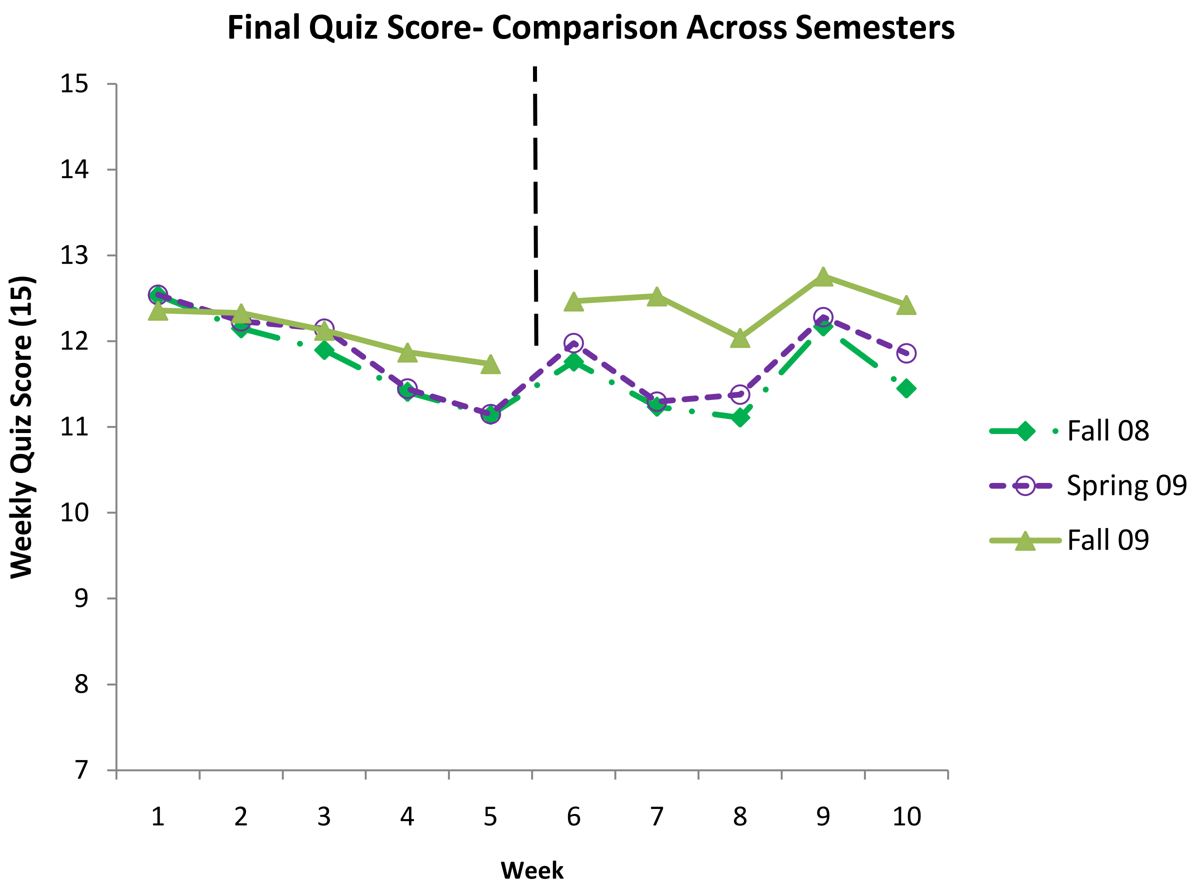


Figure 17. Comparison between the average Final Weekly Quiz Score between Fall 2008, Spring 2009 and Section 1 and Section 2 of Fall 2009.

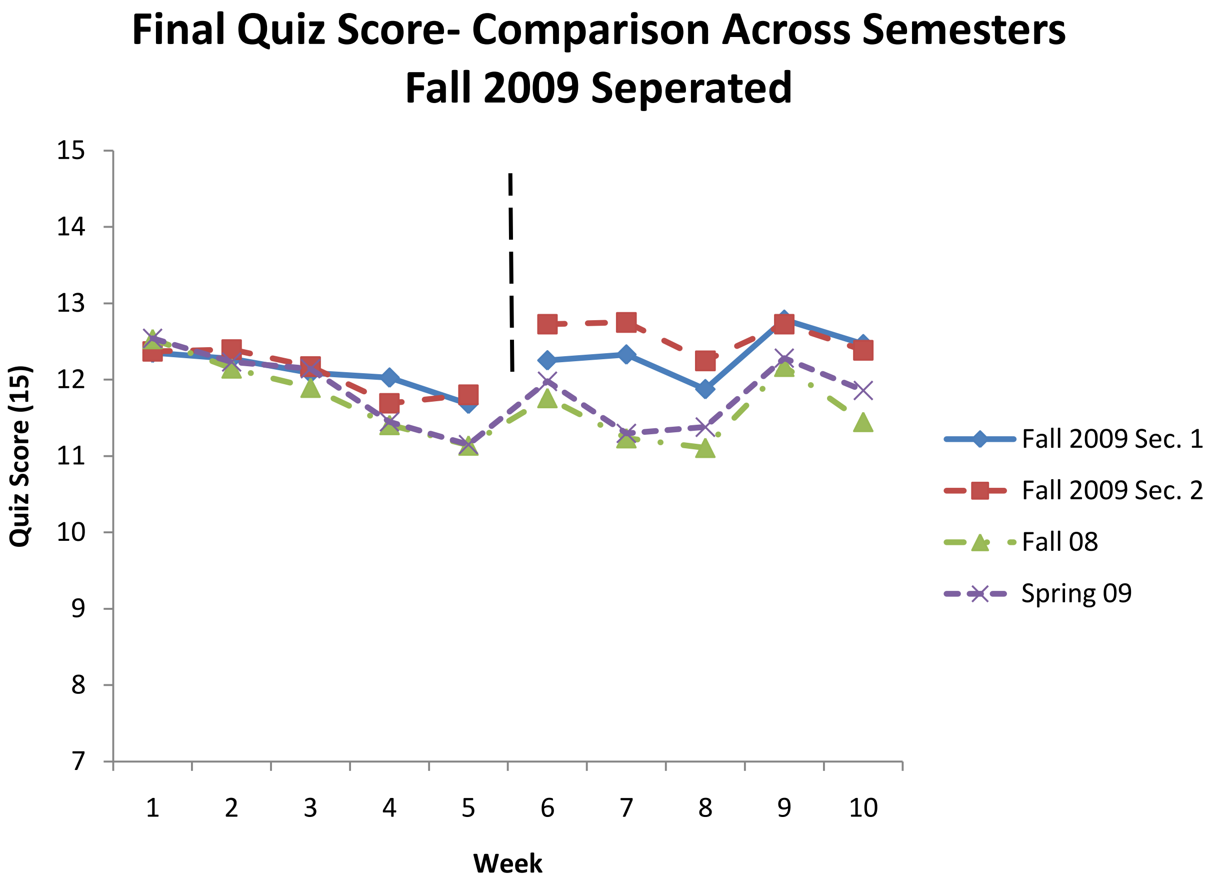
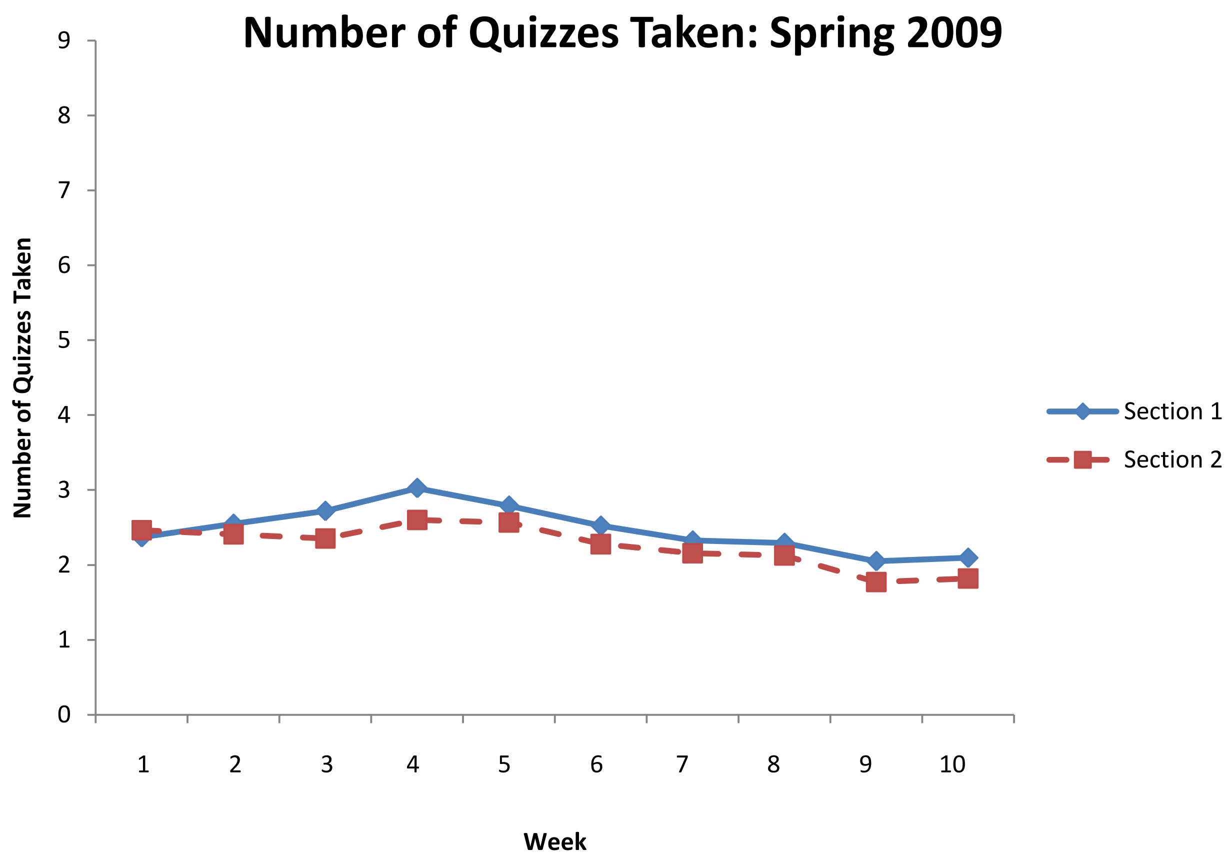


Figure 18. Average number of retakes students in each section completed during the Spring 2009 semester.



Appendix A

Protocol for Teaching Traditional Sessions

At the beginning of class you will tell the students that they will need to register their clickers for a discussion quiz that will be presented at the end of the discussion session. Ask the students to pull out their clickers, turn them on and enroll in PSY 101, using their clicker. As they register, count the number of students attending class. Once you have counted, wait until the counter on the right hand corner of the screen has reached the same number. If the number hasn't reached the number of students you counted within thirty (30) seconds, recount the number of students. If the number is still incorrect, ask the class if anyone has forgotten to bring their clicker. Count the number of students that do not have clickers and add that number to the number of clickers registered. If that number equals the number of students, then continue on with the lesson. If not, ask the class if everyone is signed in to the right class by asking "Does everybody's clicker say 'ANS:' at the top of your screen and 'PSY 101' at the bottom?" If a student (or students) is having difficulty signing on, assist them as necessary. If the number of students in the class is still greater than the number of clickers enrolled, remained them that they will only get points for the discussion quiz if their clicker is enrolled in the class. If no one speaks up, write down each student's name, the time and date of discussion and place in Daniel Reimer's box. Then continue with the lesson. Troubleshoot clicker problems as needed.

Once clicker set-up is complete (as described above) begin the Chapter presentation. Make sure that the presentation you are using is in the traditional presentation format. These will be labeled with the letter B following the chapter number (i.e. Chapter16B). Read the content on the slide, elaborating briefly on the content. You can paraphrase the exact content on the slide, but the basic content must remain the same. The elaboration can include a pertinent study that you are familiar with, and example or a personal anecdote, and should continue from between 10 seconds to about two minutes. You can also elaborate further on an experiment that is discussed in the presentation in more detail. The description of the experiment can last anywhere between ten seconds and three minutes.

After every section of content, a review slide will be presented. These review slides should function as a prompt for both you and the students. A review slide will consist of a heading titled "Questions" and the main points in the previously covered section, allowing you to ask students if they need any of the content discussed from the previous section clarified. If a question is brought up, you will present this question to the class and then wait for a response. If a student raises his or her hand, you will call on the student to answer the question. If the answer is correct, you will praise the student and briefly (in less than 30 seconds) explain why the answer was correct. If the answer was incorrect, you may ask other students to volunteer if they know the answer. If no students volunteer to answer within approximately 10 seconds, you will give the correct answer and briefly (in less than 30 seconds) explain why this is the correct answer. You will then move on to the next sub-topic in the chapter and continue presenting the material. If there aren't any questions, you will move on to the next sub-topic in the chapter.

A "late" student will be considered any student that has not been exposed to the question prompt slide. Once the first available question prompt slide is completed (i.e. the slide is no longer projected onto the screen) any students who comes in to discussion should be noted. A late student sheet is taped to the table at the front of the discussion. Check the "Late" box, write the first and last name of the student, the date, the time the discussion was scheduled to start and the student's section number and. Tell the student to register their clicker for the class.

A student that comes to a discussion, but does not have their clicker with them has two options. They can sit in on the discussion and miss out on their discussion quiz points, or they can leave and come to another discussion with their clicker. If they decide to leave, no action needs to be taken. If they decide to stay, check the "No Click" box, write the first and last name of the student, the date, the time the discussion was scheduled to start and the student's section number. Then begin discussion according to protocol.

After presenting the content for the discussion session, you will inform the class that they will by taking the discussion quiz. The quiz consists of six questions. Students will get one point for each correct answer and no points for an incorrect answer. The results of each question will be displayed after each question. If 70% or more of the class answers the question correctly, you will praise the class and briefly (in less than 30 seconds) explain why that answer was correct. If less than 70% or more of the class did not answer correctly, you will go back to the slide that is pertinent to the answer and explain (in less than 60 seconds) why a particular answer was correct.

As you teach several sessions, you will likely become familiar with the discussion quiz questions for each chapter. Therefore, you are asked not to drop hints during the presentation about which topics may show up on the discussion quiz at the end of the session. You may still emphasize any topic that you think is important for that chapter, but should not make comments that directly hint at a possible quiz question, such as "You might see this as a question later." Do not say anything that might clue students in as to what questions they can expect on the quiz.

If time management becomes an issue for you during your sessions, you may request that students bring their questions or comments to you after class so that you are able to cover all the material and get to the discussion quiz in a timely fashion.

Trouble Shooting:

Occasionally, a student's clicker will run out of power. This will most likely occur at the beginning of a discussion. If this occurs, take the student to the office. There is a screw-driver and spare batteries in the right hand drawer of the check in desk. Use the screwdriver to release the screw which will allow you to remove the battery cover. Replace the old battery with a new one, replace in the screw and make sure the clicker is working by turning it on. Follow the protocol to register the student to the class and continue with the protocol. The instructor can also ask a proctor to replace the battery.

If the student's clicker runs out of power after discussion has begun, follow the protocol to replace the battery (see above) and inform Daniel that the student's score may be split into two.

Appendix B

Protocol for Teaching Active Responding Sessions

At the beginning of class you will tell the students that they will be participating in a "clicker session" and that they will be seeing six review questions during the presentation. You will ask them to pull out their clickers, turn them on and enroll in PSY 101, using their clicker. As they register, count the number of students attending class. Once you have counted, wait until the counter on the right hand corner of the screen has reached the same number. If the number hasn't reached the number of students you counted within thirty (30) seconds, recount the number of students. If the number is still incorrect, ask the class if anyone has forgotten to bring their clicker. Count the number of students that do not have clickers and add that number to the number of clickers registered. If that number equals the number of students, then continue on with the lesson. If not, ask the class if everyone is signed in to the right class by asking "Does everybody's clicker say 'ANS:' at the top of your screen and 'PSY 101' at the bottom?" If a student (or students) is having difficulty signing on, assist them as necessary. If the number of students in the class is still greater than the number of clickers enrolled, remained them that they will only get points for the discussion quiz if their clicker is enrolled in the class. If no one speaks up, write down each student's name, the time and date of discussion and place in Daniel Reimer's box. Then continue with the lesson. Troubleshoot clicker problems as needed.

Once clicker set-up is complete (as described above) begin the Chapter presentation. Make sure that the presentation you are using is the presentation with the Active Responding slides. These will be in the AR Sessions folder and labeled with an "AR" after the chapter number (i.e. Ch 1 AR). Read the content on the slide, elaborating briefly on the content. You can paraphrase the exact content on the slide, but the basic content must remain the same. The elaboration can include a pertinent study that you are familiar with, and example or a personal anecdote, and should continue from between 10 seconds to about two minutes. You can also elaborate further on an experiment that is discussed in the presentation in more detail. The description of the experiment can last anywhere between ten seconds and three minutes.

After every section of content, an Active Responding Opportunity slide will come up. Read the question and the four possible answers aloud, and press the green “Play” button on the upper left corner of the screen. The timer will begin to count down from one minute. You will be able to see a running count of the number of votes on the upper right corner of the screen. If, after 30 seconds, all students have not voted, you will prompt students to vote by saying that everyone needs to vote. If some students have still not responded after 40 seconds (20 seconds left on the countdown clock) ask the class to make sure that their clicker’s all say “received” on the bottom line. If there are still students who haven’t responded, tell the entire class to press the “Enter” button. Once the countdown clock has reached “0:00” no more votes will be counted. If all students answer before the countdown clock reaches zero, then wait five seconds in case students want to change their answer and then hit the stop button.

If 70% or more of the class answers the question correctly, you will praise the class and briefly (in less than 30 seconds) explain why that answer was correct. If less than 70% or more of the class did not answer correctly, you will go back to the slide that is pertinent to the answer and explain (in less than 60 seconds) why a particular answer was correct. If a question is brought up, you will present this question to the class and then wait for a response. If a student raises his or her hand, you will call on the student to answer the question. If the answer is correct, you will praise the student and briefly (in less than 30 seconds) explain why the answer was correct. If the answer was incorrect, you may ask other students to volunteer if they know the answer. If no students volunteer to answer within approximately 10 seconds, you will give the correct answer and briefly (in less than 30 seconds) explain why this is the correct answer. You will then move on to the next sub-topic in the chapter and continue presenting the material. If there aren't any questions, you will move on to the next sub-topic in the chapter.

A "late" student will be considered any student who is unable to answer the first AR slide. Once the first available AR slide is completed (i.e. the instructor has pressed the stop button or the 60 second time has run out) any students who have not answered because they were not present should be noted. A late student sheet is taped to the table at the front of the discussion. Check the "Late" box, write the first and last name of the student, the date, the time the discussion was scheduled to start and the student's section number and. Tell the student to register their clicker for the class and answer the other available questions.

A student that comes to a discussion, but does not have their clicker with them has two options. They can sit in on the discussion and miss out on their discussion quiz points, or they can leave and come to another discussion with their clicker. If they decide to leave, no action needs to be taken. If they decide to stay, check the "No Click" box, write the first and last name of the student, the date, the time the discussion was scheduled to start and the student's section number. Then begin discussion according to protocol.

After presenting the content for the discussion session, you will inform the class that they will by taking the discussion quiz. The quiz consists of six questions. Students will get one point for each correct answer and no points for an incorrect answer. The results of each question will be displayed after each question. If 70% or more of the class answers the question correctly, you will praise the class and briefly (in less than 30 seconds) explain why that answer was correct. If less than 70% or more of the class did not answer correctly, you will explain (in less than 60 seconds) why a particular answer was correct.

As you teach several sessions, you will likely become familiar with the discussion quiz questions for each chapter. Therefore, you are asked not to drop hints during the presentation about which topics may show up on the discussion quiz at the end of the session. You may still emphasize any topic that you think is important for that chapter, but should not make comments that directly hint at a possible quiz question, such as "You might see this as a question later." Do not say anything that might clue students in as to what questions they can expect on the quiz.

If time management becomes an issue for you during your sessions, you may request that students bring their questions or comments to you after class so that you are able to cover all the material and get to the discussion quiz in a timely fashion.

Trouble Shooting:

Occasionally, a student's clicker will run out of power. This will most likely occur at the beginning of a discussion. If this occurs, take the student to the office. There is a screw-driver and spare batteries in the right hand drawer of the check in desk. Use the screwdriver to release the screw which will allow you to remove the battery cover. Replace the old battery with a new one, replace in the screw and make sure the clicker is working by turning it on. Follow the protocol to register the student to the class and continue with the protocol. The instructor can also ask a proctor to replace the battery.

If the student's clicker runs out of power after discussion has begun, follow the protocol to replace the battery (see above) and inform Daniel that the student's score may be split into two.

Appendix C

	Pre-test	Week 1	Week 2	Week 3	Week 4	Week 5	Mid-term	Week 6	Week 7	Week 8	Week 9	Week 10	Final
Section 1	x	AR	AR	AR	AR	AR	x						x
Section2	x						x	AR	AR	AR	AR	AR	x

Appendix D

Figure	Description
1	Discussion quiz scores, averaged at a group level. Results visually depict the differences of the discussion scores between Section 1 and Section 2.
2	First weekly quiz scores, averaged at a group level. Results visually depict the differences of the first weekly chapter quiz scores after discussion between Section 1 and Section 2.
3	Final weekly quiz scores, averaged at a group level. Results visually depict the differences of the final weekly chapter quiz scores, the highest quiz score taken during the week, between Section 1 and Section 2.
4	The number of quiz opportunities students took, averaged at a group level.
5	The average final score of students across sections when grouped according to the number of quizzes taken; 1 or more than 1
6	Discussion quiz scores, averaged at the separated achievement level. Results visually depict the differences of the discussion scores between achievement levels across Section 1 and Section 2
7	First weekly quiz scores, averaged at the separated achievement level. Results visually depict the differences of the first weekly chapter quiz scores after discussion between achievement levels across Section 1 and Section 2
8	Final weekly quiz scores, averaged at the separated achievement level. Results visually depict the differences of the final weekly chapter quiz scores, the highest quiz score taken during the week, between achievement levels across Section 1 and Section 2
9	The number of quiz opportunities students took, averaged across sections and between achievement levels

10 & 11	Comparison across instructors between Section 1 and 2 for discussion quiz scores
12 & 13	Comparison across instructors between Section 1 and 2 for first weekly quiz scores
14 & 15	Comparison across instructors between Section 1 and 2 for final weekly quiz scores
16	A historical analysis comparing the current experimental data with data collected from two semesters previous, Fall 08 and Spring 09.
17	The comparison seen in Figure 16, but includes a breakdown by section for Fall 2009, the target semester
18	Average number of retakes each section took during the Spring 2009 semester