November 2018

Effects of Interspersing Recall versus Recognition Questions with Response Cards During Lectures on Students' Academic and Participation Behaviors in a College Classroom

Leslie S. Singer
University of South Florida, singer@mail.usf.edu

Follow this and additional works at: https://scholarcommons.usf.edu/etd

Part of the Behavioral Disciplines and Activities Commons

Scholar Commons Citation

This Dissertation is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.
Effects of Interspersing Recall versus Recognition Questions with Response Cards During Lectures on Students’ Academic and Participation Behaviors in a College Classroom

by

Leslie S. Singer

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Applied Behavior Analysis
Department of Child and Family Studies
College of Behavioral and Community Sciences
University of South Florida

Major Professor: Kimberly A. Crosland, Ph.D., BCBA-D
Raymond G. Miltenberger, Ph.D., BCBA-D
Rose Iovannone, Ph.D., BCBA-D
Kevin Murdock, Ph.D., BCBA-D

Date of Approval:
October 16, 2018

Keywords: multiple-choice questions, short answer questions, active instruction, opportunities to respond, effective instruction

Copyright © 2018, Leslie S. Singer
Dedication

There have been numerous people to support my decision to begin and finish this dream to complete my graduate school experience as a Doctor of Philosophy in Applied Behavior Analysis. First, thank you to Victoria Fogel who shaped my skills as a young behavior analyst and inspired me to get involved in a career that has been so meaningful to me. To Dr. Timothy Weil, thank you for pushing me to do more than I ever though I could. When life served lemons, you taught me to make lemonade. To my mom, Ellen Tyree, thank you for answering the many phone calls that came pouring in and supporting me to move forward despite the obstacle. To Diana Yeager, thank you for mentoring me throughout my schooling experience. I am so grateful to call you my friend, Diana. To my friends, thank you for helping me learn to take a day off (you know who you are). To Elizabeth Ward, thank you for inspiring me to go back to college by just simply saying, “you got this.” To Dr. Kevin Murdock, I never really knew how much you taught me until I left your supervision. I hope to be half the behavior analyst you are one day. To Dr. Miltenberger, thank you for always wearing a smile and helping me to figure out tough obstacles with elegance. I will never hear Neil Young again without thinking of you. Lastly, thank you to Dr. Crosland for your guidance and support during this endeavor. I will forever be grateful to you for all you have taught me about research and life. I have enjoyed sharing many laughs with you. Socrates once said, “I cannot teach anybody anything, I can only make them think.” I hope to inspire future analysts and students to simply think, for themselves.
# Table of Contents

List of Tables ............................................................................................................................. iii

List of Figures ............................................................................................................................. iv

Abstract ......................................................................................................................................... v

Chapter One: Introduction ........................................................................................................... 1

Chapter Two: Methods .................................................................................................................. 33
  Participants .................................................................................................................................. 33
  Setting ......................................................................................................................................... 34
  Phase 1 ......................................................................................................................................... 34
  Reliability and validity of questions ........................................................................................... 35
  Phase 2 ......................................................................................................................................... 36
  Materials ..................................................................................................................................... 36
  Dependent variables and data collection .................................................................................... 36
    Academic behavior ..................................................................................................................... 37
    Percentage of student responding ............................................................................................ 38
  Inter-observer agreement ............................................................................................................ 38
  Experimental design ...................................................................................................................... 39
  Procedures ................................................................................................................................. 40
    Instructor training ..................................................................................................................... 40
    Baseline .................................................................................................................................... 41
    Response cards intervention ....................................................................................................... 42
      Response cards with recall ....................................................................................................... 43
      Response cards with recognition ............................................................................................. 43
  Treatment Integrity ..................................................................................................................... 43
  Social Validity ............................................................................................................................. 44

Chapter Three: Results .................................................................................................................. 46
  Phase 1 ......................................................................................................................................... 46
  Phase 2 ......................................................................................................................................... 46
  Instructor Social Validity ........................................................................................................... 49
  Student Social Validity ............................................................................................................... 50

Chapter Four: Discussion .............................................................................................................. 61

References ....................................................................................................................................... 72

Appendices ..................................................................................................................................... 83
Appendix A: Experienced Instructor Raters Questionnaire........................................84
Appendix B: Percentage of Student Response Data Recording Sheet .......................85
Appendix C: Treatment Integrity Data Sheet for Baseline Condition..........................86
Appendix D: Treatment Integrity Data Sheet for Response Card Conditions................87
Appendix E: Instructor Social Validity Questionnaire
Appendix F: Student Social Validity Questionnaire..................................................89
Appendix G: Instructor Delivered Questions During Each Class ...............................90
Appendix H: IRB Exempt Letter ................................................................................91
**List of Tables**

<table>
<thead>
<tr>
<th>Table 1:</th>
<th>Average Results of Experienced Instructor Rater Questionnaires</th>
<th>52</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2:</td>
<td>Results of All Students’ Average Immediate Quiz Scores across Conditions</td>
<td>53</td>
</tr>
<tr>
<td>Table 3:</td>
<td>Results of All Students’ Average Delayed Quiz Scores across Conditions</td>
<td>53</td>
</tr>
<tr>
<td>Table 4:</td>
<td>Statistical Analysis of Conditions</td>
<td>54</td>
</tr>
<tr>
<td>Table 5:</td>
<td>Instructor Social Validity Results</td>
<td>55</td>
</tr>
<tr>
<td>Table 6:</td>
<td>Student Social Validity Results</td>
<td>55</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1: Immediate Average Quiz Scores ................................................................. 56
Figure 2: Delayed Average Quiz Scores ..................................................................... 56
Figure 3: Average Individual Student Immediate Quiz Scores .................................. 57
Figure 4: Average Individual Student Delayed Quiz Scores ..................................... 57
Figure 5: Lower 20% Immediate Quiz Scores ............................................................ 58
Figure 6: Lower 20% Delayed Quiz Scores ............................................................... 58
Figure 7: Individual Student Immediate Quiz Scores ............................................... 59
Figure 8: Individual Student Delayed Quiz Scores ................................................... 59
Figure 9: Average Student Responses ...................................................................... 60
Abstract

Instructional design and delivery may be one tool available to teachers to increase the academic and social behaviors of all students in the classroom. *Effective instruction* is an evidence-based teaching strategy that can be used to efficiently educate our youth across all learning environments. One effective instructional strategy includes increasing students’ *opportunities to respond* to instructor-posed questions during lectures. Students may respond to questions using a response card system as a way to promote active engagement. This study examined the most common form of instructor-posed questions presented during lecture, recall and recognition questions, to determine the differential effects on students’ academic and participation behavior in a college classroom. Results found no differentiation in students’ academic behavior with respect to question type. Students’ participation behavior was greater when the instructor used class wide active responding procedures than observed in baseline conditions that represented typical college instruction.
Chapter One:

Introduction

Classrooms today are highly diverse and include students from a variety of backgrounds, cultures, abilities, learning strengths and weaknesses. Evidence of this diversity can be found in the 39th annual report to congress on the implementation of the Individuals with Disabilities Education Act (IDEA, 2004). IDEA is a law in the United States that guarantees students with a disability will receive free and appropriate public education. The legislation is divided into four parts (A-D), and part B ensures that students 3-21 years old receive services needed to facilitate their education (IDEA, 2004). The law states that all students, despite disability, will receive the same opportunity for education across the nation.

According to the report, a total of 6,050,725 students age 6-21 years old were served under IDEA part B during the 2015-2016 academic school year (U.S. Department of Education, OSEP, 2017). This number represents 8.9% of the resident population ages 6-21 years old. Despite the diversity of students, federal mandates require inclusion of all children in the classroom. Ninety-five percent of students served under IDEA part B were educated in regular classrooms for at least some portion of the school day, with 62.7% spending 80% or more of their day in regular classrooms (U.S. Department of Education, OSEP, 2017). These numbers represent almost a 10% increase in students served under the law across the last twelve years.

Since 2012, there has been an increasing trend in the number of students served under IDEA. Students served are divided into several categories, illustrating the types of diversity that exists in classrooms today. Of those served, specific learning disabilities was the largest
category of students served (38.8%), followed by speech and language impairment (17.3%),
other health impairments (15%), autism (9.1%), intellectual disabilities (6.9%), emotional
disturbance (5.7%), and other disabilities combined (7.2%) (U.S. Department of Education,
OSEP, 2017). Between 2006 and 2015, the percent of resident population ages 12-17 years old
(189%) and 18-21 years old (209%) reported under the category of autism increased drastically
(U.S. Department of Education, OSEP, 2017). Classrooms may contain students diagnosed with,
or at-risk for, learning disorders, emotional behavioral disorders (EBD), developmental
disorders, and/or intellectual disorders. Students with, or at-risk for, EBD for example, tend to
be removed from regular classrooms and placed in self-contained classrooms yielding higher
rates of problem behaviors and lower academic expectations (Scott, Alter, & Hirn, 2011;
Williams Bost & Ricconomi, 2006). While these restrictive placements may be appealing due to
the supposed structure (e.g., low student-teacher ratio, increased number of paraprofessionals,
environmental modifications to optimize space), student outcomes reveal they are less
academically capable than their general education peers and continue to emit more disruptive
behaviors (Maggin, Wehby, Moore Partin, Robertson, & Oliver, 2011). Another requirement of
IDEA is that while including diverse students in the classroom, teachers must also meet
accountability measures demonstrating that students are meeting their grade-level norms.

At the national level, student learning is measured using high stakes tests. The No Child
Left Behind Act (NCLB) requires states to develop standardized assessments in an effort to
improve education outcomes for all students (NCLB, 2008). Federal funding is provided to
schools based on students’ performance on these tests, and standards of achievement are
determined by each state. Schools are required to monitor achievement of students using
indicators of adequate yearly progress, measures of academics, and rates of students dropping
out of school in addition to rates of graduation (NCLB, 2008). Each year, schools are expected to improve on these measures in an effort to receive their federal funding for the school.

Less than half of the students served under IDEA part B participated in regular assessment based on grade-level academic achievement standards with accommodations in reading (grades 3-8: 38.4%; high school: 46.7%) and math (grades 3-8: 38.5%; high school: 48.4%; U.S. Department of Education, OSEP, 2017). Students who were assessed without accommodations occurred at even lower rates in reading (grades 3-8: 39.5%; high school: 49.3%) and math (grades 3-8: 37.5%; high school: 47.4%; U.S. Department of Education, OSEP, 2017). While the accommodations may facilitate standardized testing, some students are not graduating from high school. Of those who do graduate high school, some will continue into higher education. Considering these students is important given the fact that we are educating our youth to become productive members of society in hopes to increase all students’ quality of life.

Of the students (14 – 21 years old) who exited IDEA part B, a total of 69.9 % of students earned their regular high school degree and 18% dropped out of school (U.S. Department of Education, NCES, 2017). Once students graduate high school, some will continue their education into postsecondary classrooms. According to the National Center for Education Statistics (NCES), the 2015-2016 academic school year enrolled 20,224,069 students in postsecondary institutions, 1,075,496 in Florida (U.S. Department of Education, NCES, 2017). At the national level, 59.8% of students enrolled in universities receive their four-year degree within six years (U.S. Department of Education, NCES, 2017). Retention rate is the percent of full-time, first-time bachelor’s degree-seeking undergraduates from the previous fall who are again enrolled in the current fall semester. Retention rates for the full-time bachelor’s degree
seeking college students were 81% nationally (U.S. Department of Education, NCES, 2017). Given the diversity of the classrooms, it is important for teachers across all education environments (i.e., primary, secondary, and post secondary) to ensure learning is occurring for all students.

Classrooms across primary, secondary, and post secondary environments will likely include students with disabilities. According to the NCES (2017), 11.1% of undergraduate students in 2011-12 academic year were reported as having a disability. These numbers are similar to the 10.9% of undergraduate students reported in 2007-08 academic year. Eleven percent of undergraduate students across the three academic years represent both male and female students that self-reported either a specific learning disability, a visual impairment, hard of hearing, deafness, a speech impairment, an orthopedic impairment, or a health impairment (U.S. Department of Education, NCES, 2017).

Highlighting the diversity of the classroom is important because it is likely that most teachers will encounter students who disrupt the learning process. At the national level, approximately two students per classroom experience social or emotional difficulties that disrupt daily functioning of class (Thompson, 2011). Teachers in these environments are challenged to engage students in efforts to increase social and academic skills (Scott et al., 2011; Thompson, 2011). One way teachers can engage all students is to have an effective and efficient classroom management plan. When teachers use effective classroom management strategies, research has shown increases in academics, decreases in student problem behaviors as well as increases in social skills (Carnine, 1976; Lambert, Cartledge, Heward, & Lo, 2006; Newcomer, 2009).

A comprehensive classroom management plan includes effective environmental, behavioral, and instructional strategies (Newcomer, 2009; Simonsen, Fairbanks, Briesch, Myers
and Sugai, 2008). Simonsen and colleagues (2008) conducted an extensive literature review to determine components of effective classroom management. Results of their review found “five empirically supported, critical features of effective classroom management: a) maximize structure; b) post, teach, review, monitor, reinforce expectations; c) actively engage students in observable ways; d) use continuum of strategies for responding to appropriate behaviors; and e) use continuum of strategies to respond to inappropriate behaviors” (p. 353). When a teacher considers environmental strategies, they are observing the ways in which students interact with each other and the teacher within the classroom environment.

Effective environmental strategies may include maximizing structure in the physical arrangement of the classroom environment to ensure students have a place to store their items, considering the organization of the classroom to allow for all students to easily walk around the room, in addition to ensuring all students seated at their desks have sight of the instructor during lessons (Gilkey Hirn & Park, 2012; Simonsen et al., 2008). Other environmental strategies may include close proximity of the teacher to students who have difficulties in particular subject areas, so the teacher is able to monitor closely student performance and is available if the student needs assistance. Close proximity is the teacher’s engagement with students, for example the placement of students in relation to each other and the teacher, a teacher moving around the room, providing direct instruction and feedback (Conroy, Sutherland, Snyder, & Marsh, 2008). These environmental strategies, like close proximity, also allow the teacher to monitor student behavior in an effort to help students stay engaged in the learning process rather than disrupting the learning process.

Behavior management strategies are defined as arranging antecedents to prevent challenges and consequences to differentially reinforce expected classroom behaviors.
Behavior management strategies may include classroom rules posted, taught and regularly reviewed so students have learned what behaviors are expected in the classroom (Barbeta, Leong Norona, & Bicard, 2005; Newcomer, 2009). Posting rules for behavior, and teaching these rules, may prevent challenges of noncompliance and off-task behaviors (Simonsen et al., 2008). Once classroom rules are established, the teacher must carry out consistent implementation of earned and/or lost consequences. Students could self-monitor progress towards academic and behavioral goals, which may facilitate the teacher’s use of consistent implementation of consequences. If challenging behaviors do arise, then preplanning consequences is necessary.

Examples of preplanned consequences would include loss of tokens for inappropriate behavior, or ignoring a student who calls out for attention and provide attention for raising their hand. Student misbehaviors are learned and most often occur as a function to access something (e.g., tangible item, attention) or to escape something (e.g., difficult work, non-preferred activity; Barbeta et al., 2005). The job of the teacher is to determine the function of inappropriate behaviors so they are able to teach students more appropriate ways to solicit the help or attention (e.g., ask for a break, ask for help, ask if doing good work). Classroom management is the teachers’ implementation of environmental and behavioral strategies that maximize the learning process. The last strategy in a comprehensive classroom management plan is the pedagogy.

Effective instructional strategies have been described as one method to prevent students’ problem behaviors and are likely the best management strategy to increase students’ academic skills (Barbeta et al., 2005; Conroy et al., 2008). By manipulating antecedents (e.g., increasing opportunities for students to respond during lectures) and consequences (e.g., increasing contingent praise for correct responses), teachers are promoting a positive climate in their
classroom (Conroy et al., 2008). Classwide interventions also include a group of evidence-based teaching strategies to increase appropriate and decrease inappropriate academic and social behaviors (Newcomer, 2009; Simonsen et al., 2008). These strategies usually include a combination of praise, opportunities to respond and rules.

While students in classrooms are highly diverse, teacher-delivered instruction is one variable that most classrooms share. Using evidence-based teaching methods may be one effective strategy teachers can use to help all students achieve success (Carnine, 1976; Newcomer, 2009). Teachers are responsible for fostering positive climates while motivating students’ academic success. Teachers who use effective instructional strategies as part of their management plan are likely able to achieve this positive climate in the classroom. Effective instruction is in contrast with what traditional instruction usually involves (Faust & Paulson, 1998; Michael, 1991).

In primary education, teachers use daily lecture format with a variety of independent, small group and whole class assignments to measure student learning (Kretlow, Cooke, & Wood, 2012). In secondary education, teachers use a combination of daily lecture plus class discussion, group and/or individual assignments, homework, attendance, and participation in addition to quiz and exam grades as a measure of students’ learning (Michael, 1991). Traditional instruction at the college level typically involves an instructor-delivered lecture, some include visual displays of the material being presented (e.g., PowerPoint), and students listening while some may take notes (Faust & Paulson, 1998; Michael, 1991; Newman Thomas, Blood Pinter, Carlisle, Goran, 2015). Few instructors actively engage students by asking questions during lecture. Asking questions is a form of feedback in this traditional classroom that may be acquired by asking students to volunteer to share their answer. This method may be problematic as an individual
student response makes it difficult for the instructor to gather a precise sense of whole-class understanding, especially in a larger classroom (Narayan, Heward, Gardner, Courson, & Omness, 1990; Newman Thomas et al., 2015).

When using individual response methods, the students that do respond are typically the students who are confident in their response being accurate, resulting in the same one or two students responding to a teacher’s question (Helf, 2015). The teacher could also pose a question to the class and receive zero responses from students (Helf, 2015). When students do not know an answer to a teacher-posed question, it could serve as an abolishing operation for students to respond. Research surrounding evidence-based instruction promotes active engagement in the classroom as a way to increase academic and social behavior.

Effective instruction involves a teachers’ use of appealing and structured teaching strategies as an antecedent for students’ academic responding (e.g., participation, note-taking). Components of delivering effective instruction include a teacher increasing students’ opportunities to respond (Kritch & Bostow, 1998; MacSuga-Gage & Gage, 2015), increasing effective teacher praise and increasing the amount of time a teacher is actively engaging students’ responding (Barbetta et al., 2005; Maggin et al., 2011; Stichter et al., 2009). Effective instruction has been suggested to help prevent students from dropping out of school (Williams Bost & Riccomini, 2006), in addition to enhancing academic skills (Haydon et al., 2010).

When effective teaching practices are not used, it could result in unmotivated students who dropout of school all together (Williams Bost & Riccomini, 2006). The literature surrounding effective instruction offers a wide variety of recommendations. Williams Bost and Riccomini (2006) suggest 10 evidence-based principles of effective instruction to improve
student engagement, especially for students with disabilities, in effort to prevent students from dropping out of school:

1. Active engagement (i.e., time on task);
2. Providing the experience of success (e.g., task assignment matches level of achievement);
3. Content coverage and opportunity to learn (e.g., increased opportunities to learn increases achievement);
4. Grouping for instruction (e.g., whole group, small group, 1:1; whole offers shared learning experience);
5. Scaffold instruction (e.g., individualized instruction and/or support);
6. Addressing forms of knowledge (e.g., declarative [basic facts and vocabulary], procedural [steps used to solve problems], conditional knowledge [when and where to use certain strategies]);
7. Organizing and activating knowledge (e.g., relate new content to old content);
8. Teaching strategically (e.g., teach students how to learn);
9. Make instruction explicit (e.g., goals/objectives, structure lesson in obvious formats, present clearly and directly);
10. Teaching sameness (e.g., same within subjects and across subjects; recognize patterns and organize). (p. 305-307).

While these principles were developed from the literature surrounding special education and students with disabilities, students without disabilities might also be more likely improve their engagement if a teacher were to consider using these principles during instruction.
Instructional design and delivery may be one tool available to teachers to increase academic and social behaviors of all students in the classroom. When students begin to relate their school experiences to successful positive outcomes, they may be more likely to perform better academically and socially and may be less likely to drop out of school (Williams Bost & Riccomini, 2006). The principles provided are a set of tools for teachers to use to help improve student outcomes in the learning process. Some teachers are losing up to four hours per week managing classroom disruptive behaviors (Thompson, 2011). This reactionary approach to classroom management is detrimental to the climate of the classroom and subsequent learning achievement of all students. As one study (Oliver & Reschly, 2010) noted, teachers are trained mostly in reactive classroom management approaches rather than preventative approaches.

Oliver and Reschly (2010) reviewed 135 syllabi from 26 university special education teacher-training programs and found that programs supported more reactive procedures than preventative approaches. Ninety-six percent of the 26 university programs focused on behavior reduction strategies with little to no emphasis on classroom preventative approaches. This study provides insight into one state’s teacher training programs. Results of their study found only 7 of the 26 universities offered a class on classroom management skills, highlighting the need for better teacher preparation of classroom management skills. If teachers are providing effective environmental, behavioral, and instructional strategies then increases in academic and social skills will likely result. One evidence-based teaching practice involves the teachers’ use of opportunities to respond as a way to increase students’ active engagement (Whitney, Cooper, & Lingo, 2015). If teachers are asking students questions during lectures, this will increase students’ active responding and subsequently decrease students’ passive attending to lecture.
An opportunity to respond (OTR) is defined as a teacher’s presentation of a question (e.g., verbally, textually) to a class as an antecedent for students’ response (Schnorr, Freeman-Green, & Test, 2016; Stichter et al., 2009). When a teacher increases students’ OTR to academically posed questions, a result of increased student engagement and correct responding has been found (Conroy et al., 2008; Haydon et al., 2010). Simply increasing the OTR has a built in component of the teacher providing more feedback, error correction, contingent praise and progress monitoring, all components of effective instruction. “Increasing instructional pacing through OTR is a questioning, prompting, or cueing technique that begins a learning trial” (Conroy et al., 2008, p. 26).

According to Conroy and colleagues (2008) OTR includes five main components:

1. Increasing rates of teacher instructional talk that includes repeated verbal, visual, or both prompts for responding;
2. Present information in a manner that increases student correct responding (e.g., “This is the letter A, what letter is this?”; cue + prompt);
3. Implement individualized instructional modifications appropriate for students level of functioning and checks for understanding and accuracy;
4. Repeated instructional prompting that incorporates wait time to allow for student response;

When a teacher increases the rate of OTR, research has shown increases in academics and social behaviors (Christle & Schuster, 2003; Khan, Miltenberger, & Singer, 2015; Lambert et al., 2006; Singer, Crosland, & Fogel, 2013). For example, Singer and colleagues (2013) examined the effects of using response cards to answer 16 teacher-delivered questions on the social behaviors
of four elementary students. Results found a decrease in disruptive behaviors for all students during the response card OTR interventions.

According to Whitney and colleagues (2015), increasing OTR will likely increase student-teacher interactions. Their study randomly sampled 900 15 min direct observations of teachers’ rate of OTR across 18 elementary, 9 middle, and 7 high schools and found that as students’ age and their grade level increased, the teachers’ rates of OTR decreased. In other words, as students age, their teachers ask less and less questions. Results show teachers in elementary asked a question every 1.4 min, middle school teachers asked one question every 1.5 min and high school teachers asked one question every 2 min during reading and math instruction. The rates obtained from these 900 samples are well below the recommended norms of 4-6 responses per min for new material being presented and 8-12 OTR for review as suggested by the Council for Exceptional Children (1987).

Another suggestion by Scott et al. (2011) is for teachers to provide a minimum of three OTR per min to students. Sainato, Strain, and Lyon (1987) examined teacher-led instruction to a group of preschool children who were disabled and found as the rate of OTR increased, rates of correct responding also increased. Greenwood and colleagues (1984) were among the first to define the concept of OTR as active responding during engaged time, noting the importance of high rates of OTR affecting achievement. The literature surrounding rates of OTR also note the importance of pacing during the lecture.

Carnine (1976) evaluated the effects of a fast-paced and slow-paced lecture on the off-task behaviors, participation, and correct responding behaviors of two first grade students and found reductions in off-task behaviors for both students during small group instruction. Correct responding and participation were most frequent for one student in all three fast-paced conditions.
and most frequent in final fast-paced conditions for the second student. Fast-paced delivery was defined as zero delay between commands during reading instruction, whereas a 5 s delay between student response and delivery of the next command occurred for slow-paced instruction. Praise was held constant (90 s fixed-interval schedule) to prevent increasing amounts of praise from confounding the results. These results suggest pacing of instruction will likely control behavior in the classroom, especially academic and social behaviors.

In a university intro to psychology classroom, Grobe, Pettibone, and Martin (1973) audio recorded lectures in a study of pacing, and pace was differentiated by +/- 8 syllable counts per min (i.e., the instructor spoke slow [X=102 syllables per min], moderate [X=134 syllables per min], or fast [X=145 syllables per min]). Results found less classroom noise was attributed to moderately paced lecture as compared to a fast-paced lecture and slow-paced lecture.

Providing students with OTR is effective if they are attending to the material. Lignugaris/Kraft and Rousseau (1982) noted that academic-engaged time requires students’ attending to OTR in an effort to learn the skill or concept being taught. They note the importance of the within-trial interval (i.e., question, pause for student responses, pause before feedback), and suggest longer durations of stimuli presentations result in more accurate responses. The first pause should be long enough for students to conceptualize or “figure out” the response whereas the second pause should be short and provide immediate corrective feedback (i.e., immediacy of reinforcement to increase skill acquisition has been shown more effective than delay of feedback). They also suggest the intertrial interval (ITI) is another important component of instructional pacing that should be shorter in duration to facilitate increases in on-task behaviors. The moving on to the next concept/fact is called the ITI. Faster rates of ITI have resulted in the reduction of challenging behavior and higher accuracy and
participation (Lignugaris/Kraft & Rousseau, 1982). Task difficulty and student characteristics make it difficult to find a parameter of ITI and within-trial interval norms, but these variables are important components of effective instruction and warrant further investigation.

Despite the recommendations for evidence-based teaching strategies, some teachers are not implementing effective instruction strategies. In the study by Whitney et al. (2015), 19% of observations had zero OTR. These findings suggest teachers in both primary, secondary, and college classrooms (Newman Thomas et al., 2015) are using a lecture mostly format, requiring students to passively attend to material being presented. Some authors point out that teacher training on managing behavior in the classroom focuses mostly on reactive approaches, rather than preventative approaches (Oliver & Reschly, 2010). These findings suggest that teachers are not trained to recognize effective instruction as a way to prevent challenging classroom behavior. This reactionary approach dominates our society, including our systems of education.

When students are prompted to actively respond to a teacher’s question, increases in academics (e.g., quiz grades, exam scores) have been found for special education students (Barbetta, Heron, & Heward, 1993; Sterling, Barbetta, Heward, & Heron, 1997), general education students (Drevno et al., 1994), and higher education students (Kellum, Carr, & Dozier, 2001). Several teaching strategies used to increase student responding include: peer tutoring, self-monitoring, choral responding, guided notes, and classroom response systems (Heward, 1997; Newcomer, 2009).

Peer tutoring was developed over 40 years ago as a way to increase active student engagement in the primary and secondary classrooms and as a cost-effective way to decrease drop outs and increase retention in college settings (Topping, 1996). Peer tutoring usually involves a teacher dividing students into pairs, assigning one person to the role of the student and
one person to the role of the tutor. Sometimes this role is fixed and sometimes the role is reciprocal. Students then practice what they learned in class while receiving feedback from the tutor. This class-wide peer tutoring program has been evaluated across subjects like spelling, reading, and math in both primary and secondary education and has found increases in academic and social behaviors of students who are culturally diverse, students with disabilities and students without disabilities (Greenwood, 1997; Topping 1996).

Peer tutoring programs that exist at the college level pairs early undergraduate (e.g., freshmen, sophomore) students with a more advance undergraduate peer tutor (e.g., junior, senior) based on year of study, and some offer tutors in the same year based on grade point average (Topping, 1996). Whether peer tutoring exists in a small group or individual format, the tutor will offer structured ways to advance through the curriculum based on training received on their role as a tutor. The role of the tutor, whether in schools or universities, is typically to check the learners’ comprehension, test the learners’ skills while recording progress towards a learning outcome (Topping, 1996). Students could also self-monitor as a way to actively respond during the learning trial.

When students employ self-monitoring (or self-assessment), they are recording and monitoring their own progress towards learning a particular topic (Brieseh & Chafouleas, 2009; Crabtree, Alber-Morgan, Konrad, 2010; Nikou & Economides, 2016) or towards reducing off-task behaviors (Dalton, Martella, & Marchand-Martella, 1999; Dean, Malott, & Fulton, 1983). Some research has shown students prefer computer-based assessment (including mobile assessment), rather than paper and pencil assessment mediums, and found increases in student learning and motivation as a result of self-assessment (Nikou & Economides, 2016). Research
has shown increases in low- and medium-achieving students’ grades as a result of using electronic devices to self-monitor progress (Dean et al., 1983; Nikou & Economides, 2016).

While self-monitoring progress towards a learning goal is important, if teachers are not monitoring students’ assessments then students could be practicing errors or monitoring inaccurately. This has been an ongoing concern in the literature surrounding self-monitoring in classrooms, as Briesch and Chafouleas (2009) point out the need for external (i.e., teacher) contingencies to make the most of the intervention. Also, much of the research on self-monitoring has focused on low-achieving students, and mixed results are found with high achieving students (Crabtree et al., 2010; Dean et al., 1983; Nikou & Economides, 2016). These results suggest that this technology may not be effective for all students in a classroom, and/or effective across many educational environments.

There are two types of classroom response systems: those that involve limited technology are known as low technology systems and those that incorporate high technology usually include both hardware and software (Caldwell, 2007).

One effective way to increase active student responding in the classroom is through the use of low technology classroom response systems during instructional periods (Barbetta et al., 1993). These systems include posing questions to the class and having students respond by either raising their hand, chorally, or using response cards (Barbetta et al., 1993). These systems are low technology due to their ability to be implemented without hardware or software requirements.

Traditional lecture may involve a teacher presenting a question to the class about material just covered, or an assigned reading, followed by a prompt for students to raise their hand to
answer the teacher’s question. When teachers present a question to the class using individual responding, usually one student is called on to respond to the question.

While this method is superior to lecture only classrooms, typically the students who are confident in their responses will raise their hands (Helf, 2015). Reluctant responders may be too shy to answer aloud in class in fear they will have the wrong answer, or may not know the answer at all and just not raise their hand. The downfall of this approach is that all students in the class passively attend to the prompt while the one student called on to answer the question is provided with reinforcement in the form praise for a correct answer.

In contrast to hand raising methods, choral responding requires all students to respond to the teachers question in unison either by raising their hands simultaneously or verbally calling out the answer in unison (Haydon, Marsicano, & Scott, 2013). This response method is more effective than the individual responding that usually follows a hand raising method where the teacher selects only one student to respond (Haydon & Hunter, 2011; Haydon et al., 2013). While students are more active in their responses, the teacher is challenged to differentiate between simultaneous voices to determine which student may have provided an incorrect response. If teachers are requiring choral responding using hand raises, they are more likely to determine which students are not accurate in their responses. If teachers are asking students to vote on a correct response to a multiple-choice question by presenting cues one-at-a-time, then they will have even better chances of seeing how the group responded to a particular question.

Benefits to this type of ASR method are that all students are actively engaged while responding to the prompt. Many question formats are available to the instructor when presenting the prompt which will be discussed later in the review. The downfall to this type of ASR method
is cheating is more likely to occur. Students can easily respond slowly to the prompt based on how many hands begin to raise if they are unsure on a correct answer.

Response cards are available in two formats: write-on and preprinted (Heward, 1997). Write-on response cards are small hand-held laminated white boards students can write answers on and simultaneously hold up in response to a teacher’s question. Preprinted response cards have answers printed in advance (e.g., T/F, A, B, C, D, Y/N) and students also simultaneously hold up the cards in response to a teacher’s question. Response cards are one effective way a teacher can increase students’ opportunity to respond in the classroom (Helf, 2015). Most research surrounding response cards has found increases in participation (Khan et al., 2015; Singer et al., 2013) and some have found increases in academic gains (Kellum et al., 2001).

Another effective way to increase active student responding in the classroom is through the use of high technology response systems during instructional periods (Caldwell, 2007). These systems include hardware and software to facilitate instruction. Some of these systems include guided notes and student response systems (Caldwell, 2007; Heward, 1997). What these high technology systems all share in common is a need for a computer in the classroom, and some will require a projector and/or printing student handouts.

A teacher who wants to use guided notes during lecture will prepare in advance handouts that omit key words and/or concepts taught in class to cue students to take notes. Students will attend to lecture to hear or see the blank in the handout so they are able to fill in the correct answer to later study for an exam over the presented material. This technology works well for students that have a difficult time taking notes on important concepts presented in class (Heward, 1997). Guided notes have been found to increase the quality of notes taken during class (Austin,
Lee, & Carr, 2004), in addition to increasing academic scores when compared to students taking their own notes (Lazarus, 1993).

While guided notes have been shown to be effective to improve note taking and academics for primary and secondary students, the results are more robust than for post secondary students (Konrad, Joseph, Eveleigh, 2009). A disadvantage of this technology is that students who are well versed at taking notes, such as students in college who have a long history of note taking behaviors, may miss the lecture trying to attend to the blanks in their guided notes to be sure the blanks are filled in accurately. Also, if errors are made during the completion of guided notes, students may practice the error and receive delayed feedback in the form of a missed exam question. Another disadvantage of using this technology is that it requires much response effort on the teacher’s part. The teacher must omit key concepts, print out the handouts for all students, hand out the guided notes packets to all students and be sure to fill in the blanks for students while lecturing using verbal cues or visual cues in an overhead presentation.

Much like response cards in nature, student response systems (SRS) are becoming a popular instructional tool in higher education to increase active student responding in class. There are over 26 labels used to describe the technology (Kay & LeSage, 2009; Penuel, Kim Boscardin, Masyn, & Crawford, 2007), ranging from audience response system to personal response system. This is a huge concern for researchers and teachers interested in this technology, as multiple names for the same intervention requires an extensive search of the literature to discover benefits, challenges and best practices for using the systems (Kay & LeSage, 2009).

Using a student response system that allows all students to respond will facilitate all students to participate in class and can increase the teachers ability to group students based on
learning characteristics (e.g., group leader, level of knowledge on a topic). Since questions are anonymously answered, teachers are equipped with an ability to ask sensitive questions to a classroom (e.g., political, sexual identity, satisfaction with group members) and receive honest responses from students. “Sometimes issues of gender, culture, or disability affect who participates more and who tends to hold back” (Goldstein, 2013, p. 5). Using a student response system provides teachers with great insight on common misconceptions, attitudes and/or beliefs that students may have and provides a great opportunity to create group discussions that could break down these inaccurate views.

Additional benefits of using a classroom response system includes the ability for teachers to pose a wide variety of questions (e.g., recalling facts, conceptual understanding, application of knowledge to scenarios, critical thinking, student perspective, monitoring activity towards future assignments [e.g., who has started an outline on the paper due at the end of the semester]) to the class, in addition to using the system for various activities (e.g., attendance, quizzes, homework, discussion), all likely strategies to enhance learning at the classroom level (Penuel et al., 2007). While SRSs have been evaluated extensively in higher education, limited research exists in the primary and secondary contexts (DeSorbo, Noble, Shaffer, Gerin, & Williams, 2013; Penuel et al., 2007).

Penuel and colleagues (2007) surveyed 584 elementary and secondary school teachers (209 elementary, 174 middle, 201 high) that were using one brand of SRS (i.e., eInstruction’s Classroom Performance System) to better understand how they were being used in K-12 classrooms. This study was a first attempt to investigate teaching with SRS in K-12 settings. As of 2007, the database indicated there were around 1,000 users across the United States, over one-half responded to the survey. Since the effectiveness of use is determined by the purpose (e.g.,
facilitate discussion, learning checks), the survey asked questions about teachers’ goals for using the system and instructional strategies employed while using the system. Goals of using the system fell into two categories: 1) to improve learning and instruction, and 2) to assess learning. Similar instructional strategies were used when implementing the SRS (i.e., posing questions, sharing a graphical representation of student response data, discussion, feedback to adjust instruction). The findings of using the system to improve assessment and instruction are also reported in the higher education literature, in addition to instructional strategies employed.

Whole-class instruction bears an important role in the learning process. Equally important is the role of questions delivered by the teacher during instruction. The important role of questioning increases engagement, and also allows the teacher to check for understanding. Questions that elicit many responses have been found to be more effective than easy questions or those types of questions that lead students to a single answer (Bruff, 2009). Questions presented before lecture will facilitate tailoring the lecture. Questions presented post-lecture will check student understanding. Teachers need a broad array of questions mapped to their curriculum to make effective use of asking questions during lecture.

There are many types of questions an instructor can pose during lectures to assess students’ comprehension and overall learning of the material. This question presentation method during the learning trial dates back to the Ancient Greeks. Socrates would use a question-based approach to learning, known as an elenchus approach, during his teachings (Plato, trans. 1941). Some examples of question formats would include: recall, conceptual understanding, application, critical thinking, student perspective, confidence level, and monitoring (Bruff, 2009; Smith & Karpicke, 2014).
Recall (i.e., short answer) questions are fairly straightforward in that they require students to recall and produce factual and/or conceptual responses to the instructor's prompt (Bruff, 2009). On the other hand, conceptual understanding questions may evoke students’ misconceptions about a topic to be learned in class. When an instructor poses a conceptual understanding question, the selection of answer choices are important in that they include basic principles that have been misunderstood by students based on the instructor’s past experience teaching a particular concept (Bruff, 2009). For example, an instructor may ask the class which definition best describes the basic principle of reinforcement: a) a stimulus that increases a behavior, b) an increase in behavior in the future, c) in the presence of a stimulus, a behavior is more likely, or d) a stimulus that is added or removed and results in an increase in that behavior in the future. Students who select answer d are likely to have a more accurate understanding of the concept as the other answer choices are correct but missing key components.

Another type of question an instructor may pose during lecture is an application-based question, requiring students to translate their knowledge to a particular scenario (Bruff, 2009). The Behavior Analysis Certification Board exam is primarily made up of application-based questions, requiring students to translate their knowledge of the basic principles of applied behavior analysis and their comprehension of ethical considerations to select the best outcome for a client in a given scenario. Critical thinking questions could be posed as another variation of multiple-choice formats where the instructor must attend to the answer choices provided. This format is more concerned with why students select a particular answer, rather than if they were accurate or not, as a way to foster class discussion (Bruff, 2009). Using critical thinking questions during lecture will require students to respond with the best answer choice among
several correct responses to determine how concepts relate and/or evaluate their own responses based on a given criteria.

Student perspective questions may be presented at the beginning of a semester, for example, to evaluate the type of learners in a given classroom based on demographic information and/or graduate program information (Bruff, 2009). Answers to these types of questions will help instructors tailor their lectures to cover a wide variety of learners. Asking students to rate their level of confidence in a particular answer may provide insight to an instructor, and students, on how students are mastering the content covered during a lecture or from a reading (Bruff, 2009). Confidence level questions may also be useful to the instructor to find concepts taught that are difficult for students to understand. If students rate their level of confidence low, an instructor can be prepared with more examples, or find new ways, to relate the new information to old information during lectures.

Monitoring questions may help an instructor to determine if students are on track with class assignments throughout the semester (Bruff, 2009). For example, a class that requires students to conduct a functional behavior assessment and behavior intervention plan as an end of semester project may ask in the middle of the first few weeks if students have completed their indirect and direct assessments. There are many more types of questions instructors may pose to the class during lecture, and only a few were mentioned above, but all require either a recognition or subsequent selection of a response or a recall of a response.

Multiple-choice and short answer questions are typically used when instructors are interested in measuring a students’ learning (Alba & Pennypacker, 1972; Ozuru, Briner, Kurby, & McNamara, 2013; Roediger III & Marsh; 2005; Smith & Karpicke, 2014). Research suggests that the two types of questions are potentially assessing different comprehension processes.
(Ozuru et al., 2013). Each question format provides different information to the student in the form of a cue, or antecedent stimulus, in that more information is provided using a multiple-choice format, and less information is provided in open-ended format requiring the student to recall the answer from memory (i.e., the strength of stimulus control). Question type has been evaluated in education literature in both basic (Kritch & Bostow, 1998; Nakata, 2016; Roediger III & Marsh, 2005; Smith & Karpicke, 2014) and applied (Alba & Pennypacker, 1972; Mayer et al., 2009; Ozuru et al., 2013) settings.

One role of the instructor is to engage students during lecture, fostering student-instructor interactions. This may be achieved using various questioning methods during lecture. A student who uses active processing during learning is one who attends to relevant material while organizing new content and making relations to prior knowledge (Mayer et al., 2009). “If students do not feel they are involved in the learning situations, they are less likely to work hard to make sense of the presented material and therefore less likely to perform as well as they could on assessments measuring their learning” (Mayer et al., 2009, p. 51).

The elenctic, or Socratic, method during instruction is supported by three broad categories of literature: adjunct question effects, testing effects, and self-explanation effects (Mayer et al., 2009). The literature on adjunct question effects is concerned with the placement (e.g., pretest, posttest) and type (e.g., recall, recognition, factual, conceptual) of questions posed during the learning trial. The remainder of this review will focus on multiple-choice and short answer questions interspersed throughout lectures.

The literature surrounding testing effects has shown that students perform better on exams when they have had a practice test as compared to reviewing the lesson (Alba & Pennypacker, 1972; Mayer et al., 2009; Roediger III & Marsh, 2005; Roediger & Karpicke,
Posing questions during the learning trial is a type of practice for students to assess their learning and for instructors to formatively assess students’ comprehension. The literature surrounding self-explanation effects has shown improved exam scores when students would think aloud, or type, a rationale for a sentence from a text (Chi, DeLeeuw, Chi, & Lavancher, 1994; Magliano & Millis, 2003; Ozuru et al., 2013). Answering teacher-posed questions is a form of thinking aloud during the learning trial. Basic research has found some evidence to support both types of questions (i.e., multiple-choice, short answer) as an assessment of student learning, yet the literature has found mixed results exist on which question type is promoting academic gains in the applied realm. Roediger III and Marsh (2005) warn that there are faults using multiple-choice testing as a measure of assessment during the learning trial as students may walk away from the learning trial with false knowledge. That is, students may attribute the lures (i.e., incorrect answers) on a multiple-choice test as true knowledge especially when students do not study the material.

Nakata (2016) examined the effects of multiple-choice and open-ended questions during the learning phase of Swahili as a second language for English-speaking college students by measuring learning outcomes using same day and next week post tests in an analogue setting. Four different posttests were issued: 1) receptive recognition (i.e., Swahili word presented and students selected the correct response from four options), 2) productive recognition (i.e., English word presented as a cue and students selected the correct response from four options), 3) receptive recall (i.e., Swahili word presented as a cue and students typed the correct English response), and 4) productive recall (i.e., English word presented as a cue and students typed the correct Swahili response) using computer-based software.
Sixty-four students who studied 60 Swahili-English word pairs were randomly divided into four groups. The 60 word pairs were also divided into four groups (i.e., 5 pairs in each group) and each group of students learned each group of word pairs under different learning conditions (i.e., group 1 learned set A under condition 1, set B under condition 2, Set C under condition 3, set D under condition 4, etc.). The four types of learning conditions were: 1) recognition, 2) recall, 3) hybrid (i.e., both), and 4) productive recall only (i.e., receptive recognition format twice, productive recognition format twice). Students were provided with as much time as needed to respond and feedback was given during the learning phase only. As students were learning, number of correct responses and time studying were measured as well as correct response on same day and one week later posttests. Results found as they were learning, recognition conditions in treatment produced the most correct responses, and all four learning conditions produced statistically significant differences from each other.

Posttest performance found a small advantage of the two recall conditions in the immediate and delayed tests. Each posttest contained 60 items, all four posttests were given in the same order (i.e., productive recall, productive recognition, receptive recall, receptive recognition) and all four were the same as in the learning phase (i.e., retrieval formats). All groups experienced all learning conditions and all posttests. There were no significant differences between recall and recognition conditions on the receptive recall posttest (i.e., Swahili word provided as a cue and type the English word). While the recall condition resulted in higher scores on the productive recall posttest (i.e., English word presented as a cue and students typed the correct Swahili response), it did not on the receptive recall posttest and there were no differences between the conditions on recognition posttests. The recognition condition was as effective as the other three conditions on three of the four posttests, was less time
consuming and produced more correct responding during the learning phase. When learners are required to spell the newly learned word correctly, however, recall condition was better. Recall was as effective as recognition if spelling was not considered. Both recall conditions decreased performance in the learning phase, but were effective on the posttests. Since the words were learned in 45 min, the authors suggest extending the time period of study when examining the effects of recall and recognition on learning performance. When a student is learning a new language, recognition questions may be better during the learning phase but recall may be better to transfer knowledge to posttest.

When acquiring new knowledge, readers often self-explain by linking new information to old information through notes in the margin (Ozuru et al., 2013). As an example of applied research, Ozuru and colleagues (2013) examined 41 undergrads text comprehension as measured by both types of questions. Students read a short passage about sexual reproduction while pausing to explain seven pre-selected sentences. They also answered a series of questions about their prior knowledge about the topic. Results found that the quality of notes (i.e., self-explanation) was correlated with performance on open-ended questions. In contrast, the level of prior knowledge related to text was correlated with performance on multiple-choice questions. Accuracy of self-explanations was also correlated to performance on open-ended questions; that is, more accurate explanations were positively correlated, inaccurate explanations were negatively correlated to comprehension. This finding was not correlated to performance on multiple-choice questions.

The authors state that “when participants with low levels of prior knowledge self-explain texts while using effective reading strategies, they reword the text, make connections between ideas and make use of whatever logic and common knowledge they have at their disposal” (p.
These strategies help learners that are new to a particular topic improve their comprehension and knowledge. Whether reading a text about a new topic, or listening to a lecture, active student responding may be a form of actively processing the information to maximize comprehension. The authors suggest future researchers examine effects of specific types of processing (i.e., active, passive) on performance specific questions in more detail.

In another applied example, Mayer et al. (2009) examined the effects of three questioning methods during lecture on exam scores. Across three years, the author delivered the same instruction to education psychology majors in a large lecture hall while manipulating questioning methods (i.e., group questions with clickers, group questions with paper and pencil, no question control). Using a quasi-experimental design, results showed an increase in exam scores for the group questions plus clicker condition, a 1/3 increase in final grades, as compared to the other two conditions. There were no significant differences between the control condition and group question condition. The two groups with in-class questions experienced 2-4 questions per lecture; whereas, the control group was asked if there were any questions at several points throughout the lecture. Both treatment groups also received points (up to 40 points) for correct responding, although this additional points benefit did not show differences between treatment groups without clickers and the control group. It is surprising that the two groups who received questioning during lecture differed on learning outcomes, and the authors suggest that a student response system may have facilitated lecture more seamlessly than a pencil and paper format as the flow of lecture was less interrupted through the use of clickers; whereas, paper and pencil question groups had to pause to pass out the question sheets and hand in the sheets. In contrast, Desrochers and Shelnutt (2012) found writing short answers produced a significantly higher gain on scores (i.e., pretest to posttest) as compared to multiple-choice answer formats.
Active instruction has been shown to improve students’ learning when compared to passive instruction (Kellum et al., 2001; Marmolejo, Wilder, & Bradley, 2004). In an analogue study by Desrochers and Shelnutt (2012), 70 undergraduate students in an intro psychology course were randomly assigned to either a manual (i.e., write letters or short answers on an index card) or automated condition. Students completed a 12-item pretest, viewed a 17 min video instructing over various single subject research designs, and then answered six review questions while receiving feedback from the researcher, with a rationale provided, on the correct answer. Then they viewed a 16 min video of the instructor teaching on single subject designs, answered six review questions using the other answer format (i.e., multiple-choice, short answer). All short answer responses were eight characters in length to facilitate clicker use. Posttests followed the second video and each session lasted 2 hrs total. All participants experienced both answer formats and groups differed in an automated vs. manual condition. Over half of students (i.e., 67%) had higher gain scores in the short answer condition as compared to the multiple-choice condition. This can be seen as a letter grade difference (B to A). Despite the increase in academic scores that resulted from short answer questions, social validity results indicated that students preferred answering questions with a letter rather than a word. While results are impressive, a loss of ecological validity is assumed as students were participants in the study outside of normal class time and the instruction was delivered on a video rather than in vivo.

Active versus passive lectures have been extensively evaluated with active instructional approaches demonstrating superior gains in academics (Gardner, Heward, & Grossi, 1994; Kellum et al., 2001; Marmolejo et al., 2004; Narayan et al., 1990) and social behaviors (Lambert et al., 2006; Khan et al., 2015; Singer et al., 2013). The number of correct answers provided after instruction is also increased using active responding procedures like guided notes and
choral responding compared to lecture only and lecture plus notes taken on blank sheets of paper only (Kreiner, 1997). Whether questions were presented at the end of a lecture in a review format (Desrochers & Shelnutt, 2012) or during a lecture (Kellum et al., 2001; Marmolejo et al., 2004) the academic benefits surpass traditional lecture alone utilizing hand raising. One variable that could have affected the increase in scores is the number of responses. Malanga and Sweeny (2008) accounted for this potential confound by holding the number of responses constant during both active lecture conditions, asking five questions during the short answer condition and 4 questions during the multiple-choice condition. While students scored slightly higher on the end-of-week quizzes in the short answer condition, results may be due to the additional question presented in the condition. Both active student response conditions did produce higher quiz scores as compared to baseline, although procedures used during baseline were not clear and it may have been the review sessions alone, despite type of question presented, that accounted for the increase in quiz scores.

Both Kellum et al. (2001) and Malanga and Sweeny (2008) evaluated active student responding strategies in special education classes at the university level and suggest that future researchers investigate the effects of ASR in other disciplines. Also, few would argue that a review session at the end of a lecture would improve quiz scores, but it is not clear if questions interspersed throughout the lecture would result in similar gains while holding the number of responses constant. Kellum and colleagues (2001) provided support for review questions during lecture at increasing same day quiz scores and Malanga and Sweeny (2008) support for end-of-the-week quiz scores. Both studies preprinted response cards only offered two response options, giving students a 50/50 chance of guessing a correct response. More response options could have lowered the ability to just guess a correct answer, and perhaps the increase in academics
would have been more profound and lower quiz scores during the multiple-choice conditions could have improved.

Marmolejo et al. (2004) interspersed questions throughout lecture, asking six questions per 75 min lecture and having students respond by either holding up a preprinted response card with four multiple-choice options or true/false options or by raising their hands simultaneously to multiple-choice questions they felt were accurate (i.e., “if answer A is correct, raise your hand…”). Results found a 54% increase in quiz scores in the response card condition and 75% of low achieving students increased their performance. A 10% increase in quiz scores could be a letter grade difference, and these results demonstrate a 12% increase in scores as compared to baseline conditions where zero questions were presented. Simultaneous hand raising, or polling with hand raising, resulted in a 2% increase in quiz grades as compared to baseline, suggesting it was a combination of using response cards plus six questions that improved scores rather than questions alone. All quizzes and questions in this study were multiple-choice, requiring students to respond to the cues provided to recognize a correct response.

As mentioned above, over six million students were served under IDEA in the 2015-2016 academic school year with 95% of students spending some portion of the day being educated in regular classrooms (U.S. Department of Education, OSEP, 2017). This diversity in the classroom results in challenges for teachers to manage their classrooms while meeting accountability measures. Teachers are encouraged to engage all students using effective and efficient classroom management. Included in a classroom management plan is effective instruction. When a teacher uses appealing and structured teaching strategies, increases in academic and social behaviors have been found. These strategies are suggested to enhance, rather than replace, traditional lecture. These suggestions stand in contrast to a typical lecture
that is presented in college classrooms where students mostly listen to a lecture presented by an instructor while taking notes. Having students actively respond to an instructor’s question during lecture will involve students in the learning process and improve learning. There are several types of questions that may be presented during the learning trial with mixed results on which question format is superior to increase academic achievement and participation. If students are responding to recognition questions during the learning trial, they must select the correct response from a variety of options. If students are responding to recall questions, they must remember the answer option. If quizzes include both types of questions, then outcomes will likely be a direct result of question type delivered during the learning trial. The first research question this study attempts to answer is which type of instructor-delivered question during the learning trial will enhance academic performance as measured by immediate and delayed quizzes? The second research question this study will attempt to answer is which question type is preferred by both the instructor and students. The purpose of phase one of this study was to evaluate quiz questions’ level of difficulty and content validity by sharing questions with four experienced instructors in the field of ABA who rated each question based on concepts to be taught in each class. The purpose of phase two of this study was to examine two types of questions (i.e., recall and recognition questions) delivered by the instructor during the learning trial on students’ academic and responding behaviors in a college classroom.
Chapter Two:  

Method  

Participants  

During phase one of this study, four experienced instructors in the field of applied behavior analysis (ABA) were recruited using a face-to-face format by the first author. Upon agreeing to participate, instructors assessed the reliability and content validity of quiz questions to be presented throughout the study. Experienced instructors were included if they had at least three years of experience teaching a college course in the field of ABA. The author approached the experienced instructors, explained the nature of the study and asked if they would like to participate in the study. The first four instructors approached by the first author agreed to participate. Three experienced instructors had five years of teaching experience, two of the three instructors had taught the introductory course in prior semesters, and all three had been trained on best practices for delivering active student responding (ASR) procedures including using response cards (RC’s) while delivering lectures. A fourth experienced instructor had 12 years of teaching experience in various ABA courses that spanned from undergraduate to master’s level and experience using and researching ASR procedures during lectures.  

During phase two of this study, one university instructor was selected to participate. This instructor was not a participant in phase one. The instructor was included based on willingness to partake in the study in addition to teaching an introductory course in ABA. The instructor was informed that their participation was voluntary and that they could end their participation in the study at any time. The male instructor was a Caucasian graduate teaching assistant in an ABA
doctoral program at a local university and had taught the same course for five semesters. He had been trained through the university on best teaching practices for an undergraduate course, including how to use ASR procedures and RC’s while delivering instruction.

Forty-four undergraduate students also participated in phase two of this study. Students were either seeking a minor in ABA or taking the course as an elective as part of their major. All students were exposed to the procedures in this study. Since the intervention was a research-supported procedure that has been used in educational settings, informed consent was not necessary on the teaching strategy in which students were exposed. Verbal informed consent was only required from students on their opinions of the teaching procedures gathered for social validity purposes. Students who successfully completed the course received three credits. Students were not informed that they were taking part in a research study, including the nature of the study, until the final lecture day. The section was one of two sections offered during the semester.

Setting

The study was conducted in a university classroom in Florida. The classroom was equipped with rows of tables that allowed two students to sit per table, a white board, a computer, an overhead projector and a screen to pull down for the overhead to be displayed. All observation and intervention sessions were conducted within the natural environment of the classroom during normal university scheduled academic classes. The instructor delivered instruction from the front of the room using PowerPoint slides, a projector and screen.

Phase 1

The primary investigator (PI) conducted an indirect assessment of question type using a survey administered to four experienced instructors in the field of ABA. The survey consisted of
reliability items related to ASR questions to be delivered during instruction and content validity items related to the quiz questions to be administered following instruction.

**Reliability and validity of questions.** Experienced instructors were provided with each class’ lecture materials in the form of PowerPoint presentations that included the 20 class-wide ASR questions to be posed during the lecture and corresponding lecture quiz questions to be administered following the lecture. Next, they were provided with 21 experienced instructors-rating questionnaires (see Appendix A). Each questionnaire was exactly the same but was rated differently based on the 21 quizzes (i.e., both immediate and delayed) and corresponding ASR questions to be delivered across the 21 lectures. On the questionnaire, experienced instructors rated the level of difficulty of quiz questions using a 5-point Likert-type scale (1 = very easy, 2 = easy, 3 = neutral, 4 = difficult, 5 = very difficult) to ensure similar levels of difficulty existed across all questions. If the level of difficulty was similar across sessions, then fluctuations in quiz scores were more likely attributed to question type. In addition to level of difficulty, we asked instructors if the quiz questions reflected the learning objectives of each class. Experienced instructors rated their level of agreement to quiz questions reflecting the learning objectives of the class using a 5-point Likert-type scale (1 = strongly agree, 2 = agree, 3 = neutral, 4 = disagree, 5 = strongly disagree). This ensured quiz questions were relevant to the material and concepts being taught. Experienced instructors also confirmed the reliability of the frequency of recognition and recall questions on all quizzes (i.e., 10 total questions, 5 recall questions, 5 recognition questions) using a yes/no response and the reliability of the frequency of class-wide ASR questions (i.e., 20 questions) using a yes/no response. This helped to ensure fluctuations in quiz scores were not attributed to differing amounts of questions on each quiz or during lectures.
Phase 2

Materials. One of the requirements of the introductory course to ABA was that students purchase the textbook, *Behavior Modification: Principles and Procedures* (Miltenberger, 2016, 6th Ed.). Another requirement of the course was for students to purchase and bring RC’s, blank sheets of paper and/or iPads with them to class along with a dry-erase marker or ink pen to respond to instructor-delivered questions. Materials included up to 44 student RC’s (e.g., dry-erase white boards, blank sheets of paper, iPad application), dry-erase markers and erasers, and/or ink pens. The instructor verbally prompted students to take out their RC’s during each class in which they were used. During classes in which RC’s were not used, the instructor announced to the class that there would be no RC’s used on that particular day. As a way for students to differentiate conditions, the ASR slides were colored green when recall conditions (i.e., short answer) were in place, blue when recognition conditions (i.e., multiple-choice) were in place and white when baseline conditions were in place.

Dependent variables and data collection. The instructor was the primary observer for phase two the study who recorded participation data and the primary investigator (PI) was the secondary observer for the study and sat in the back of the classroom in clear view of all students and the instructor. On days where interobserver agreement data were collected, the instructor and the PI sat in the front of the room in an effort to see each RC answer. Two dependent variables were measured throughout this study: a) academic behavior and b) percentage of student responses. The instructor used quiz scores to record academic behavior. All quizzes were first graded by the instructor’s teaching assistant and then reviewed for grading accuracy by the instructor. The instructor also recorded student responding data during all RC conditions.
using a rating scale and recording the number of students who posed questions during class when baseline conditions were in place.

The rating scale was based on the percentage of students that kept their RC’s on their desk (i.e., did not raise their response cards) following an instructor-delivered class-wide ASR question. The instructor scored a 1 for 1-4 (i.e., 10%) students, a 2 for 5-9 (i.e., 20%) students, a 3 for 10-13 (i.e., 30%) students, a 4 for 14-18 (i.e., 40%) students, or a 5 for 19-22 (i.e., 50%) students that kept their RC’s on the desk rather than raising their cards with an answer to the class-wide ASR questions.

Research was conducted 1-2 days per week on Tuesday and Thursday during the same scheduled academic class time from 9:30-10:55am. The lecture period ranged from 50-60 min, allowing up to 7 min for the delayed lecture quiz at the start of class and up to 7 min for the immediate lecture quiz at the end of class. There were five exam dates in which no lectures were delivered and students only received a delayed lecture quiz. There was also one week off for spring break throughout the spring semester.

**Academic behavior.** The primary dependent variable that was measured in this study was academic behavior. This was defined as students’ answers to various quiz questions. Two quizzes were administered each day to determine the academic behavior of students in the class. The first quiz was a delayed lecture quiz that was administered at the start of class and students had no more than 7 min to complete the 10-item quiz. Students answered 10 questions from the prior class’ lecture, which served as a measure of retention of the material, learned either two or five days prior.

Following each lecture, the instructor administered an immediate lecture quiz. Here, students answered 10 questions (i.e., 5 recall questions, 5 recognition questions) about the
material covered during lecture. Both quizzes (i.e., immediate lecture quiz, delayed lecture quiz) had the exact same 10 questions and only the time of administration differed (i.e., same day, next class).

**Percentage of student responding.** During baseline, the instructor did not ask any class-wide ASR questions, but did encourage students to raise their hands to ask questions at any point throughout the lecture. Those students who did raise their hands to ask questions were recorded using a checklist-recording sheet (see Appendix B) and pen. The instructor also encouraged students to respond if they wanted to by raising their hand or calling out answers. During the RC lectures, the percentage of student responding (i.e., raising the RC) was observed and recorded by the instructor using the same checklist-recording sheet and pen. The recording sheet had 20 intervals representing the 20 class-wide ASR questions to be posed during the RC lectures.

Within each instructional interval, the instructor counted the number of students who did not respond by raising the RC to each question and subtracted that number from the total number of students in the class to determine the number of students who did respond per question. Next, we added the number of students who did respond and divided by the number of opportunities to respond to find the class average of student responding per question. The total number of student responses (i.e., raising the RC) was added each session and divided by the total number of opportunities to respond and then multiplied by 100 for a percentage of students responding per session.

**Inter-observer agreement.** Inter-observer agreement (IOA) was collected for 48% of all conditions (n = 4 for baseline [57%], n = 3 for recall [43%], n = 3 [43%] for recognition). During IOA checks, the PI and the instructor sat in clear view of all students in the front of the room and independently recorded percentage of students who kept their RC on the desk (i.e., not
IOA was calculated as the percent of agreement on the nonoccurrence of students’ responding. During the RC conditions, observers rated the percentage of students who kept the RC’s on the desk for each ASR question (i.e., rate 1 for 10% = 1-4 students, rate 2 for 20% = 5-9 students, rate 3 for 30% = 10-13 students, rate 4 for 40% = 14-18 students, rate 5 for 50% = 19-22 students). Next, ratings were summed for each class per observer. Finally, sums were divided to determine the percentage of agreement per RC condition.

The formula for determining the percentage of agreement was calculated by dividing the smaller rating by the larger rating and then multiplying by 100 for an overall percentage (S / L x 100). Results of IOA checks for participation during RC and baseline conditions averaged 90.4% (range = 70% - 100%) while recording the nonoccurrence of student responding using RC following a class-wide ASR question and while recording the number of students who raised their hands during baseline conditions. IOA per condition was averaged at 100% during BL, 80% (range = 70% - 95%) during recall and 88.3% (range = 84% - 96%) during recognition.

IOA was also calculated on academic scores for 100% of immediate quiz grades and delayed quiz grades by having a second observer grade each quiz to determine if both observers recorded the same score per quiz. The formula for determining the percentage of agreement was calculated by dividing the smaller quiz grade by the larger quiz grade and then multiplying by 100 for a percentage (S / L x 100). Inter-observer agreement for accuracy of quiz grading was 100% for every quiz.

**Experimental design.** An alternating treatments design with a baseline condition (BL) was used to evaluate the effects of question type on academic behavior and student responding. Because the goal of the study was to examine the effects of two interventions on individual
students’ behaviors, this single-subject design was most appropriate for answering the research questions since each participant served as their own control within this design (Cooper, Heron, & Heward, 2007). Using this design allowed data to be analyzed individually with respect to student responding behavior during lecture.

The BL, RC with recognition question, and RC with recall question conditions were rapidly alternated across sessions. To prevent predictability of the order, all sessions were randomized in sets of three. The three conditions (i.e., BL, RC with recognition questions, RC with recall questions) were placed in a hat and selected without replacement to determine which session experienced which condition.

**Procedures**

The instructor implemented all of the BL and RC conditions during normal class dates and times. The material from the curriculum was not affected; only the delivery of instruction changed throughout all sessions.

**Instructor training.** The instructor was trained by the PI on how to deliver lecture during all conditions using behavioral skills training. During a face-to-face meeting, the PI provided a rationale to the instructor for the intervention while verbally instructing plus modeling the procedures involved. Following instruction plus modeling, the PI provided the instructor an opportunity to rehearse the procedures for all three conditions (i.e., BL, RC with recall, RC with recognition). Steps taught for BL included:

1. Verbally and visually present 10 recall statements to the class (e.g., “Differential reinforcement includes what two basic principles of applied behavior analysis?”) and verbally and visually present 10 recognition statements to the class (e.g., “Differential reinforcement includes: a] reinforcement, b] punishment, c] extinction, d] both a and c”).
2. Verbally present the answer to the class.

Steps taught for both intervention conditions included:

1. Verbally and visually present a class-wide ASR question to the class (e.g., “Differential reinforcement includes what two basic principles of applied behavior analysis?” or “Differential reinforcement includes: a) reinforcement, b) punishment, c) extinction, d) both a and c”).

2. Provide adequate wait time for students to use response cards (e.g., enough time for ¾ of students in classroom to respond [10-15 s]).

3. Verbally present the answer to the class.

4. Verbally provide a general praise statement or corrective feedback to the whole class for responses (e.g., “I see all answered this question correctly,” “I see most answered this question correctly,” “I see none of you answered this question correctly, let us review the concept”).

If needed, corrective feedback was provided until the instructor demonstrated mastery of the procedures (i.e., 100% of steps performed correctly).

**Baseline.** During this condition, the instructor delivered lecture only. This condition was typical of college classrooms that administer lecture only formats (Faust & Paulson, 1998; Michael, 1991; Newman Thomas et al., 2015). The instructor started the class with a delayed lecture quiz (i.e., retention questions from the prior lecture). Following the delayed lecture quiz, the instructor asked students if there were any questions that needed addressed before starting lecture. Then the instructor delivered a lecture for approximately 60 min while reading off a combination of 10 recall and 10 recognition statements with corresponding answers as a way to review the material just taught. Following lecture, an immediate lecture quiz was administered.
If students raised their hand during the lecture to ask questions, the instructor answered student questions as they normally would in class. The instructor was advised to teach as he normally would with the addition of ensuring 20 questions (i.e., 10 recall, 10 recognition) were interspersed throughout the delivery of lecture. The instructor allowed students to respond using call-outs or hand raises during baseline conditions if students wanted to respond. On average, 1-3 students would respond to most questions by calling out answers.

**Response cards intervention.** There were two different RC conditions: Response cards with recall questions and RC with recognition questions. The PI and the instructor using the instructor’s selected textbook (Miltenberger, 2006) developed content questions for both conditions. All questions were delivered vocally by the instructor and textually on the slides. Prior to each lecture, and following the delayed lecture quiz, the instructor cued students verbally to take out their RC’s in an effort to help students differentiate between baseline and intervention conditions. In order to differentiate between RC conditions, the background colors of the ASR slides were different (i.e., blue for recognition questions, green for short answer questions).

For both conditions, an instructional interval consisted of an instructor-posed class-wide ASR question, a pause for student responding (e.g., 10-15 s), followed by the instructor’s visual scan of student responses and feedback to the entire class (e.g., “I see most answered the question correctly,” “perhaps we need to review this question”).

Steps for each instructional interval included: 1) vocally presenting the question to the class while visually displaying the question on the slide, 2) provide adequate wait time for students to respond, 3) glance around the room at each card with no individual feedback provided, 4) vocally reveal the answer to class, and 5) provide positive and corrective feedback if needed. Corrective feedback was delivered by the instructor if 50% of students responded
incorrectly and the concept was taught again. Throughout the study, there was a low rate of student error when responding using RC; therefore, corrective feedback from the instructor occurred at near zero levels. Spelling of concepts on RC was not assessed nor was accuracy of response, only student responding was recorded. Students were permitted to look at others’ response cards. If students were late and the instructor was on the first 1-5 slides (i.e., no ASR slides were presented) then their immediate quiz grade was kept. If a student was late to class after ASR slides were presented then their immediate and delayed quiz scores were not included.

**Response cards with recall.** When the instructor was teaching using RC with recall questions (i.e., short answer questions), the background color of the slide was green. The instructor presented a question to the class and then the students responded by writing a one or two word answer on their RC’s, displaying their cards immediately, and then awaiting instructor feedback.

**Response cards with recognition.** When the instructor was teaching using RC with recognition questions (i.e., multiple-choice questions), the background color of the slide was blue. During this condition, all questions were a recognition option requiring students to respond by writing the letter A, B, C, or D on the card.

**Treatment Integrity**

The PI conducted treatment integrity checks for at least 29% of all intervention conditions to ensure that the instructor was following the structured format for conducting the intervention. If integrity fell below 100%, the PI provided a refresher training that included positive and corrective feedback before the next intervention session. Steps for instructor-delivered questions for each RC condition included:
1. Verbally and visually present a question to the class (e.g., “Differential reinforcement includes what two basic principles of applied behavior analysis?” or “Differential reinforcement includes: a) reinforcement, b) punishment, c) extinction, d) both a and c”).
2. Provide adequate wait time for students to use response cards (e.g., 10-15 s).
3. Verbally present the answer to the class.
4. Verbally provide a general praise statement or corrective feedback to the whole class for responses (e.g., “I see all answered this question correctly,” “I see most answered this question correctly,” “I see none of you answered this question correctly, let us review the concept”).

A checklist was used for fidelity measures (see Appendices C and D), outlining the steps for implementing the BL conditions and RC conditions during each instructional interval. Treatment integrity data was collected by the PI and calculated by dividing the number of steps performed correctly by the total number of steps and then multiplied by 100 for a percentage.

The PI recorded the instructor’s behavior during 29% of the intervention conditions (n = 6). Treatment integrity was calculated to be 95% (range = 80% - 100%) across all three conditions. Treatment integrity was recorded during three baseline conditions at 100%, during three recall conditions at 90% (range = 80-100) and during one recognition condition at 100%. The only time treatment integrity fell below 100% was when the instructor did not provide a praise statement following a class-wide ASR question, resulting in corrective feedback and brief refresher training.

Social Validity

This study evaluated both instructor (see Appendix E) and student (see Appendix F) social validity using open-ended and forced-choice survey questions. The forced-choice items
were anchored on a 5-point Likert-type scale ranging from 5 (*strongly agree*) to 1 (*strongly disagree*), with the middle option representing a neutral anchor. A separate student and instructor survey was administered following completion of the study. At the end of the study, the PI administered the brief questionnaire to the students and data from only those students who agreed to the informed consent script were obtained. The instructor survey was administered via email following completion of the study with a request to return the survey within one week.

Students were asked to rate their perception of participation when all three conditions were in place: using RC’s with recognition questions, using RC’s with recall questions, listening to lecture. Students were also asked what they liked best about using RC’s and what they liked the least about using RC’s during lectures and which question type they preferred.

The instructor’s questionnaire ratings ranged from 1 (*strongly disagree*) to 5 (*strongly agree*) with a neutral anchor in the middle. Items on the survey asked the instructor to rate the procedures’ ease of use, likeability of the intervention to be used in the course and if students answered more questions correctly when using RC’s with recall or recognition questions during the learning trial. The instructor was also asked what was the best part of implementing RC’s during lectures, what was the worst part about implementing RC’s during lectures and to rate the overall utility of the intervention.
Chapter Three:

Results

Phase 1

The average results of the four experienced instructor raters’ questionnaires are displayed in Table 1. There were 21 total questionnaires answered by each experienced instructor rater that corresponded to the 21 quizzes and lectures to be delivered. When instructors were asked if quiz questions for each lecture reflected the learning objectives of the class, they strongly agreed by scoring a 1 (range = 1 – 1.18). Experienced instructors rated similar levels of difficulty for all quiz questions by rating an average score of 2 (range = 2.18 – 3.08) indicating an average of easy questions (i.e., 1 = very easy, 2 = easy, 3 = neutral, 4 = difficult, 5 = very difficult). In other words, quiz questions ranged in ratings from easy (i.e., 2) to middle of the road (i.e., 3) on level of difficulty. All four experienced instructors answered yes when asked if there were 10 quiz questions and yes when asked if there were 20 active student responding (ASR) questions per lecture.

Phase 2

The results of phase two of the study were analyzed for the main dependent variable (i.e., academics) five ways: the mean immediate and delayed quiz scores for all students across quiz numbers, the mean immediate and delayed quiz scores per condition across all individual students, the mean immediate and delayed quiz scores for the lower 20% of students (as determined by exam scores) across quiz numbers, individual student quiz scores and statistically. Results are displayed in Figures 1 to 9.
The results of the effects of question type delivered during instruction across all students’
average immediate quiz score were 8.34 (range = 5.8 – 9.1) during baseline (BL), 8.72 (range =
7.9 – 9.7) during recognition and 8.38 (range = 7.26 – 9.26) during recall conditions (see Figure
1). The results of the effects of question type delivered during instruction across all students’
delayed quiz scores was 8.47 (range = 5.97 – 9.65) during BL, 8.75 (range = 7.92 – 9.88) during
recognition and 8.38 (range = 7.06 – 9.47) during recall conditions (see Figure 2). Visual
analysis revealed no differentiation per condition with all three data paths overlapping in both
immediate and delayed quiz scores. Table 2 displays the average results of all students’
immediate quiz scores for each condition across all 21 quizzes and Table 3 displays the average
results of all students’ delayed quiz scores for each condition across all 21 quizzes.

The results of the effects of question type delivered during instruction across individual
students per condition on immediate quizzes were 8.38 (range = 5.71 – 10) during BL, 8.77
(range = 6.86 – 10) during recognition and 8.38 (range = 5.29 – 10) during recall conditions (see
Figure 3). The results of the effects of question type delivered during instruction across
individual students per condition on delayed quizzes were 8.56 (range = 5 – 10) during BL, 8.77
(range = 6.5 – 10) during recognition and 8.42 (range = 5.21 – 10) during recall conditions (see
Figure 4). Visual analysis revealed no differentiation per condition with all three data paths
overlapping in both immediate and delayed quiz scores.

Exam scores were analyzed and students that received a C or lower (i.e., 70% or less) on
all five exams were included in the next analysis. These students (n = 9) represented the lower
20% of the class and their data were further analyzed to determine the effects of the intervention
on quiz scores. The results of the lower 20% of students’ average immediate quiz scores were
7.54 (range = 3.67 – 8.75) during BL, 7.96 (range = 6.86 – 9.25) during recognition and 7.8
(range = 6.5 – 9) during recall conditions (see Figure 5). The results of the lower 20% of students’ average delayed quiz scores were 7.64 (range 4.25 – 9.29) during BL, 8.1 (range = 4.25 – 9.29) during recognition and 7.69 (range = 5.8 – 9.33) during recall conditions (see Figure 6). Visual analysis revealed no differentiation per condition with all three data paths overlapping in both immediate and delayed quiz scores.

One student’s quiz score data (i.e., student 2) was analyzed individually across quiz numbers for both immediate and delayed quizzes. Student 2’s average quiz immediate score in BL was 7.92 (range = 5 – 10), in recognition condition was 8.71 (range = 7 – 10) and in recall condition was 7.33 (range = 6 – 9; see Figure 7). Student average delayed quiz score in BL was 8.6 (range 6 – 10), in recognition condition was 9.33 (range = 8 – 10) and recall condition was 8 (range = 6 – 10; see Figure 8). Student 2 was also in the lower 20% of students in the class. Visual analysis revealed no differentiation per condition with all three data paths overlapping in both immediate and delayed quiz scores.

A series of unpaired two-tailed t-tests were conducted to determine the statistical significance of comparing all three conditions in both immediate (i.e., BL I, recall I, recognition I) and delayed (BL D, recall D, recognition D) quiz scores to each other. Results of all statistical analyses were not significant (see Table 4).

Participation per condition was analyzed to determine the effects of instructor-delivered ASR questions on students’ participation behaviors (i.e., raising the response cards during recognition and recall conditions, asking questions during baseline conditions). During the BL condition, students asked an average of 1.14 questions per class (range = 0 – 5). During the recognition condition, students participated in responding to the instructor-delivered ASR question an average of 90.7 % (range = 87% – 96.5%) and participated in responding to the
instructor-delivered ASR question during the recall condition an average of 88.94% (range = 82% - 94.8%). Visual analysis revealed that while both recognition and recall conditions data paths overlapped, baseline condition data demonstrated low levels of student participation that were on a decreasing trend (see Figure 9).

**Instructor Social Validity**

Results of the instructor social validity revealed he strongly agreed that students were more engaged when using response cards, he agreed to use recall and recognition questions during this subject area when taught in the future and that he agreed to enjoying using response cards during lectures. Neutral ratings were provided when asked if the procedures were easy to use in the classroom, if students were answering more recall questions correctly and when asked if students were answering more recognition questions correctly (see Table 5). The best part about implementing the intervention in the classroom was reported to be the use of response cards during lecture, although he did not feel it was important which question type was presented and felt fewer questions would be equally effective. When asked what the worst part about implementing the intervention in the classroom responses included too many questions being prepared and presented during lectures, too many quizzes which he felt negatively impacted the grades of some students compared to his baseline rates across other semesters. He felt the many questions and quizzes were overwhelming and boring to some students. When asked what could be done differently to have instructors implement the procedures during their lectures he responded:

“Often times I found that students who did not participate in the active student responding still did well on the quizzes and tests. Therefore, it’s important to consider everyone’s’ learning capabilities before a person goes gung ho on which
teaching strategy will be used. Because there are normally numerous students in class, I think it is important to not overwhelm them with white board questions, or ask too few white board questions. Instead, consider a happy medium in which relevant examples, white board questions, normal hand-raising questions, and other strategies (including technology, videos, and more). Also, some material actually works better with multiple-choice, whereas other material works better with short answer. Therefore, one should not pigeonhole all material into being taught in exactly the same fashion.”

Overall, he rated the instructional approach useful and had no preference on question type but stated short answer questions were easier to prepare.

**Student Social Validity**

Of the 44 students in the class, 39 students were in attendance on the day social validity data were obtained. Of the 39 students in attendance, all but one student responded to most items on the survey. When asked which question-type students preferred during lectures, one of the 38 students preferred both question types and one of the 38 students preferred neither question type. Fourteen students preferred recall questions to be delivered during instruction. When asked why they preferred this type of instruction, student responses varied from: the professor provided a more in-depth explanation for why an answer may have been right or wrong (1 student), helped me to know if I understood the material (5 students), helped me to know if I am learning the material (5 students), it was more engaging (1 student), easier and faster to respond and less pressure and anxiety (1 student) and one student left the open-ended question blank. When the same students were asked what they liked best about the recall condition, responses included: using response cards (1 student), the variety of techniques used to teach lecture (2 students), how
the quizzes were set up in the beginning of class (1 student), everything (1 student), the amount of practice (1 student), sufficient amount of time given to write answers down (1 student), class was more interesting, interactive and fun (3 students), and four students left the response blank. When the same students were asked what they liked least about the recall condition, responses included: they did not like the amount of quizzes delivered (4 students), I liked everything (2 students), bringing response cards to class (1 student), sometimes it took a long time (1 student), five students left the answer blank and one student’s response was not legible.

Twenty-two of the 38 students preferred recognition questions to recall questions. When asked why they preferred this type of class-wide ASR question delivered during instruction, student responses varied from: it was difficult recalling the correct vocabulary (6 students), easier to retain information and/or narrow down the correct answer (14 students), it allowed me to think about each choice in relation to the question (1 student) and it forced me to pay attention (1 student). When the same students were asked what they liked best about the intervention, responses included: I was more actively engaged (5 students), I could show comprehension without talking (1 student), all the prompts became examples (1 student), learning about the best way to teach/learn (1 student), I could retain the information better (3 students), using response cards (1 student), not knowing about the intervention (2 students), positive reinforcement when you get the answer right (1 student), opportunity to study outside of class (1 student), quick and easy to discuss (1 student), and five students left the answer blank. When the same 22 students were asked what they liked least about the recognition condition, responses included: the amount of quizzes (3 students), I got frustrated when I couldn’t think about the correct words (1 student), room for more error (1 student), the amount of questions delivered and recall questions (1 student), I liked everything (2 students), and the grade impact resulting from the quizzes (1
student), I did not remember recognition questions as much as I did recall questions (1 student)
and twelve students left the answer blank.

Students were asked to rate from 1 (strongly disagree) to 5 (strongly agree) their level of
agreement to participating more during lectures when recall conditions were in place, to
participating more during lectures when recognition conditions were in place and to participating
more during lectures when baseline conditions were in place. Table 6 displays students’ average
responses to each question.

**Table 1.** Average Results of Experienced Instructor Rater Questionnaires

<table>
<thead>
<tr>
<th>Quiz Number</th>
<th>Rating of quiz questions matching learning objectives of each class</th>
<th>Rating of level of difficulty for each quiz question</th>
<th>Are there 10 quiz questions with 5 recall and 5 recognition</th>
<th>Are there 20 ASR questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.2</td>
<td>2.3</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2.5</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>3</td>
<td>1.1</td>
<td>2.2</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>2.7</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2.6</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>2.6</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>2.4</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>2.6</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>2.6</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>2.6</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>2.8</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>3.1</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>2.4</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>2.6</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>2.9</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>2.9</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>2.4</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>3</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>2.6</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>2.5</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
<td>2.9</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

The data in Table 1 represents the average of ratings from four experienced instructors across
each quiz number (e.g., 1 = strongly agree, 3 = neutral, 5 = strongly disagree).
Table 2. Results of All Students’ Average Immediate Quiz Scores across Conditions

<table>
<thead>
<tr>
<th>Quiz Number</th>
<th>Average Baseline Quiz Score (range)</th>
<th>Average Recognition Quiz Score (range)</th>
<th>Average Recall Quiz Score (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>9.7 (range = 6 - 10)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>9.3 (range = 6 - 10)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>8.9 (range = 5 - 10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>8.3 (range = 5 - 10)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>8.8 (range = 6 - 10)</td>
<td></td>
<td>8.8 (range = 6 - 10)</td>
</tr>
<tr>
<td>6</td>
<td>9 (range = 5 - 10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>7.9 (range = 2 - 10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>8.4 (range = 3 - 10)</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>9 (range = 5 - 10)</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>8 (range = 3 - 10)</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>8.7 (range = 5 - 10)</td>
<td></td>
<td>8.8 (range = 6 - 10)</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>8.7 (range = 6 - 10)</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>8.3 (range = 4.5 - 10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>5.8 (range = 0 - 10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>8 (range = 5 - 10)</td>
<td></td>
<td>8 (range = 4 - 10)</td>
</tr>
<tr>
<td>16</td>
<td>8.8 (range = 5 - 10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>9.3 (range = 7 - 10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>7.3 (range = 3 - 10)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The data in Table 2 represents students’ average immediate quiz score across baseline, recognition and recall conditions.

Table 3. Results of All Students’ Average Delayed Quiz Scores across Conditions

<table>
<thead>
<tr>
<th>Quiz Number</th>
<th>Average Baseline Quiz Score (range)</th>
<th>Average Recognition Quiz Score (range)</th>
<th>Average Recall Quiz Score (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>9.9 (range = 8 - 10)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>9.5 (range = 5 - 10)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>9.7 (range = 5 - 10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>9 (range = 6 - 10)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>9 (range = 6 - 10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>9 (range = 2 - 10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>9 (range = 5 - 10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>7.9 (range = 4 - 10)</td>
<td></td>
</tr>
</tbody>
</table>
The data in Table 3 represents students’ average delayed quiz score across baseline, recognition and recall conditions.

Table 3. (Continued)

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td>8.4 (range = 5 - 10)</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>8.8 (range = 4 - 10)</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td>8.1 (range = 5 - 10)</td>
</tr>
<tr>
<td>12</td>
<td>8.6 (range = 5 - 10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td>8 (range = 2 - 10)</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td>8.3 (range = 5.5 - 10)</td>
</tr>
<tr>
<td>15</td>
<td>8.4 (range = 3 - 10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>6 (range = 0 - 10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
<td>8 (range = 4 - 10)</td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
<td>8.6 (range = 3 - 10)</td>
</tr>
<tr>
<td>19</td>
<td>8.4 (range = 5 - 10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td>9.4 (range = 6 - 10)</td>
</tr>
<tr>
<td>21</td>
<td></td>
<td></td>
<td>7.1 (range = 3 - 10)</td>
</tr>
</tbody>
</table>

The data in Table 3 represents a series of t-tests comparing each condition to another. There were no significant results of all tests completed with limited variability between conditions.

Table 4. Statistical Analysis of Conditions

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Immediate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline Delayed</td>
<td>7</td>
<td>8.55</td>
<td>1.90</td>
<td>p &lt; 0.84</td>
</tr>
<tr>
<td>Recognition Immediate</td>
<td>7</td>
<td>8.74</td>
<td>1.45</td>
<td>p &lt; 0.47</td>
</tr>
<tr>
<td>Recognition Delayed</td>
<td>7</td>
<td>8.77</td>
<td>1.42</td>
<td>p &lt; 0.94</td>
</tr>
<tr>
<td>Recognition Immediate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recall Immediate</td>
<td>7</td>
<td>8.38</td>
<td>1.68</td>
<td>p &lt; 0.36</td>
</tr>
<tr>
<td>Recognition Delayed</td>
<td>7</td>
<td>8.77</td>
<td>1.42</td>
<td>p &lt; 0.94</td>
</tr>
<tr>
<td>Recall Immediate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recall Delayed</td>
<td>7</td>
<td>8.38</td>
<td>1.65</td>
<td>p &lt; 0.99</td>
</tr>
<tr>
<td>Baseline Delayed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recognition Delayed</td>
<td>7</td>
<td>8.77</td>
<td>1.42</td>
<td>p &lt; 0.61</td>
</tr>
<tr>
<td>Recall Delayed</td>
<td>7</td>
<td>8.38</td>
<td>1.65</td>
<td>p &lt; 0.86</td>
</tr>
<tr>
<td>Recognition Delayed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recall Delayed</td>
<td>7</td>
<td>8.38</td>
<td>1.65</td>
<td>p &lt; 0.36</td>
</tr>
</tbody>
</table>

The data in Table 4 represents a series of t-tests comparing each condition to another. There were no significant results of all tests completed with limited variability between conditions.
Table 5. Instructor Social Validity Results

<table>
<thead>
<tr>
<th>Instructor Rating</th>
<th>Ease of use</th>
<th>Continued use of recall</th>
<th>Continued use of recognition</th>
<th>Enjoyment</th>
<th>Engagement</th>
<th>Recognition condition produced more correct answers</th>
<th>Recall condition produced more correct answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rating</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

The data in Table 5 represents the instructor's ratings of procedures used in the study (i.e., 1 = strongly disagree, 3 = neutral, 5 = strongly agree).

Table 6. Student Social Validity Results

<table>
<thead>
<tr>
<th>Preferred question method</th>
<th>I participated more with recognition</th>
<th>I participated more with recall</th>
<th>I participated more when no questions were posed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preferred recall (n = 14)</td>
<td>X = 3.5 (range = 1-5)</td>
<td>X = 3.7 (range = 1-5)</td>
<td>X = 2.2 (range = 1-4)</td>
</tr>
<tr>
<td>Preferred recognition (n = 22)</td>
<td>X = 4.4 (range = 3-5)</td>
<td>X = 3.1 (range = 1-4)</td>
<td>X = 2.2 (range = 1-5)</td>
</tr>
<tr>
<td>Preferred both (n = 1)</td>
<td>4</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Preferred neither (n = 1)</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

The data in Table 6 represents the average quantitative results of students’ ratings of procedures used in the study (i.e., 1 = strongly disagree, 3 = neutral, 5 = strongly agree). Students either preferred recall conditions, preferred recognition conditions, preferred both types of questions or preferred neither condition as listed under the preferred question method column. Based on their preference of condition, their results were then averaged and presented as the mean result throughout the table. Of the 44 students that were in the class throughout the semester, only 38 students were in attendance the day social validity results were obtained. The remaining six students were asked if they would like to submit a survey via email and no response was received following completion of the study.
Figure 1. Immediate Average Quiz Scores
Immediate class averages quiz scores across baseline (BL), recall (short answer [SA]), and recognition (multiple choice [MC]) conditions. The asterisk above quiz 2 denotes a question mistake on number 7 of 10 in which all students were awarded points.

Figure 2. Delayed Average Quiz Scores
Immediate class averages quiz scores across baseline (BL), recall (short answer [SA]), and recognition (multiple choice [MC]) conditions. The asterisk above quiz 3 denotes a procedural mistake in which the instructor reviewed the immediate quiz prior to delivering the same quiz as a delayed quiz that could have potentially influenced the quiz scores.
Figure 3. Average Individual Student Immediate Quiz Scores
All students’ average quiz scores for each condition (i.e., BL, recognition [MC], recall [SA]) across students (n = 44). There were 7 immediate quizzes per condition.

Figure 4. Average Individual Student Delayed Quiz Scores
All students’ average quiz scores for each condition (i.e., BL, recognition [MC], recall [SA]) across students (n = 44). There were 7 delayed quizzes per condition. The x-axis represents students’ mean quiz score and the y-axis represents student number.
Figure 5. Lower 20% Immediate Quiz Scores
The above graph represents the lower 20% of students (n = 9) who scored a C or lower on all exams. Data represents average quiz scores per quiz number for each condition (i.e., BL, MC, SA). The asterisk above quiz 2 denotes a question mistake on number 7 of 10 in which all students were awarded points.

Figure 6. Lower 20% Delayed Quiz Scores
The above graph represents the lower 20% of students (n = 9) who scored C or lower on all exams. Data represents average quiz scores per quiz number for each condition (i.e., BL, MC, SA). The asterisk above quiz 3 denotes a procedural mistake in which the instructor reviewed the immediate quiz prior to delivering the same quiz as a delayed quiz that could have potentially influenced the quiz scores.
Figure 7. Individual Student Immediate Quiz Scores
The above graph represents an individual student’s immediate quiz scores per condition. Missing quiz grades are due to either an absence or late arrival to class resulting in data to be omitted.

Figure 8. Individual Student Delayed Quiz Scores
The above graph represents an individual student’s delayed quiz scores per condition across quiz numbers. The asterisk above quiz 3 denotes a procedural mistake in which the instructor reviewed the immediate quiz prior to delivering the same quiz as a delayed quiz that could have potentially influenced the quiz scores. Missing quiz grades are due to either an absence or late arrival to class resulting in data to be omitted.
Figure 9. Average Student Responses
Average student responses across recall (short answer [SA]) and recognition (multiple-choice [MC]) conditions. The x-axis represents the class number and the y-axis represents the average percentage of class participation of students (i.e., raising the response card during conditions).
Chapter Four:  
Discussion

The research questions for this study were which type of instructor-delivered question (i.e., recall [short answer], recognition [multiple-choice]) during the learning trial would enhance students’ academic performance as measured by immediate and delayed quizzes in addition to enhancing students’ participation. Another research question for this study is which question type was preferred by both the instructor and students. This study was divided into two phases. The purpose of phase one was to evaluate quiz questions’ level of difficulty and quiz questions’ content validity by sharing questions with four experienced instructors in the field of applied behavior analysis (ABA) who rated each item based on concepts to be taught in each class. Phase one also looked at the reliability of quiz items to ensure 10 questions were on each quiz (5 recall questions, 5 recognition questions) and 20 active student responding (ASR) items were interspersed throughout each of the 21 lectures.

Results of the experienced instructor raters found similar levels of difficulty existed across all quiz items by rating each question on average as an easy question (i.e., range of 2 = easy to 3 = neutral on level of difficulty). Since the level of question difficulty was similar across sessions, fluctuations in quiz scores might be more likely attributed to question type. Additional results of the survey found that experienced instructors strongly agreed that all quiz items did reflect the learning objectives of the class (e.g., all questions from quiz 2 were about observing and measuring behavior, all questions from quiz 3 were about graphing and measuring behavior change). This ensured quiz items were relevant to the material and concepts being
taught. Raters also ensured that there were 10 quiz items (5 recall questions, 5 recognition questions) delivered on each quiz to confirm that the assessment of learning did not differ across quizzes. Lastly, raters confirmed that there were 20 ASR questions to be delivered by the instructor during every lecture. Experienced instructors also confirmed the reliability of the frequency of recognition and recall questions on all quizzes (i.e., 10 total questions, 5 recall questions, 5 recognition questions) using a yes/no response and the reliability of the frequency of class-wide ASR questions (i.e., 20 questions) using a yes/no response. This helped to ensure fluctuations in quiz scores were not attributed to differing amounts of items on each quiz or during lectures.

The purpose of phase two of this study was to examine the effects of two types of questions (i.e., recall, recognition) delivered by the instructor during the learning trial on students’ academic and responding behaviors in a college classroom. Anticipated results were expected to demonstrate that both intervention conditions would be superior to the standard lecture style in baseline with respect to students’ academic behavior and students’ responding. Unfortunately, the baseline in this study was not standard lecture style as the instructor provided multiple opportunities for individual students to respond across all conditions. That is, the instructor asked an average of 31 relevant individual questions per class despite the condition. It was expected that students’ academic behavior would be improved as a result of asking students questions during the lectures as this provided students with an opportunity to practice the skill just learned from listening to the lecture. When an instructor increases the rate of opportunities to respond, research has shown increases in academics and social behaviors (Christle & Schuster, 2003; Khan et al., 2015; Lambert et al., 2006; Singer et al., 2013). Anticipated results of this study were that academic behaviors may be slightly lower in the response card (RC) with recall
condition on the immediate lecture quiz, but retention of material learned would likely be higher during this condition as measured by the delayed quizzes and as demonstrated in research (Alba & Pennypacker, 1972). Results of the study differed from research supporting the inclusion of recall questions during the learning trial to enhance academics (Alba & Pennypacker, 1972), as both question types demonstrated no differentiation in students’ academic behavior. This finding in the research may be due to the response effort required to produce a response during the learning trial rather than simply recognizing the response (Alba & Pennypacker, 1972; Nakata, 2016). The results of our study may have differed due to all students learning the material similarly and subsequent quiz scores reflecting the similar learning conditions, rather than dividing the students into groups and teaching using one of the question types during instruction.

In contrast, quiz scores during the RC with recognition condition were anticipated to be initially higher as students’ response effort of recognizing the correct answer would be lower. This outcome was expected to affect retention of the material on delayed quizzes by recording a reduction in scores as demonstrated in research (Alba & Pennypacker, 1972; Nakata, 2016) but our study did not find this same effect. Results of our study differed from some findings in research as our study found no differentiation in academic behavior despite the type of question delivered during the learning trial. The baseline levels for quiz grades were higher than anticipated possibly due to a ceiling effect. This finding may be a result of the type of assessment being delivered following instruction, rather than a posttest delivered at completion of the semester as done in prior research. Perhaps achievement should be measured using exams following units of learning rather than quizzes each day. Also, if quiz questions had higher levels of difficulty, differentiation in scores may have been achieved.
To further analyze the academic data, we used exam scores as a measure of class performance due to the overall final score being influenced by many extra credit opportunities (e.g., for students participating as a research assistant, participating in the pretest, participating in the posttest). All exams were analyzed and students who scored a C (i.e., 70%) or lower on all five exams were considered the lower 20% of the class (n = 9). Results found no differentiation between conditions for this lower 20% of students as averages were found to be similar on both the immediate (BL = 7.54, recognition = 7.96, recall = 7.8) and delayed (BL = 7.64, recognition = 8.1, recall = 7.69) average quiz outcomes.

While there were no differentiations in students’ academic performance related to question type, there were differences in student participation behaviors. More students participated in lectures when both RC conditions were in place as compared to baseline. Data were not collected on students’ responses to instructor-delivered questions during baseline; rather, data were collected on the frequency of students posing a question to the instructor during this condition. It was recorded by the secondary investigator that when the instructor posed a question in baseline, an average of 1-3 students would call out a response to some questions while other questions were met with zero student responding. This is problematic for the instructor who is changing their teaching style in the moment based on students’ responses indicating that they were not learning the concept just taught. While our study did find higher levels of students participating during RC conditions, there was no differentiation in data paths for question type delivered. That is, student responding was at similar levels in both recall and recognition conditions. It was anecdotally recorded that short answer questions may have required more time during the learning trial, although all lectures averaged 52.24 min (range = 25 min – 70 min) and all lectures were completed in that amount of time despite the type of
question delivered during instruction. Some classes included group activities that took 8-15 min (i.e., class 2, 12). Four classes included exam review that took up to 15 min.

Short answer questions were anecdotally recorded by the secondary investigator to require more time (e.g., 5 s – 10 s) during the learning trial for students to recall an answer, write the answer and for the instructor to deliver the correct response. One student responded on an open-ended social validity question about the best part of the intervention was that the instructor did provide an adequate wait time to recall and write the answers to this question type. The instructor did follow procedures to wait an average of 10 s -15 s for students to respond which did result in at least 75% of students responding to the instructor-delivered ASR question. It was anecdotally recorded by the secondary investigator that some students were writing answers to the recall ASR questions when the instructor provided the correct answer to the class resulting in those students lacking the opportunity to display their response cards. This did not seem to negatively affect results of the study, as the average class participation had high levels across both recall and recognition conditions (i.e., 88.94% and 90.7% respectively).

There were some challenges controlling all variables within an applied classroom setting subsequently resulting in some limitations to the study. Research on using response cards during lectures suggests that there should be a rule asking students to hold cards at least head high so an instructor can clearly see all students’ answers. In this study, there were no rules to hold cards at least head high because the instructor was not comfortable having this rule. As a result, all student responses were difficult to see from the front of the room resulting in the instructor moving his body in an effort to see all student responses and provide corrective feedback. This lack of rule also made it difficult to count the number of responses from the back of the room resulting in the secondary investigator having to move to the front of the room to collect
interobserver (IOA) data. Future research should incorporate this rule so all student responses are visible from the front of the room allowing the instructor to easily provide corrective feedback if needed and allow for IOA data to be collected in a less obtrusive manner.

Research on using RC’s during lectures recommends that instructors cue students to present their response cards simultaneously. Throughout this study, the instructor did not feel comfortable cueing college students to present answers simultaneously resulting in students presenting and lowering cards at different times. When cues to hold cards up simultaneously do not exist, research suggests students have an easier opportunity to cheat/looking at a classmates RC. Another effect of students raising and lowering cards at different times made IOA data difficult to capture on student participation for each learning trial. Future research should incorporate cues for students to hold cards up simultaneously in an effort to collect participation data in a less obtrusive manner and as a way to prevent cheating. An attendance roster was not used during class, as the instructor did not feel comfortable using a sign in sheet. Future studies are suggested to use an attendance sign-in sheet, or deny make up quizzes, in order to keep up with which grades to omit/include more easily.

Another limitation to the study included the instructor telling the students that both delayed and immediate quizzes were exactly the same. The study was designed to keep students blind to the procedures that they were experiencing. During the second class of the semester, the instructor told students that the same exact quiz would be delivered next class. Feedback to the instructor was provided following lecture to not tell students the same exact quiz would be delivered at the start of the next class to better measure retention. The order of questions was changed on the delayed quizzes in an effort to prevent cheating. During class three, the instructor told the students that the order of quiz questions was changed on the delayed quiz
compared to the immediate quiz and told the students again that it is still the same exact quiz. This information could have negatively impacted the results of the study as students could have just memorized their responses from the immediate quizzes to respond better on the delayed quizzes.

There were some limitations on quizzes administered due to mistakes in the delivery of the quiz. For example, the instructor told students while delivering lectures that certain questions might be a quiz question (i.e., one question from class 2, two questions from class 3, one question from class 6). Feedback was provided to the instructor following lecture to omit cueing students on which questions could be on the quiz as the statement could influence quiz scores. Other mistakes included quiz questions being thrown out due to mistakes on the quiz question (i.e., question 7 on both the immediate and delayed quiz 2, one question class 10 on 2/20). Another difference across quizzes included multiple answers being accepted (i.e., both immediate and delayed quiz 7 contained 1 question in which two answers were accepted) and some quiz answers receiving partial credit if students responded with part of the correct answer. During class three, the instructor reviewed the immediate quiz answers with students. Feedback was immediately provided to review only delayed quiz answers as reviewing immediate quiz answers could influence scores on the delayed quiz. Future researchers are urged to follow this procedure, as there was a likely increase in delayed quiz scores following review of the immediate quiz answers.

While disruptive behavior was not recorded during instruction, it was anecdotally reported that disruptive behavior did occur during some classes (e.g., texting, Facebook, internet surfing) despite using active responding procedures. Future research should collect data on disruptive behaviors when implementing this study’s procedures to determine effects on the
competing responses during lectures. This would likely require researchers to use video recording of student behavior from many viewpoints to capture all college students’ behaviors. This study did not examine the effects of question type on disruptive behavior due to the difficulty of distinguishing if talking to a peer was on/off topic, inability to see all students’ computer screens to see if they were taking notes in a Word document or using the internet to surf non-class related websites, etc. Despite the high levels of opportunities to respond, anecdotally it was recorded that the same 1-3 students refused to participate in class wide ASR questions. This may be due to the easy level of quiz questions being delivered each class. This type of study may be better suited for an instructor who is reluctant to pose questions during the delivery of instruction. The instructor in this study asked an average of 1 question per min, and 1 relevant question per 2 min. This high level of opportunities to respond during all three conditions could have negatively impacted the results of the study as the instructor provided many opportunities for students to practice the material just learned and could have been the result of most student academic quiz scores being in the 8-10 questions correct range. Although not a primary dependent variable for this study, the researcher collected frequency data for the total number of questions the instructor asked and the number of questions relevant to the class material. Appendix G lists the class number and the total number of instructor delivered questions per class and the total number of relevant questions delivered to the class. Relevant questions were those that expanded on the concepts being taught each class (i.e., introduction to behavior modification, observing and measuring behavior, graphing and measuring change, respondent behavior, reinforcement, extinction, punishment, stimulus control, prompting, shaping/chaining, functional assessment, functional non-aversive treatments, differential reinforcement, antecedent control procedures, behavioral skills training, time-out and response
cost, using positive punishment, promoting generalization). Some examples of questions that were not considered relevant to the class material included: Does anyone have a dumb phone? Why are you wearing a Tiger’s jersey? Anyone need more time on this slide? Who went to Gasparilla? Some examples of questions that were considered relevant to the class material being taught included: What is reactivity (during the class that covered observing and measuring behavior)? Any questions on the quiz? An unconditioned stimulus elicits what (during a class that covered respondent behavior)? Is this scenario an example of reinforcement or not (during a class that covered reinforcement)?

This purpose of this study was to examine the effects of question type delivered during instruction on students’ academic and participation behaviors. The instructor selected to participate in this study used high amounts of opportunities to respond in his class that may not be typical of lecture only classes. The goal of the study was to manipulate question type only, not how the instructor delivered lectures; therefore, the high levels of opportunities to respond and practice the correct answers to questions may have impacted the study’s results.

Prior to the last quiz (i.e., immediate) being delivered, the instructor told the students that he would drop the lowest quiz scores and keep only the highest five quiz scores of the semester. This statement could have served as an abolishing operation that influenced student motivation to try hard on both the last immediate and last delayed quiz (X = 7.26 and X = 7.05 respectively) as these scores were the second lowest of the semester. Also, quiz 16 was the lowest immediate (X = 5.8) and delayed (X = 5.97) quiz scores of the semester and likely due to the administration of the quiz following a week off from the semester due to spring break. While the study did experience some limitations, the instructor followed most procedures outlined in the treatment integrity checklist.
The only time treatment integrity fell below 100% was when the instructor did not provide a praise statement following an ASR question, subsequently resulting in corrective feedback. The instructor stated that he did not feel as though each question needed positive feedback to the group as a whole. Anecdotal data collected by the PI found more students responded following the positive feedback delivered by the instructor. Once the data were shared with the instructor, he began providing positive feedback to the group following each ASR question. Future researchers are encouraged to use the positive feedback to the group when incorporating ASR procedures during lecture to ensure students are actually responding during the learning trial.

Lastly, it was expected that both the instructor and students would rate the active responding procedures higher than passive lecture procedures and results from social validity data confirmed this expectation. Students rated the baseline condition the lowest in social validity measures (X = 2.19) as compared to recognition (X = 4.36) and recall (X = 3.14) conditions when asked which condition they were more likely to participate. When asked which question type was preferred, most students (n = 22) preferred the recognition questions likely due to it being easier to select the correct response during the learning trial. No students preferred the baseline condition that represented standard lecture style in a college classroom.

When college instructors are delivering lectures, especially at the introductory level, results from this study suggest that it does not matter which question type (recall, recognition) is delivered during lectures. Using active student responding procedures during lectures, while providing students with multiple opportunities to respond, has been shown to improve academics and participation in classrooms. It has been suggested that using recall questions for assessment of student learning is better than using recognition questions so students are not leaving the
assessment with false understanding (Roediger & Marsh, 2005). This phenomenon has been attributed to students’ confusing the distractors in a multiple-choice answer bank with the true answer and subsequently leaving the class with inaccurate learning of concepts. During the learning trial, it appears both types of questions produce similar results in academic scores and instructors are encouraged to determine which question type best suits the material to be taught. Our college classrooms are diverse; students come from different backgrounds, different cultures, and have different learning abilities. Tailoring instruction to meet the individual needs of our diverse student population is important to foster a learning environment that benefits all students.
References


doi:10.5210/bsi.v7i1.299


doi:10.1177/1063426609333448


doi:10.1080/1045988X.2012.682184


doi:10.1177/10983007060080020701


doi:10.1177/016264348200500302


doi:10.1207/s1532690xci2103_02

doi:10.1007/s10864-007-9056-8


doi:10.1080/1045988x.2014.919138


doi:10.1177/07419325060270050501
Appendix A: Experienced Instructor Raters Questionnaire

Date: __________     Time: __________     Class #: __________     Rater #__________

Instructions: Please circle the response that best matches your expert opinion. If you would like to provide additional feedback, a space has been provided.

<table>
<thead>
<tr>
<th>Question 1: This question reflects the learning objectives of the class:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Please rate the level of difficulty:</td>
<td>strongly agree</td>
<td>agree</td>
<td>neutral</td>
<td>disagree</td>
<td>strongly disagree</td>
</tr>
<tr>
<td>corrective feedback:</td>
<td>very easy</td>
<td>easy</td>
<td>neutral</td>
<td>difficult</td>
<td>very difficult</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Question 2: This question reflects the learning objectives of the class:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Please rate the level of difficulty:</td>
<td>strongly agree</td>
<td>agree</td>
<td>neutral</td>
<td>disagree</td>
<td>strongly disagree</td>
</tr>
<tr>
<td>corrective feedback:</td>
<td>very easy</td>
<td>easy</td>
<td>neutral</td>
<td>difficult</td>
<td>very difficult</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Question 3: This question reflects the learning objectives of the class:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Please rate the level of difficulty:</td>
<td>strongly agree</td>
<td>agree</td>
<td>neutral</td>
<td>disagree</td>
<td>strongly disagree</td>
</tr>
<tr>
<td>corrective feedback:</td>
<td>very easy</td>
<td>easy</td>
<td>neutral</td>
<td>difficult</td>
<td>very difficult</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Question 4: This question reflects the learning objectives of the class:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Please rate the level of difficulty:</td>
<td>strongly agree</td>
<td>agree</td>
<td>neutral</td>
<td>disagree</td>
<td>strongly disagree</td>
</tr>
<tr>
<td>corrective feedback:</td>
<td>very easy</td>
<td>easy</td>
<td>neutral</td>
<td>difficult</td>
<td>very difficult</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Question 5: This question reflects the learning objectives of the class:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Please rate the level of difficulty:</td>
<td>strongly agree</td>
<td>agree</td>
<td>neutral</td>
<td>disagree</td>
<td>strongly disagree</td>
</tr>
<tr>
<td>corrective feedback:</td>
<td>very easy</td>
<td>easy</td>
<td>neutral</td>
<td>difficult</td>
<td>very difficult</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Question 6: This question reflects the learning objectives of the class:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Please rate the level of difficulty:</td>
<td>strongly agree</td>
<td>agree</td>
<td>neutral</td>
<td>disagree</td>
<td>strongly disagree</td>
</tr>
<tr>
<td>corrective feedback:</td>
<td>very easy</td>
<td>easy</td>
<td>neutral</td>
<td>difficult</td>
<td>very difficult</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Question 7: This question reflects the learning objectives of the class:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Please rate the level of difficulty:</td>
<td>strongly agree</td>
<td>agree</td>
<td>neutral</td>
<td>disagree</td>
<td>strongly disagree</td>
</tr>
<tr>
<td>corrective feedback:</td>
<td>very easy</td>
<td>easy</td>
<td>neutral</td>
<td>difficult</td>
<td>very difficult</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Question 8: This question reflects the learning objectives of the class:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Please rate the level of difficulty:</td>
<td>strongly agree</td>
<td>agree</td>
<td>neutral</td>
<td>disagree</td>
<td>strongly disagree</td>
</tr>
<tr>
<td>corrective feedback:</td>
<td>very easy</td>
<td>easy</td>
<td>neutral</td>
<td>difficult</td>
<td>very difficult</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Question 9: This question reflects the learning objectives of the class:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Please rate the level of difficulty:</td>
<td>strongly agree</td>
<td>agree</td>
<td>neutral</td>
<td>disagree</td>
<td>strongly disagree</td>
</tr>
<tr>
<td>corrective feedback:</td>
<td>very easy</td>
<td>easy</td>
<td>neutral</td>
<td>difficult</td>
<td>very difficult</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Question 10: This question reflects the learning objectives of the class:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Please rate the level of difficulty:</td>
<td>strongly agree</td>
<td>agree</td>
<td>neutral</td>
<td>disagree</td>
<td>strongly disagree</td>
</tr>
<tr>
<td>corrective feedback:</td>
<td>very easy</td>
<td>easy</td>
<td>neutral</td>
<td>difficult</td>
<td>very difficult</td>
</tr>
</tbody>
</table>

Are there 10 quiz questions that contain 5 recall and 5 recognition questions? yes  no
Are there 20 ASR questions? yes  no
Appendix B: Percentage of Student Response Data Recording Sheet

Observer Initials: ____  Start Time: ______  End Time: ______  Date: ______

Intervention condition only:
Instructions: For each question, please tally the number of students who do not raise the RC during intervention.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
</tr>
</tbody>
</table>

Baseline Condition only:
Instructions: Tally the number of students who ask questions anytime during class.

How many students attended class today? ________ students
Multiply # of students in class by 20 Opportunities To Respond _______ OTR
Subtract total number of students who did not respond ________ student responders
Divide responders by OTR and multiply by 100 for percentage ________ %

(Example: 40 students in class x 20 OTR = 800 OTR – 20 non responders = 780/800 = .975 x 100 = 97.5%)
Appendix C: Treatment Integrity Data Sheet for Baseline Condition

Treatment Integrity for Baseline Condition

Date: ______

Start Time: ______   End Time: ______

Observer Initials: ____

Instructions: Please indicate if steps were performed during each instructional trial by placing a plus (+) to indicate step complete and minus (-) to indicate step not complete. An instructional trial begins with a teacher posed question and ends with a subsequent answer revealed.

1. Teacher presents 10 multiple questions to the class.
2. Teacher presents 10 short answer questions to the class.
3. Teacher reveals the answer to class.

<table>
<thead>
<tr>
<th>Instructional Interval</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Teacher MC question presented</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Teacher SA question presented</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Teacher reveals answer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix D: Treatment Integrity Data Sheet for Response Card Conditions

Treatment Integrity for Response Card Condition

Date: ______
Start Time: ______  End Time: ______
Observer Initials: ____

Instructions: Please indicate if steps were performed during each instructional trial by placing a plus (+) to indicate step complete and minus (-) to indicate step not complete. An instructional trial begins with a teacher posed question and ends with a subsequent answer reveal plus praise.

1. Teacher presents question to the class.
2. Teacher provides adequate wait time for students to use response card (e.g., enough time for ¾ of students in classroom respond).
3. Teacher reveals the answer to class.
4. Teacher provides praise statement for responses (e.g., “Great job answering everyone”).

<table>
<thead>
<tr>
<th>Instructional Interval Step</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Teacher question presented</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Teacher provided adequate wait time (e.g., ¾ of students responded)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Teacher reveals answer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Teacher provides praise statement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix E: Instructor Social Validity Questionnaire

Instructions: Please answer all the following questions with your honest expert opinion.

1. The procedures used in this study were easy to use in my classroom:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strongly disagree</td>
<td>Disagree</td>
<td>Neutral</td>
<td>Agree</td>
<td>Strongly agree</td>
</tr>
</tbody>
</table>

2. I will continue to use recall questions during this subject area:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strongly disagree</td>
<td>Disagree</td>
<td>Neutral</td>
<td>Agree</td>
<td>Strongly agree</td>
</tr>
</tbody>
</table>

3. I will continue to use recognition questions during this subject area:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strongly disagree</td>
<td>Disagree</td>
<td>Neutral</td>
<td>Agree</td>
<td>Strongly agree</td>
</tr>
</tbody>
</table>

4. I enjoyed using white boards in my class:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strongly disagree</td>
<td>Disagree</td>
<td>Neutral</td>
<td>Agree</td>
<td>Strongly agree</td>
</tr>
</tbody>
</table>

5. Students were more engaged when I used white boards:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strongly disagree</td>
<td>Disagree</td>
<td>Neutral</td>
<td>Agree</td>
<td>Strongly agree</td>
</tr>
</tbody>
</table>

6. When recall questions were used, students answered more questions correctly:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strongly disagree</td>
<td>Disagree</td>
<td>Neutral</td>
<td>Agree</td>
<td>Strongly agree</td>
</tr>
</tbody>
</table>

7. When recognition questions were used, students answered more questions correctly:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strongly disagree</td>
<td>Disagree</td>
<td>Neutral</td>
<td>Agree</td>
<td>Strongly agree</td>
</tr>
</tbody>
</table>

8. What was the best part of implementing this intervention in your classroom?

9. What was the worse part of implementing this intervention in your classroom?

10. What could be done differently to have instructors implement these procedures?

11. What grade would you give your experience with the intervention: Circle one.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Very useful instructional approach</td>
<td>Useful instructional approach</td>
<td>Neutral</td>
<td>Not very useful instructional approach</td>
<td>Not useful instructional approach</td>
</tr>
</tbody>
</table>

9b. If a grade of C or lower was reported, please provide us with feedback on what made this instructional approach not useful in the classroom:

12. Which question type did you prefer to use, recall (i.e., short answer) or recognition (i.e., multiple choice)?
Appendix F: Student Social Validity Questionnaire

Date: ____________  Time: ______

1. Which type of questions during lecture did you prefer, short answer or multiple-choice?

2. Why did you prefer the above question type?

3. What did you like best about the intervention?

4. What did you like least about the intervention?

Please rate the following questions:

5. I participated more during classes when multiple-choice questions were posed:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly disagree</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>Agree</td>
<td>Strongly agree</td>
</tr>
</tbody>
</table>

6. I participated more during classes when short answer questions were posed:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly disagree</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>Agree</td>
<td>Strongly agree</td>
</tr>
</tbody>
</table>

7. I participated more during classes when no questions were posed:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly disagree</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>Agree</td>
<td>Strongly agree</td>
</tr>
</tbody>
</table>

Additional feedback:
**Appendix G: Instructor Delivered Questions During Each Class**

<table>
<thead>
<tr>
<th>Class Number</th>
<th>Total Instructor Delivered Questions</th>
<th>Total Instructor Delivered Questions that were Relevant to the Class Objectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>56</td>
<td>39</td>
</tr>
<tr>
<td>2</td>
<td>43</td>
<td>25</td>
</tr>
<tr>
<td>3</td>
<td>54</td>
<td>32</td>
</tr>
<tr>
<td>4</td>
<td>55</td>
<td>39</td>
</tr>
<tr>
<td>5</td>
<td>95</td>
<td>89</td>
</tr>
<tr>
<td>6</td>
<td>65</td>
<td>46</td>
</tr>
<tr>
<td>7</td>
<td>49</td>
<td>20</td>
</tr>
<tr>
<td>8</td>
<td>36</td>
<td>25</td>
</tr>
<tr>
<td>9</td>
<td>50</td>
<td>32</td>
</tr>
<tr>
<td>10</td>
<td>30</td>
<td>25</td>
</tr>
<tr>
<td>11</td>
<td>40</td>
<td>24</td>
</tr>
<tr>
<td>12</td>
<td>38</td>
<td>27</td>
</tr>
<tr>
<td>13</td>
<td>35</td>
<td>26</td>
</tr>
<tr>
<td>14</td>
<td>39</td>
<td>21</td>
</tr>
<tr>
<td>15</td>
<td>43</td>
<td>34</td>
</tr>
<tr>
<td>16</td>
<td>47</td>
<td>32</td>
</tr>
<tr>
<td>17</td>
<td>34</td>
<td>24</td>
</tr>
<tr>
<td>18</td>
<td>39</td>
<td>30</td>
</tr>
<tr>
<td>19</td>
<td>39</td>
<td>30</td>
</tr>
<tr>
<td>20</td>
<td>32</td>
<td>23</td>
</tr>
<tr>
<td>21</td>
<td>33</td>
<td>20</td>
</tr>
</tbody>
</table>
Appendix H: IRB Exempt Letter

12/5/2017

Leslie Singer
ABA-Applied Behavior Analysis
Child and Family Studies
13301 Bruce B. Downs Blvd., MHC 2321
Tampa, FL 33612

RE: Exempt Certification
IRB#: Pro00031706
Title: Effects of Interspersing Recall versus Recognition Questions with Response Cards During Lectures on Students’ Academic and Participation Behaviors in a College Classroom

Dear Ms. Singer,

On 12/4/2017, the Institutional Review Board (IRB) determined that your research meets criteria for exemption from the federal regulations as outlined by 45CFR46.101(b):

(1) Research conducted in established or commonly accepted educational settings, involving normal educational practices, such as (i) research on regular and special education instructional strategies, or (ii) research on the effectiveness of or the comparison among instructional techniques, curricula, or classroom management methods.

As the principal investigator for this study, it is your responsibility to ensure that this research is conducted as outlined in your application and consistent with the ethical principles outlined in the Belmont Report and with USF HRPP policies and procedures.

Please note, as per USF HRPP Policy, once the Exempt determination is made, the application is closed in ARC. Any proposed or anticipated changes to the study design that was previously declared exempt from IRB review must be submitted to the IRB as a new study prior to initiation of the change. However, administrative changes, including changes in research personnel, do not warrant an amendment or new application.

Given the determination of exemption, this application is being closed in ARC. This does not limit your ability to conduct your research project.

We appreciate your dedication to the ethical conduct of human subject research at the University of South Florida and your continued commitment to human research protections. If you have