Winter 2018

Emergence of Stimulus Equivalence and Topography-Based Responding Following Lecture Instruction

Justin Dale Krzmarzick
Central Washington University, justinkrzmarzick@gmail.com

EMERGENCE OF STIMULUS EQUIVALENCE AND TOPOGRAPHY-BASED RESPONDING FOLLOWING LECTURE INSTRUCTION.

A Thesis

Presented to

The Graduate Faculty

Central Washington University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

Applied Behavior Analysis

---

by

Justin Dale Krzmarzick

March, 2018

CENTRAL WASHINGTON UNIVERSITY

Graduate Studies

We hereby approve the thesis of

Justin Dale Krzmarzick

Candidate for the degree of Master of Science

APPROVED FOR THE GRADUATE FACULTY

______________ ______________________________
Dr. Sadie Lovett, Committee Chair

______________ ______________________________
Dr. Susan Lonborg

______________ ______________________________
Dr. Richard Marsicano

______________ ______________________________
Dean of Graduate Studies

ABSTRACT

EMERGENCE OF STIMULUS EQUIVALENCE AND TOPOGRAPHY-BASED RESPONDING FOLLOWING LECTURE INSTRUCTION.

By

Justin Krzmarzick

March, 2018

The present study examined the emergence of stimulus equivalence using both selection-based and topography-based tests following a lecture or control condition. This study also evaluated generalization to novel stimuli in both selection-based and topography-based response formats, and evaluated the social validity of the instructional procedure. Twenty undergraduate students who were at least 18 years of age were assigned to a lecture or control condition. Participants in the lecture condition were exposed to a lecture on the topic of generalization. Participants in the control condition watched the video *Martin Seligman: The New Era of Positive Psychology* that did not relate to the content of the tests. Participants were given multiple choice pre- and post-tests, intraverbal pre- and post-tests, and emergent relation pre- and post-tests. When selection-based tests were compared to topography-based tests, neither group performed significantly better on one type of test or the other. As for generalization, both the lecture and control groups showed an increase in correct responding. Since both groups had an increase in correct responding, the generalization that occurred was likely due to a testing effect and not the specific condition that the participants were exposed. Participants in this study moderately preferred the instructional format.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONTINUED)

Chapter Page

CHAPTER I

INTRODUCTION

Online courses are becoming more prevalent than ever in our education system. More than 6.1 million college students were taking at least one online course during the fall of 2010, and 31% of all higher education students are taking an online course (Allen, & Seaman, 2011). Due to tight school budgets and a push for more students to graduate, schools have been looking for alternatives to the traditional classroom format (Gabriel, 2011). Straying from the traditional classroom style has sparked debate as to whether these online courses are as effective as traditional classroom courses (Gabriel, 2011). Before assessing the effectiveness of online courses, the elements that are necessary to have an effective education system should be discussed.

According to Austin (1999), to provide the highest-level education possible, the education college students are receiving should be “effective, efficient, and empirically validated” (pg. 449). These principles are central ideas in behavior analysis (Baer, Wolf, & Risley, 1968) and are seen in the work of Skinner (1984), who based his analysis of improving education on finding ways for students to learn faster with the same amount of effort as traditional classes. Skinner (1984) believed that if students could learn faster with the same amount of effort we would solve many problems associated with our poor education system (Skinner, 1984). This solution is attractive because it would not require students to spend any additional time in the classroom; we would simply change the way concepts are taught.

As stated by Skinner (1984), the first step in improving the rate at which students learn is being clear about what is being taught. Skinner found that many of his critics believed that, although he could teach students how to correctly solve a wide range of academic tasks, they did not believe that the students knew the concepts. For example, using teaching machines Skinner could teach students how to solve algebra problems. Skinner's teaching machines were automated devices that presented curriculum to students, and gave immediate consequences, which would allow the students to immediately check their mastery of the material. Some educators criticized Skinner's accomplishments by stating that, although the students could solve the problems, they did not "know" algebra. Skinner argued that if students develop "intuition" related to a subject area when they are provided with high-quality instruction for basic concepts, it would lead to greater proficiency in that area. For example, learning many algebra problems will result in it being easier to solve new problems, which to the average person may appear to be "intuition" (Skinner, 1984).

The second solution that Skinner proposed was that students should be taught "first things first." What this means is that educators too often want to teach concepts such "creativity" and "excellence," which are not basic skills but something acquired once the learner is taught several variations of skills. These variations of skills are what comprise "creativity" or "excellence," and to reach these goals it is important that clear steps are provided on how to progress there. For example, if you wanted someone who knew nothing about cars to "innovate" a new vehicle, you would first need to teach the person the basics mechanics of how a car works, and how different cars are put together. Once the person understands the variations of different variables he or she would be able to combine them in a new and "innovative" way (Skinner, 1984).

The third and fourth solutions provided by Skinner relate to the rate of instruction and presentation of the instructional material. It was suggested that individual learners progress at their own rate instead of having all learners progress together. Having students progress together causes learners who master skills quickly to have to wait for other students to catch up and those who need more time to fall even further behind. Finally, academic subjects can be programmed to promote learner engagement with the subject matter, prompt correct responses as needed, and then fade the prompting until the learner is able to answer the question independently. When the learner can answer the question on his or her own, the reinforcing consequences of working independently helps sustain the behavior. Effective programming also helps tackle the problem of motivation. Instead of students learning to avoid the punishment associated with not answering correctly, a more effective way to teach is to program positive reinforcers to be contingent on the target academic behavior (Skinner, 1984).

There have been several behavior analytic instructional methodologies inspired from Skinner’s analysis. This study will evaluate a phenomenon called stimulus equivalence in relation to a traditional lecture teaching format as compared to a control condition for undergraduate students. In the next section, there will be an overview of several behavior analytic instructional methods, and the specific behavioral processes used in instructional method for this study.

CHAPTER II

LITERATURE REVIEW

Behavior Analytic Instruction

Within the field of behavior analysis, there are several instructional methods that have been found to be successful with college students, including precision teaching, active responding, interteaching, personalized system of instruction (PSI), and computer-aided personalized system of instruction (CAPSI). One of the earliest behavior analytic instructional methods developed was precision teaching, which was based on four principles derived from the works of Skinner (Austin, 1999). The first principle is that the learner knows best. What this means is that the behavior of the learner can tell us more about how he is learning than any other source of information. If a student is picking up the material quickly then the teacher is teaching correctly, but if the learner is not picking up the material quickly, then the teacher's program is not on the right track. The next principle is using the rate of response as the standard unit of measurement, which differs from how teaching is traditionally conducted where the emphasis is placed on accuracy. The third principle is emphasizing observable behavior and using direct and frequent measurement by implementing frequent, short, timed trials. The last principle of precision teaching is that the data collected from the trials should be graphed to maximize feedback and motivation. Graphs showing the celeration of performance over time, either increasing or decreasing are used to gauge how effective the instructional method is at teaching the learner academic concepts. Analyzing these graphs is important when deciding whether changes need to be made to the instructional method (Austin, 1999).

Another behavior analytic instructional method used to aid student learning is active student responding. Active student responding is based on the idea that students learn by doing (Austin, 2000). If teachers wish to maximize the amount of student learning, they need to set up contingencies that maximize the amount of active participation that occurs in the classroom. Three ways in which active student responding can be improved are through guided notes, response cards, and choral responding. Guided notes involve providing lecture notes to students in advance with essential information omitted. Students are then able to actively respond by filling in the missing information while listening to the lecture. Guided notes help students retain information just as regular note taking does but with the added benefit of having fewer errors that traditional note taking. Response cards are a tool to help promote active student responding by having students raise a card with their answer on it once a teacher poses a question. Response cards can be categorized as write in response cards or color response cards. With write in response cards answers are written in and shown to the teacher, while with color response cards students can be hold up a color card corresponding to the answer. Choral responding is another type of active student responding where the entire class answers a teacher's question simultaneously and the teacher can give feedback to the class as a whole. This method gives teachers a way to quickly assess several student responses (Austin, 2000).

Another behavior analytic instructional method is the PSI (Austin, 1999). PSI contains five main components: self-pacing, unit mastery requirement, using lectures and demonstration for motivational purposes, using written communication, and using undergraduates as proctors (Austin, 1999). Self-pacing is when the students can take exams when they feel that they have mastered the material. The tests are always available, which allows students to work at their own pace. Unit mastery requirement means that for a learner to advance to the next unit the learner needs to achieve a minimum test score. There is no limit to the number of times that a learner can take the tests, and there is no penalty for receiving a subpar score. The third component of PSI uses lectures and demonstrations solely for motivational purposes and to provide clarification on the material. The fourth component of PSI focuses on written communication. All the important information that is needed to master the material is provided in the form of written communication. The final component of PSI uses undergraduates as proctors, which allows for the continuous grading needed to make PSI a viable instructional arrangement.

Interteaching is another method based on behavior analytic methods that addresses some of the weaknesses associated with PSI. There are several components that comprise this method, including providing students with a preparation guide to complete before class, arranging small group discussions during class, writing down interesting or difficult information following class discussions, providing brief lectures based on feedback from student discussions, and administering frequent tests. This method can be most effective when contingencies are arranged to promote quality in-class discussion and work (Bernstein & Chase, 2013).

Another instructional method that is based on PSI is CAPSI. CAPSI was developed by Pear and associates at the University of Manitoba (Bernstein & Chase, 2013). It includes the same elements as PSI, but uses computers as a means of administering the material and providing feedback. Like PSI, CAPSI includes small units of study, study guides to direct learning in each unit, self-pacing, the requirement of demonstration of mastery, review of mastery by instructors or peer reviewers, and feedback on each test question (Bernstein & Chase, 2013). CAPSI has been found to enhance students' progress over traditional lecture-based teaching methods (Pear & Novak, 1996). In addition to CAPSI, researchers have begun to develop other computer-based instructional methods based on the principles of behavior analysis, including an examination of the promotion of stimulus equivalence as part of instruction.

Stimulus Equivalence

Stimulus equivalence is said to have emerged when accurate untrained responding to stimulus-stimulus relations occurs after the reinforcement of responding to other stimulus-stimulus relations (Cooper, Heron, & Heward, 2008). For a class of stimuli to be considered equivalent, it must have the properties of symmetry, reflexivity and transitivity. Reflexivity is when, without any prior history of reinforcement or training, a learner can accurately select a comparison stimulus in the presence of an identical sample stimulus. This means that the learner would respond to stimulus A by selecting an identical stimulus A from a choice of options (A=A). An example of this is a learner being able to match a picture of a bird (the sample stimulus) to an identical picture of the bird (the comparison stimulus). Symmetry occurs when a learner can respond accurately to the reversal of a trained stimulus-stimulus relation without a previous reinforcement history for doing so. For example, if you trained that the written word "bird" (sample stimulus) matches a picture of a bird (comparison stimulus), the learner would be able to correctly respond when the order of presentation is revered. This means when shown the picture of a bird (sample stimulus), the learner will select the written word "bird" (comparison stimulus; A=B entails B=A). Transitivity occurs when two prior stimulus-stimulus relations have been directly trained. For example, if you trained that the written word "bird" is equal to a picture of a bird, and a picture of a bird is equal to a description of a bird then the learner would correctly respond to questions presenting the written word "bird" is and a description of a bird (A=B, B=C, so A=C).

The type of responding described above is a conditional discrimination, and the teaching and testing format is called matching-to-sample or MTS when related to stimulus equivalence (Green & Saunders, 1998). A single MTS trial consists of a sample stimulus being presented first. Next, several comparison stimuli are presented. If the comparison stimulus that is equivalent to the sample stimulus is selected, in the learner receives reinforcement. Selection of the other comparison stimuli does not result in reinforcement (Green & Saunders, 1998). An example of a conditional discrimination would be the word bird presented as a sample stimulus. Next a picture of a bird would be presented as a comparison stimulus, which when selected would result in reinforcement. Along with the picture of a bird, a picture of a cat and a picture of a dog would be presented as comparison stimuli, which if selected would result in either extinction or an error correction procedure. The MTS instructional arrangement has been used to promote the emergence of stimulus equivalence to teach several different concepts within higher education by a variety of researchers (Fienup, Covey, & Critchfield, 2010; Ninness, Rumph, McCuller, Harrison, Ford, & Ninness, 2005; Fields, Travis, Roy, Yadlovker, De

Aguiar-Rocha, & Sturmey, 2009).

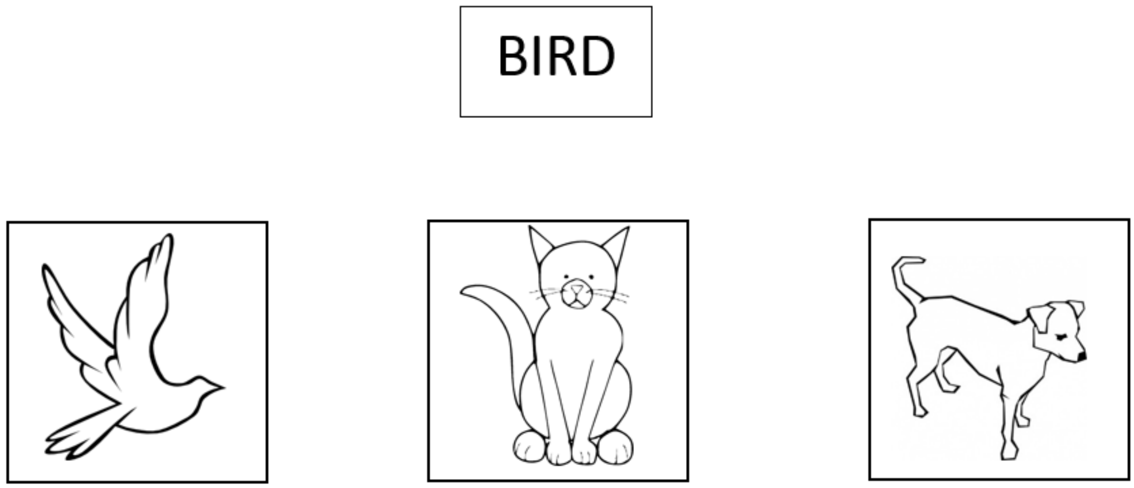


Figure 1. An example of a conditional discrimination or MTS trial.

Stimulus Equivalence in Higher Education

The stimulus equivalence paradigm has been used to teach a wide range of skills such as reading (De Rose, De Souza, & Hanna, 1996), establishing derived mands in adults with severe intellectual disabilities (Rosales & Rehfeldt, 2007), statistical interactions (Fields et al., 2009), and single-subject designs (Lovett, Rehfeldt, Garcia, & Dunning, 2011). Due to the success of previous stimulus equivalence studies, researchers have begun to extend this research to higher education to identify more effective and efficient instructional methods. There have been several applications in teaching students in higher education, such as teaching statistical interactions (Fields et al., 2009), brain-behavior relations (Fienup, Covey, & Critchfield, 2010), and single-subject research design (Lovett et al., 2011).

A study by Fields, Travis, Roy, Yadlovker, De Aguiar-Rocha and Sturmey (2009) used stimulus equivalence to teach statistical interactions to college students. Fields et al.

(2009) used a pretest-posttest research design with control and experimental groups. Participants in both groups were given a pencil-and-paper pretest and posttest. Between the pre- and posttests, the control group was given a break from the experiment while the equivalence group completed a computer program designed to promote the emergence of stimulus equivalence. The computer-based stimulus equivalence training consisted of four four-member equivalence classes. The A stimuli were graphs depicting each of the four basic statistical interactions. A stimulus class is identified by the number of stimuli in the class. The B stimuli were vignettes that described clinical situations that corresponded with the four types of statistical interactions. The C stimuli were the definitions of the four types of statistical interactions. The D stimuli were definitions of the four types of statistical interactions. The computer-based stimulus equivalence training employed a MTS format to teach the initial stimulus-stimulus relations. Equivalence was observed emerge in experimental group following the MTS instruction. When students in the experimental group were compared with the control group, the students in the experimental group scored higher on the paper-and-pencil posttest. The results of this study confirm that computer-based stimulus equivalence protocols can be used to teach academic concepts in higher education.

Another study that used stimulus equivalence to teach academic concepts to students in higher education was conducted by Lovett, Rehfeldt, Garcia, and Dunning (2011). In this study, researchers sought to teach undergraduate college students enrolled in a research methods class single-subject designs using a computer-based stimulus equivalence protocol and compare it with traditional lecture-based teaching. Participants were assigned to either the computer-based stimulus equivalence group or the traditional lecture group. All participants were given a selection-based pretest and posttest, as well as a multiple-choice paper-and-pencil posttest. The participants in the lecture group were exposed to a 56-min lecture accompanied by a PowerPoint presentation teaching single-subject designs. The computer-based stimulus equivalence protocol, much like the Fields et al. (2009) study, used a MTS procedure. There were four stimulus classes containing four stimuli each that were taught to the participants. The A stimuli in this study were the four types of fundamental single-subject design. The B stimuli were the definitions of those four-fundamental single-subject designs. The C stimuli were graphs showing the implementation of the different types of single subject design. The D stimuli were vignettes describing a clinical situation, with each vignette describing a scenario that would require use of a specific type of single-subject design. Participants in this group were additionally given tests for transitivity, equivalence, and a test for generalization (Lovett et al., 2011). The social validity assessment was included because, according to Skinner (1968), students should prefer the active responding of the computer-based protocol to the lecture. However, the study conducted by Lovett et al (2011) found that participants had no preference towards either teaching method.

There was also a multiple-choice test used to assess generalization in this study. The first test assessed for relations between the names of the types of single-subject design and novel graphs corresponding to the types of single subject design. The second selection-based test assessed for relations between the names of the types of single-subject design and novel clinical vignettes. The study also included a topography-based test to assess whether participants could orally name the stimuli. This topography-based test consisted of the experimenter showing the participant flash cards with either a novel vignette or a picture of a novel graph corresponding to one of the basic types of single-subject design. The experimenter then asked the participants “What design is this?” and they were given 10 seconds to respond. This study found that 3 of the 4 participants who took the topography-based test answered correctly to the trained and novel stimuli. However, on the selection-based multiple-choice test, there was not a significant difference between the equivalence and lecture groups. The equivalence group had an average increase in scores from pre-to posttest of 2.9, and the lecture group had an average increase in scores from pre-to posttest of 2.4 points. Due to the varying result of the selection-based vs. topography-based tests, these data suggest there may be important differences in response form.

Selection-Based vs. Topography-Based Responding

Stimulus equivalence has been shown to be effective at teaching academic concepts but there are several different ways in which correct responding can be assessed. One way that learners can respond to questions is a selection-based response. Sundberg and Sundberg (1990) define selection-based responding as pointing, touching, looking or in some way indicating a particular stimulus (Sundberg, & Sundberg, 1990). An example of this would be filling in a bubble on a multiple-choice exam. This can be contrasted with topography-based responding where each different response will have a unique form. Topography-based responding refers to the physical response that a behavior takes and distinguishes it from other verbal responses (Sundberg, & Sundberg, 1990). For example, a learner could write an answer to a math problem, the writing would be the topography or form of the behavior. If a student verbally emitted the answer to the math problem, then speaking would be the topography of the behavior. With topography-based responding there is much more variation due to the many forms that the behavior can take.

Sundberg and Sundberg (1990) compared these two types of responding to determine which would result in faster acquisition, higher accuracy, generality, maintenance, spontaneous usage and formation of equivalence classes for nonverbal individuals. In this study, the researchers had four nonverbal individuals and taught tact (name), and intraverbal relations using either selection-based or topography-based responding. Each participant was taught 2 of the 3 relations for one paradigm (i.e. selection-based tacts and intraverbals) then they were tested for the emergent relation. Next, the participant was taught 2 of the 3 relations for the other paradigm (i.e. topography-based tacts and intraverbals) then tested for the emergent relation (Sundberg, & Sundberg, 1990). Their research found significant differences in relation to equivalence class formation. Results of this study showed that selection-based responses required more training and the percentage of correct responses emitted were lower. This research showed that participants could be taught topography-based responding more rapidly/efficiently compared to selection-based responding, which demonstrated that topography-based responding may be more desirable to teach than selection-based.

A study by Polson and Parsons (2000) looked specifically at selection-based versus topography-based responding with undergraduate students and how it applies to stimulus equivalence. In their first experiment, the researchers taught 7 participants how to respond to French words by selecting (selection-based) their English counterparts and to respond to other French words by typing out their English counterparts (topography-based). Equivalence was tested for half of the items using selection-based responding and the other half using topography-based responding. Their second experiment aimed to see if equivalence could be observed through repeated no reinforced testing. Emerge of relations from repeated non-reinforced testing is known as delayed emergence (Polson, & Parsons 2000). For this experiment 5 subjects were paid $8 each to go through the experiment. The procedure for experiment 2 was identical to experiment 1 except for 2 procedural differences. One of the procedural differences in experiment 2 was that instructions were reworded to make them more clear and friendly (Polson, & Parsons 2000). Another difference was that the pretraining was duration was shorted the 10 minutes from 15 minutes to reduce boredom and save time so that the participants could spend more time on other phases of the experiment. This experiment found better results for topography-based symmetry when items were trained from English to French as opposed to French to English. In experiment 3 researchers used English rather than French words as the stimuli for trained relations. Experiment 3 was identical to experiment 2 except English words were presented as the sample stimuli and participants had to either select or type the corresponding French word. The main results of these studies were that symmetry emerged more reliably with selection-based responding than with topography-based responding. The researchers also found that participants increased word accuracy when they were required to write English words rather than French words. The results of this study contrast the results of the Sundberg and Sundberg (1990) study, which found that topography-based responses required less training than selection-based responses.

Evaluations of selection-based and topography-based responding have been conducted as part of studies examining stimulus equivalence in higher education as well.

24The study by Fields et al. (2009), which taught statistical interactions, also included selection-based training and a selection-based paper-and-pencil test. On the paper-and-pencil pre- and posttests the participants were given 24 items and instructed to select the correct answer from four multiple-choice options (a, b, c, and d). During the training portion of the experiment, the participants used the MTS method to teach stimulus-stimulus relations. Participants in the experimental condition scored 37% better of their posttest than they did on their pretest, while the control group only had a 2% increase. No topography-based responding was used in this experiment. To assess the differences between selection-based and topography-based responding, the study by Lovett et al. (2011) included tests for both types of responding. To assess section-based responding, the researchers used a MTS format. For topography-based responding the researchers provided a tact test where a researcher would hold up a flash card with a novel vignette or graph on it and "What design is this?" The participants would then have to verbally which research design corresponded with the card. For the topography-based test 3 of the 4 participants could answer correctly to trained and novel stimuli while the selection-based test found not significant difference between the equivalence and lecture groups.

Generalization of Stimulus Equivalence Classes

Another important question when considering the utility of stimulus equivalence in an educational environment is if what is being taught will be generalized beyond the instructional examples. Some studies have shown that once stimulus relations are being trained there is a varying amount of accuracy when it comes to generalization of those responses (Lovett et al., 2011). In a study by Lynch and Cuvo (1995), the researchers examined stimulus equivalence to teach fractions and decimal relations to 5th and 6th

graders. This study was conducted because many school age individuals have difficulty with math concepts, which may be due to educators teaching math strategies instead of math concepts. Critics argue that teaching math strategies only teaches rote memorization instead of comprehension. Although stimulus equivalence classes emerged with participants, there were mixed results as to whether generalization occurred to novel examples. The authors proposed several reasons why generalization did not occur with some participants, such as lack of exposure to the posttest and the emergence of incorrect relations. This provides reasons to test for generalization because if participants are not able to generalize than the effectiveness of this procedure will be severely limited.

In the study by Fields et al. (2009), which taught statistical interactions using stimulus equivalence, participants could generalize what they had learned in the stimulus equivalence protocol to novel examples with a 37% increase in correct responding. The test for generalization was presented with novel exemplars and novel formats. This is significant when compared with only a 2% increase in the control group, which provides evidence that more exposure to the test alone does not account for the increase in correct responding. An important distinction to make in this study is that not only did participants in the equivalence group score higher on the selection-based posttest compared to participants in the control group, but they did so using novel examples in the posttest. This finding shows that equivalence relations emerged and accurate responding to emergent relations generalized to novel stimuli.

Unlike the study conducted by Fields et al. (2009), the study previously mentioned by Lovett et al. (2011) found different results. In the study by Fields et al. (2009) the researchers found that participants could form equivalence classes and generalize to novel stimuli. This contrasts the study by Lovett et al. (2011) where the researchers found that while generalization occurred with some stimuli, it was not observed with others. This study also used selection-based tests and a topography-based tact test conducted using flash cards, the results of the tact test showed that three out of four participants could establish novel graph to design name generalized relations, and two out of four were able to establish novel vignette to design name. These results are interesting because the graphs and vignettes as discriminative stimuli are very different, with the graphs being pectoral and the vignettes being textual. The results show that more research needs to be conducted to find the relationships between stimulus equivalence and generalization, and selection-based and topography-based responding and stimulus equivalence. However, Lovett et al. (2011) only evaluated topography-based generalization responses for the group exposed to stimulus equivalence training. Participants in the lecture condition did not receive this test for generalization.

The purpose of the current study is to evaluate the emergence of stimulus equivalence following lecture instruction using both selection-based and topography-based tests with college age students. Participants in this study will have to match stimuli and read and interpret clinical vignettes. This study will also use a computer-based format for testing which may have implications for the utility of its use in online instruction. There are three main purposes of the present study.

1. Examine the emergence of stimulus equivalence using both selection-based and topography-based tests following a lecture or control condition.
2. Evaluate generalization to novel stimuli in both selection-based and topography-based response formats.

3. Evaluate social validity of instructional procedure.

CHAPTER III

METHOD

Participants, Setting, Apparatus

Participants were 20 undergraduate students who were at least 18 years of age. Participants were recruited using Sona at Central Washington University, and received extra credit in a psychology course for participation along with $10 cash. To confirm that individuals participating in this study were at least 18 years of age, the informed consent document included the following sentence, “By signing this document I affirm that I am at least 18 years of age”. Sessions were conducted on a desktop computer in the Psychology building at Central Washington University. At the start of the study, participants were asked to remove potential distractions, such as food, music, phones, and homework. Removing distractions was stressed to minimize the likelihood that participants were disrupted during the study.

Chris Buchanan, a computer engineering staff member at Central Washington University, created the program used in this study. The program was created using a mixture of high-level computer languages including C++, Object Pascal, and ANSI C, with content provided by Justin Krzmarzick. The program was considered “stand-alone.” It did not rely on Internet connectivity and was therefore relatively immune from remote security vulnerabilities. A demo program was reviewed by Justin Krzmarzick prior to implementing the computer program with study participants to maintain fidelity.

Equivalence Stimuli

Three stimulus classes, which contained five stimuli each, were presented during the lecture. The stimuli were representative of the three types of generalization commonly taught in applied behavior analytic coursework, and they included the name of each type of generalization, the definition of each type of generalization, and vignettes describing an application of generalization. All stimuli were developed using a graduate-level applied behavior analysis textbook (Cooper, Heron, & Heward, 2008). The A stimuli were the names of the three types of generalization: setting/situation generalization, response generalization, and response maintenance. The B stimuli were definitions corresponding to the three types of generalization. The C stimuli were written vignettes that describe scenarios that represented each type of generalization. There were three different C stimuli used in this experiment, C, C', and C". The reason three types of C stimuli were included is because in previous research (Lovett et al. 2011), researchers suggested that multiple exemplars may help improve generalization to novel stimuli. The same basic scenarios were used for each vignette (C, C', and C'') for all three types of generalization. All the written vignettes were designed to have similar length and have the same content with the only difference being the type of generalization described in the vignette. The stimuli that were used in this study are presented in Appendix A. These stimuli were reviewed by a professor fluent in applied behavior analysis to ensure they accurately represented the various types of generalization.

General Procedure

The design used was a 2x2 factorial design. There were two between-subject factors, which were the lecture and control conditions, and there were two repeated-measure factors, which were the pre-test and post-test evaluations. Before the participants started their condition, they were given a pre-session checklist telling them to turn off their cell phones and put food/drink away (Appendix B). Paper and pencil were provided for the participants in the lecture and control condition to take notes during the study. Participants in the lecture group viewed a video lecture on the topic of generalization, and participants in the control group viewed a video covering a topic that is not directly related to this study. Participants were randomly assigned to one of the two groups after providing informed consent. Ten participants were assigned to each group. The dependent measures for this experiment were the multiple-choice test, a fill-in-the-blank intraverbal test, and an equivalence test evaluating emergent relations. The lecture group also completed a survey inquiring about their opinion of the utility of the instructional method to which they were exposed.

Multiple Choice Pre- and Post-test

The multiple-choice test included 9 multiple choice questions related to the topic of generalization. Three questions tested definition-name (B-A) relations and three questions tested the vignette-name (C-A) relations. Three additional questions evaluated novel vignette-name (Cg,,-A) relations to evaluate generalization to novel exemplars. The test was presented via a computer program as a pre-test at the start of the experiment following informed consent and again as a post-test at the end of the experiment. Each item on the test had a written question centered at the top of the screen. Under the question there were three response options with a radio button corresponding to each response. At the bottom, right of the screen there was a “Next” button that allowed the participant to advance to the next question. The questions were presented in random order and no feedback was provided following the responses. The test can be viewed in Appendix C. The test started with the following instructional statement:

> The next part of the study will ask you to answer nine questions. Please read each question carefully. Once the question has been read, select the button that best corresponds with the correct answer. When finished press the “Next” button to proceed. You will not receive feedback following your response, but please do your best.

Intraverbal Pre- and Post-test

The intraverbal test consisted of 15 fill-in-the-blank questions and evaluated the emergence of topography-based responding in contrast to the selection-based responding evaluated using the multiple-choice test. Three questions evaluated the definition-name (B-A) relations, three questions evaluated the vignette-name relations (C’-A), three questions evaluated name-vignette relations (A-C), and three questions evaluated the name-definition (A-B) relations. Three additional questions evaluated novel vignette-name (Cg’-A) relations to evaluate generalization to novel stimulus exemplars. The definition-name (B-A) and vignette-name (C-A) questions were identical to those in the multiple-choice test. The A-B, and A-C relations required lengthier responses than the B-A or C-A relations. The Cg’-A questions included novel vignettes that were not presented during any of the instructional conditions or in the multiple-choice test. These test items were presented in random order and feedback was not provided following responses. The participants read each question, typed an answer then clicked on the “Next” button to continue. Test question can be found in Appendix D. A scoring used for the intraverbal test can be found in Appendix E. Before this test began participants were shown the following instructions:

The next part of the study will ask you to answer 15 questions. Please read the question at the center of the screen. Then, type an answer in the text box that you believe best answers the question. You will need to hit the “Next” button to move on to the next question. You will not receive feedback following your response, but please do your best.

Emergent Relations Pre- and Post-test

The emergent relations test evaluated equivalence relations (B-C’’ and C’’-B) and generalization (C_g-A). This test consisted of 15 questions. There were three questions testing definition-vignette (B-C”), three questions that tested vignette-definition (C’’-B), and three questions that tested novel vignette-name (Cg-A). See Appendix A for the emergent relations novel vignette-name example. There were three questions evaluating (B-A) and three questions evaluating (C’’-A). Each individual relation (e.g., B1-C”1) was presented once during this test, and no feedback was provided following responses. This test was programmed to present the questions in random order. Questions were presented using a matching-to-sample arrangement with a sample stimulus presented at the top center of the screen and three comparison stimuli were presented below the sample stimulus. The participants were required to click on a comparison stimulus to respond. Clicking on a comparison stimulus caused the program to advance to the next question. The pre- and post-tests were given at the start of the experiment. Each individual relation was presented one time to have this test more closely resemble a test that a student may see in their actual classes and to keep the participation time to a reasonable length. Before the participants began this test, they were shown the following instructions:

In this part of the study, you will be presented with 15 questions. There will be an image with text presented at the top center of the screen and three images with text presented in a row at the bottom of the screen. Your job is to identify which image at the bottom of the screen that goes with the image at the top of the screen. Click on one of the images at the bottom of the screen to respond. Clicking on an image will cause the program to advance to the next question. You will not receive feedback following your response, but please do your best.

Social Validity Survey

A questionnaire was modeled after the survey used by Lovett et al. (2011) was used to assess participant opinion on the teaching procedure to which they were exposed. The survey included four questions that were rated using a 7-point Likert scale with higher ratings indicating a more positive opinion of the instructional method. Questions asked about the participant's confidence in his or her knowledge of generalization, how much he or she would prefer to be taught using that instructional method, and his or her opinion on the time commitment required for instruction. Two additional questions inquired about the participant's computer skills were given to the control group. The social validity survey can be found in Appendix F. Before the participants began this test, they were shown the following instructions:

In this part of the study, you will be presented with 15 questions.

Lecture Group

After the initial pre-tests, the participants were shown a video of a lecture on the concepts of generalization. The video was approximately 9:08 minutes long with a PowerPoint slide show that provides examples of stimulus generalization. The lecture was divided into three sections covering the three types of generalization: response generalization, response maintenance, and setting/situation generalization. The lecture included the name of each type of generalization, the definitions of each type of generalization and real-world examples of generalization. The examples of generalization corresponded to the C, C', and C'' stimuli. The content of the lecture was based on the textbook *Applied Behavior Analysis* (Cooper, Heron, & Heward, 2008). Participants were given the intraverbal test, emergent relations test, and multiple-choice test followed by the social validity survey after the lecture video was shown. The pre- and post-test given to the participants were identical.

Control group

The control group viewed 9:12 minutes of the video *Martin Seligman: The New Era of Positive Psychology*. The video time was edited down to match the approximate time of the lecture.

Data analysis

The data analysis was done using 3, 2x2 mixed ANOVA with observed power reported. A statistical analysis was conducted for all pre- and post-tests including the multiple-choice test, the intraverbal test, and the emergent relations test. Statistical power was reported using the SPSS results. In addition to statistical analysis, the appropriate figures such as bar graphs for each individual participants and relation (see Appendix G) to aid the analysis. A flow chart outlining the sequence of the study can be found in Appendix H.

CHAPTER IV

RESULTS

Emergent Relations Tests

The means and standard deviations were calculated for the lecture and control groups on the emergent relations pre- and posttests. For the lecture group, the mean score was 66% (*SD*=0.28) on the pretest and 82% (*SD*=0.25) on the posttest. For the control group the mean score was 62% (*SD*=0.25) on the pretest and 66% (*SD*= 0.21) on the posttest. A two-way ANOVA was conducted with the lecture and control conditions as the between-subjects variables and the emergent relations pretest and emergent relations posttest as the within-subjects variables. The within subjects pre- and posttest analysis did not yield a significant difference $F(1,19) = 3.924$, $p = 0.062$. The observed power was 0.468. The between-subjects test of the control and lecture conditions did not demonstrate a significant difference $F(1,19) = 1.198$, $p = 0.287$. The observed power was 0.180.

Intraverbal Tests

The means and standard deviations were calculated for the lecture and control groups on the intraverbal pre- and posttests. For the lecture group, the mean score was 26% (*SD*= 0.24) on the pretest and 67% (*SD*=0.21) on the posttest. For the control group, the mean score was 23% (*SD*=0.28) on the pretest and 39% (*SD*=0.33) on the posttest. A two-way ANOVA was conducted with the lecture and control conditions as the between-subjects variables and the intraverbal pretest and intraverbal posttest as the within-subjects variables. The within subjects pre- and posttest analysis yielded a significant difference $F(1,19) = 30.682$, $p = 0.000$. The observed power was 0.999. The between subjects test of the control and lecture groups did not demonstrate significant results $F(1,19) = 1.1943$, $p = 0.179$. The observed power was 0.263.

Multiple Choice Tests

The means and standard deviations were calculated for the lecture and control groups on the multiple-choice tests. For the lecture group, the mean score was 56% (*SD*= 0.13) on the pretest and 83% (*SD*=0.10) on the posttest. For the control group, the mean score was 51% (*SD*=0.19) on the pretest and 60% (*SD*=.27) on the posttest. A two-way ANOVA was conducted with the lecture and control conditions as the between-subjects variables and the multiple-choice pretest and multiple-choice posttest as the within-subjects variables. The within subjects pre- and posttest analysis yielded a significant difference $F(1,19) = 12.104$ $p = 0.003$. The observed power was 0.910. The between subjects test between the control and lecture groups did not yield significant results $F(1,19) = 4.098$, $p = 0.057$. The observed power was 0.485.

Social Validity Survey

Six questions were presented on the social validity survey. For the first question, participants in the lecture condition rated that they somewhat preferred the instructional method being used (M = 3.11, *SD* = 1.45). Participants in the control condition also rated that they somewhat preferred the instruction method used (M = 3.67, *SD* = 1.78). When asked how appropriate the time commitment was in relation to the amount the participants learned, the participants in the lecture group rated it as somewhat appropriate (M = 4, *SD* = 1.58), as did participants in the control group (M = 4.25, *SD* = 1.42). For the third question, when asked about the length of the instruction method, participants in the lecture group rated it as slightly less than somewhat preferred (M = 3.00, *SD* = 1.22), and participants in the control group rated it as somewhat preferred (M = 3.92, *SD* = 1.56). For the fourth question, when asked how confident the students were on their computer skills participants in the lecture group rated they were between somewhat confident and very confident (M = 5.78, *SD* = 2.49). Participants in the control group rated that they were slightly above somewhat confident (M = 4.75, *SD* = 1.29). For question five, when asked how much time a week they spent on computers participants in the lecture group reported an average of approximately 8 hours (M = 8.22, *SD* = 2.49) and participants in the control group spent approximately the same amount of time (M = 8.17, *SD* = 1.53). The sixth and final question was to ensure that participants were watching the videos that they were shown. Each of the participants listed three relevant comments about the videos they were exposed to.

Visual Inspection of Data

Mastery criterion for the emergent relations and intraverbal tests was 13 out of 15 or 87% correct, and the mastery criterion for the multiple-choice test was 8 out of 9 or 89% correct. In the lecture group six participants met criterion for the emergent relations posttest, two participants met criterion for the intraverbal posttest, and five participants met criterion for the multiple-choice posttest. For the control group, three participants met criterion for the emergent relations posttest, two participants met criterion for the intraverbal posttest and three participants met criterion for the multiple-choice posttest.

Mastery criterion for each individual relation was three out of three correct. For the multiple-choice posttest, five participants in the lecture group met criterion for the B-A relation, seven participants met criterion for the C-A relation, and three participants met criterion for the Cg”-A relation. For the control group, five participants met criterion for the B-A relation, five participants met criterion for the C-A relation, and two participants met criterion for the Cg”-A relation.

For the intraverbal posttest, the lecture group had three participants who met criterion for the B-A relation, five participants who met criterion for the C’-A relation, three participants who met criterion for the A-C relation, two participants who met criterion for the A-B relation, and four participants who met criterion for the Cg’-A relation. For the control group, three participants met criterion for the B-A relation, two participants met criterion for the C’-A relation, one participant met criterion for the A-C relation, two participants met criterion for the A-B relation, and two participants met criterion for the Cg’-A relation.

For the emergent relations posttest, lecture group had six participants who met criterion for the B-C’ relation, five participants who met criterion for the C”-B relation, five participants who met criterion for the Cg-A relation, six participants who met criterion for the B-A relation, and eight participants who met criterion for the C”-A relation. For the control group, five participants met criterion for the B” relation, five participants met criterion for the C’B relation, three participants met criterion for the Cg-A relation, three participants met criterion for the B-A relation, three participants met criterion for the C’-A relation.

CHAPTER V

DISCUSSION

The purpose of this study was to examine the emergence of stimulus equivalence using both selection-based and topography-based tests following a lecture or control condition, evaluate generalization to novel stimuli in both selection-based and topography-based response formats, and evaluate the social validity of the instructional procedure. Three pre- and posttests were given to the participants to evaluate the three research questions: an emergent relations test, a multiple-choice test, and an intraverbal test.

The emergent relations tests showed that the lecture group performed better on the posttest than the pretest with a mean score of 66% on the pretest and a mean 82% on the posttest. Participants in the control group did not perform as well scoring a mean of 62% on the pretest and 66% on the posttest. The results of the ANOVA showed that there was not a significant difference in pre- and posttest analysis or the between subjects ANOVA between the lecture and control groups.

On the multiple-choice tests both groups performed better on the posttests than on the pretest. The lecture group had a mean score of 56% on the pretest and a mean score of 83% on the posttest. The control group had a mean score of 51% on the pretest and 60% on the posttest. The within subjects analysis did yield a significant difference between pretest and posttest for the lecture group, but did not yield significant results for the control group, but the between subject test did not yield a significant difference between the lecture and control groups. When examining the relations testing for generalization (Cg”-A) for the multiple-choice test, participants in the lecture condition had a mean score of 53% on the pretest and 73% on the posttest. For the control condition participants had a mean score of 33% on the pretest and 52% on the posttest. Participants in the lecture group were able to perform better on the selection-based test than the control group. This may indicate that the lecture was effective at teaching the content it was intended to when learners were assessed using the selection-based test. Both groups showed similar increases in scores for generalization demonstrating that the lecture was not more effective than the control an increasing generalization even though participants in the control group were not exposed to the educational content in the lecture.

For the intraverbal tests, both groups performed better on the posttests than on the pretest. The lecture group, had a mean score of 26% on the pretest and 67% on the posttest. For the control group, the mean score on the pretest was 23% and the mean score for the posttest was 39%. The within subjects test did yield a significant difference between pretest and posttest for both lecture and control groups, but the between subject test did not yield a significant difference between the lecture and control groups. When examining the relations testing for generalization (Cg"-A) for the intraverbal test, participants in the lecture condition had a mean score of 33% on the pretest and 66% on the posttest. For the control condition participants had a mean score of 24% on the pretest and 42% on the posttest. Participants in both able to perform significantly better on the on the posttest when compared to the posttest. This means that both the lecture and control groups were able to do better on the topography-based posttest even though the control group was not exposed to the educational content. Both groups also resulted in similar increases in the scores for generalization, showing that both groups were equally effective at promoting generalization despite the content they were presented.

One of the results in this study that was not expected was the high performance on the pre-test scores of the control group on the emergent relations test. Participants in this study were not expected to have a baseline knowledge of the material presented. However due to the high scores on the emergent relations pre-test participants may have had some knowledge about the material presented. One of the reasons that this may have occurred is that participants did not adhere to the selection description provided in SONA. Although the selection criteria stated that participants needed to be undergraduate students to qualify for this experiment, it is possible that participants disregarded the description and signed up for the study. If a participant was in a special education graduate program, they would have taken classes that covered the presented material. Another possible explanation for the high scores could've been that participants already possessed knowledge on the subject from other areas such as work experience in the field of ABA or personal educational pursuits.

Finally, this study evaluated the social validity of the instructional procedure. Both the lecture and control groups both somewhat preferred the instructional method being used. Both groups also found that the time commitment was somewhat appropriate in relation to the amount the participants learned. There was a slight difference between the lecture and control group regarding the length of the instructional method. The lecture group rated the instructional method as slightly less than somewhat preferred, and the control group rated it as somewhat preferred. This could be due to the redundancy in the lecture video compared to the control video. The control video was a lecture on the current state of psychology and may have intrigued students, especially if they are currently studying psychology. In the lecture, each stimulus was presented two times to the participants to ensure they had sufficient time to be exposed to the stimuli. The lecture group was more confident than the control group on their computer skills. Both groups spent approximately 8 hours a week on computers.

Relating Findings to Previous Research

The present study examined how participants would perform on topography- and selection-based tests given to them after they were exposed to a lecture or control condition. Previous research by Sundberg & Sundberg (1990) found that a selection-based response format required more training in order to reach mastery, and the percentage of correct responses emitted was lower compared to topography-based training, which suggests that topography-based responding may be more desirable to teach than selection-based responding. A study by Lovett et al. (2011) used selection-based training tested for the emergence of topography-based responding. They found that topography-based responses did emerge in the participants that were assessed, which were college students. Sundberg & Sundberg (1990) compared selection-based training to topography-based training with individuals with disabilities and found that topography-based training was superior to selection-based training. A study by Polson and Parsons (2000) looked specifically at selection-based versus topography-based responding with undergraduate students and how it applies to stimulus equivalence. The main results of these studies were that symmetry emerged more reliably with selection-based responding than with topography-based responding. This provides evidence that selection-based Accurate topography-based responding is more difficult to emit than topography-based responding when using selection-based training.

In the present study, when given the topography-based intraverbal test participants in both the control and lecture condition performed significantly better on the posttest than the pretest. The mean posttest scores were higher for the participants in the lecture group than participants in the control group. However, there was not a significant difference between participants' scores in the control and lecture condition. For the selection based multiple-choice test, participants in both the control and lecture condition performed significantly better on the posttest than the pretest. The mean posttest scores were higher for the participants in the lecture group than participants in the control group. However, there was not a significant difference between participants scores in the control and lecture condition. Participants in the lecture group and control groups both scored higher on the pre- and posttests for the multiple-choice test than they did for the topography-based test. This provides evidence that selection-based responses are more accurately emitted following a lecture and control condition than topography-based responses. Neither group performed significantly better on one test or the other. The higher mean scores in the selection-based test are likely due to the participants being able to conditionally discriminate a selection-based test between a few choices by simply clicking compared to the simple discrimination of the topography-based responses where participants were required to emit lengthier responses by typing.

Another theme that this study investigated was the participants' ability to generalize what they learned to novel stimuli. Previous research using stimulus equivalence instructional methods has had varying results in regard to generalization to novel stimuli. Fields et al. (2009) demonstrated that participants taught using stimulus equivalence generalized to novel examples with a 37% increase in correct responding compared to a 2% increase in correct responding in the control group. In a study conducted by Lynch and Cuvo (1995), researchers used stimulus equivalence to teach fractions to 5th and 6th graders. Researchers obtained mixed results as to whether generalization occurred. Like the Lynch and Cuvo (1995) study, Lovett et al. (2011) found that generalization occurred with some stimuli but not others.

In the current study, when looking the generalized relation (Cg”-A) for the multiple-choice test participants in the lecture condition had a mean score of 53% on the pretest and 73% on the post test. For the control condition participants had a mean score of 33% on the pretest and 52% on the posttest. For the intraverbal test, when looking the generalization relation (Cg”-A), participants in the lecture condition had a mean score of 33% on the pretest and 67% on the post test. For the control condition, participants had a mean score of 24% on the pretest and 42% on the posttest. Since both the lecture and control condition had an increase in correct responding it does not provide strong evidence that the generalization that occurred was due to the particular condition that participants were exposed to.

Another theme that was examined in this study was the emergence of stimulus equivalence following a lecture or control condition. In a previous study by Lovett et al. (2011) researchers found that participants in the lecture condition performed similar to participants in the equivalence group on the paper and pencil test. In the present study, participants were given an emergent relations test to see if equivalence classes were formed. The emergent relations tests showed that the lecture group performed better on the posttest than the pretest with a mean score of 66% on the pretest and a mean 82% on the posttest. Participants in the control group did not perform as well scoring a mean of

62% on the pretest and 66% on the posttest. The results of the ANOVA showed that there was not a significant difference in pre- and posttest analysis or the test between the lecture and control. This test showed that participants in the lecture condition did not score significantly better than participants in the control group even though the mean scores in the lecture condition increased by 16%.

Limitations

One of the main limitations of this study was the small number of participants that were recruited in this study. Although some of the findings in this study were significant, more research needs to be conducted in order to be able to extend the results found here to a larger population. Another limitation of this study was some of the participants scored well on their pretests showing that they may have already had a background knowledge on the subject before taking this study. More research needs to be conducted with learners of different ages and history of learning.

Another limitation that may have impacted the study was the length of the lecture video, which could have given more in depth information on the subject. A typical lecture does not always state the facts without providing additional background information. Another possible limitation of this study is that the participants may have been fatigued by the time they completed the last test in the sequence. The total time of the study was around one hour to complete which may have been too long for each participant. Future research should investigate how total time of lecture and testing effect outcomes.

An additional possible limitation of this study was that some participants may have already possessed knowledge on the subject; this may have been reflected in the high scores of the emergent relations test. To control for this in future studies, demographic information should be obtained on each participant after they complete the study. This will help detect possible differences between groups such as classes taken, years attended college, and if they are graduate or undergraduate students.

Another confound that could have affected this study was the attention that the participants paid to the presented material. After given the initial instructions for the study, participants were rarely checked on until they completed the study. It is possible that during the study, participants were distracted by using their phones, drawing, etc. Although participants were asked questions about the information they were shown in the final question of the social validity survey, additional measures could be taken to ensure that participants were paying attention. One way in which future researchers could implement this is by having participants make an active response during the lecture (clicking a button, moving mouse, etc.) rather than taking notes.

The social validity test was given to the participants to assess how they rate each instructional method. Participants in this study gave very similar rating as to how much they preferred each instructional method. Participants in the control group preferred the length of their instructional method over those in the lecture group. Overall participants gave slightly favorable response to each instructional format. In the study by Lovett et al. (2011) researchers used a social validity survey to compare a lecture format of teaching to a computer-based stimulus equivalence protocol, which found that participants had no preference towards either teaching method.

Conclusion

The present study investigated the emergence of stimulus equivalence using both selection-based and topography-based tests following a lecture or control condition, evaluated generalization to novel stimuli in both selection-based and topography-based response formats, and evaluated the social validity of the instructional procedure. When selection-based tests were compared to topography-based tests, neither group performed significantly better on one type test or the other. As for generalization, both the lecture and control groups had an increase in correct responding. Since both groups had an increase in correct responding, the generalization that occurred was likely due to a testing effect and not the specific condition that the participants were exposed. The current study also showed that, like the Lovett et al. (2011) study, participants did not dislike using a lecture as an instructional format.

REFERENCES

Allen, I. E., & Seaman, J. (2011). *Going the distance: Online education in the United States, 2011*. Newburyport, MA. Sloan Consortium.

Austin, J. L. (2000). Behavioral approaches to college teaching. *Handbook of applied behavior analysis*, 449-472. Reno, NV, US: Context Press.

Bernstein, D., & Chase, P. N. (2013). Contributions of behavior analysis to higher education. In G. J. Madden, W. V. Dube, T. D. Hackenberg, G. P. Hanley, K. A. Lattal, G. J. Madden, ... K. A. Lattal (Eds.), *APA handbook of behavior analysis, Vol. 2: Translating principles into practice* (pp. 523-543). Washington, DC: American Psychological Association. doi:10.1037/13938-021

Cooper, J., Heron, T. and Heward, W. (2008). *Applied behavior analysis*. Upper Saddle River, N.J.: Parsons/Merrill-Prentice Hall.

De Rose, J., De Souza, D., & Hanna, E. (1996). Teaching Reading and Spelling: Exclusion And Stimulus Equivalence. Journal of Applied Behavior Analysis, 29(4), 451–469. http://doi.org/10.1901/jaba.1996.29-451

Fields, L., Travis, R., Roy, D., Yadlovker, E., De Aguiar-Rocha, L., & Sturmey, P. (2009). Equivalence class formation: A method for teaching statistical interactions. *Journal of Applied Behavior Analysis*, *42*(3), 575-593. doi:10.1901/jaba.2009.42-575

Fienup, D. M., Covey, D. P., & Critchfield, T. S. (2010). Teaching, brain-behavior relations economically with stimulus equivalence technology. *Journal of Applied Behavior Analysis, 43*(1), 19-33. doi:10.1901/jaba.2010.43-19

Gabriel, Trip. (2011, April 5) More Pupils Are Learning Online, Fueling Debate on Quality. *The New York Times. The New York Times*. Web. 23 Sept. 2015.

Green, G., & Saunders, R. R. (1998). Stimulus equivalence. In K. A. Lattal, M. Perone, K. A. Lattal, M. Perone (Eds.), *Handbook of research methods in human operant behavior* (pp. 229-262). New York, NY: Plenum Press.

Lovett, S., Rehfeldt, R. A., Garcia, Y., & Dunning, J. (2011). Comparison of a stimulus equivalence protocol and traditional lecture for teaching single-subject designs. *Journal of Applied Behavior Analysis*, *44*(4), 819-833.

Lynch, D. C., & Cuvo, A. J. (1995). Stimulus equivalence instruction of fraction-decimal relations. *Journal of Applied Behavior Analysis*, *28*(2), 115-126.

Ninness, C., Rumph, R., McCuller, G., Harrison, C., Ford, A. M., & Ninness, S. K. (2005). A functional analytic approach to computer-interactive mathematics. *Journal of Applied Behavior Analysis*, *38*(1), 1-22. doi:10.1901/jaba.2005.2-04

Pear, J. J., & Novak. M. (1996) Computer-aided personalized system of instruction: A program evaluation. *Teaching of Psychology*, *23*, 199-123. Doi:10.1207/s15328023top2302_14

Polson, D. A., & Parsons, J. A. (2000). Selection-based versus topography-based responding: An important distinction for stimulus equivalence? *The Analysis of Verbal Behavior*, *17*, 105.

Rosales, R., & Anne Rehfeldt, R. (2007). Contriving Transitive Conditioned Establishing Operations to Establish Derived Manding Skills in adults with

Severe Developmental Disabilities. *Journal of Applied Behavior Analysis*, *40*(1), 105–121. http://doi.org/10.1901/jaba.2007.117-05

Skinner, B. F. (1960). Teaching machines. *The review of economics and statistics*, 42(Suppl.), 189-191.

Skinner, B. F. (1968). *The technology of teaching*. New York: Appleton-Century-Crofts.

Skinner, B. F. (1984). The shame of American education. *American Psychologist*, *39*(9), 947.

Sundberg, C. T., & Sundberg, M. L. (1990). Comparing topography-based verbal behavior with stimulus selection-based verbal behavior. *The Analysis of Verbal Behavior*, *8*, 31.

Walker, B. D., Rehfeldt, R. A., & Ninness, C. (2010). Using the stimulus equivalence paradigm to teach course material in an undergraduate rehabilitation course. *Journal of Applied Behavior Analysis*, *43*(4), 615-633.

Wraikat, R., Sundberg, C. T., & Michael, J. (1991). Topography-based and selection-based verbal behavior: A further comparison. *The Analysis of Verbal Behavior, 9*, 1.

APPENDIXES

Appendix A

Equivalence Stimuli

A1	B1	C1
Response Generalization	The extent to which a leaner emits untrained responses that are functionally equivalent to the trained target behavior	A student is taught to hand write lecture notes in class to receive a better grade. Without being taught the student types notes in class to receive a better grade.
A2	B2	C2
Response Maintenance	The extent to which a learner continues to perform the target behavior after the intervention responsible for the initial appearance of the behavior is removed.	A student is taught to hand write lecture notes in class to receive a better grade. A month later the student is still able to hand write lecture notes.
A3	B3	C3
Setting/situation Generalization	The extent to which a leaner emits the target behavior in a setting or stimulus situation that is different from the instructional setting.	A student is taught to hand write lecture notes in class to receive a better grade. The student is then able to hand write lecture notes in other classes.

C1’	C1’’	C1g
A girl is taught to wash her hands before dinner at home. The girl then starts to use hand sanitizer before she eats dinner at home.	A girl is taught how to paint the house using a roller. Without being taught, sometimes the girl paints the house using a paint brush.	A child is taught how to tie his shoe a certain way. The child is able to tie his shoe a different way without any additional training.
C2’	C2’’	C2g
A girl is taught to wash her hands before dinner at home. A year later the girl still washes her hands before she eats dinner.	A girl is taught how to paint the house using a roller. The next summer she repaints the house and is still able to use a roller.	A child is taught how to tie his shoe a certain way. A week later the child is still able to tie his shoe without any additional training.
C3’	C3’’	C3g
A girl is taught to wash her hands before dinner at home. The girl then washes her hands before meals at school without prior training at school.	A girl is taught how to paint the house using a roller. She then is able to help paint her neighbors’ houses using the roller.	A child is taught how to tie his shoe a certain way. When given a different pair of shoes the child is able to tie them with no additional training.

C1g'	C1g"
A grocery store clerk is taught how to use a cash register. When the cash register breaks, the grocery store clerk is able to make transactions using a ledger without any additional training.	A boy is taught to wash the car counter clockwise in order to clean it. The boy starts to wash the car counter clockwise because it is more comfortable.
C2g'	C1g"
A grocery store clerk is taught how to use a cash register. Without any further instruction, a year later the clerk still knows how to use the cash register.	A boy is taught to wash the car counter clockwise in order to clean it. The boy is able to wash the car the same way when he tries a month later.
C3g'	C3g"
A grocery store clerk is taught how to use a cash register. The grocery store clerk is then able to use a different cash register without any additional training.	A boy is taught to wash the car counter clockwise in order to clean it. The boy is able to wash a different car the same way he was taught to wash the original car.

Appendix B

Pre-Session Checklist

Thank you for participating in this study. To minimize distractions please:

- Turn off your cell phone
- Remove any food/or drink
- Turn off any music listening devices

Appendix C

Multiple-Choice Test Questions

Definition-name (B-A) relations:

1. The extent to which a leaner emits untrained responses that are functionally equivalent to the trained target behavior
    a. Response Generalization
    b. Response Maintenance
    c. Setting/Situation Generalization

Answer: a. Response Generalization

2. The extent to which a learner continues to perform the target behavior after a portion or all of the intervention responsible for the behavior's initial appearance in the learner's repertoire has been terminated.
    a. Response Generalization
    b. Response Maintenance
    c. Setting/Situation Generalization

Answer: b. Response Maintenance

3. The extent to which a leaner emits the target behavior in a setting or stimulus situation that is different from the instructional setting.
    a. Response Generalization
    b. Response Maintenance
    c. Setting/Situation Generalization

Answer: c. Setting/Situation Generalization

Vignette-name (C-A) relations:

4. A student is taught to hand write lecture notes in class to receive a better grade. Without being taught the student types notes in class to receive a better grade.
    a. Response Generalization
    b. Response Maintenance
    c. Setting/Situation Generalization

Answer: A. Response Generalization

5. A student is taught to hand write lecture notes in class to receive a better grade. A month later the student is still able to hand write lecture notes.
    a. Response Generalization
    b. Response Maintenance
    c. Setting/Situation Generalization

Answer: b. Response Maintenance

6. A student is taught to hand write lecture notes in class to receive a better grade. The student is then able to hand write lecture notes in other classes.
   a. Response Generalization
   b. Response Maintenance
   c. Setting/Situation Generalization

Answer: c. Setting/Situation Generalization

Novel vignette-name (Cg''-A) relations:

7. A boy is taught to wash the car counter clockwise in order to clean it. The boy is able to wash the car the same way when he tries a month later.
   a. Response Generalization
   b. Response Maintenance
   c. Setting/Situation Generalization

Answer: b. Response Maintenance

8. A boy is taught to wash the car counter clockwise in order to clean it. The boy starts to wash the car counter clockwise because it is more comfortable.
   a. Response Generalization
   b. Response Maintenance
   c. Setting/Situation Generalization

Answer: a. Response Generalization

9. A boy is taught to wash the car counter clockwise in order to clean it. The boy is able to wash a different car the same way he was taught to wash the original car.
   a. Response Generalization
   b. Response Maintenance
   c. Setting/Situation Generalization

Answer: c. Setting/Situation Generalization

Appendix D

Intraverbal Test

Definition-name (B-A) relations:

1. The extent to which a leaner emits untrained responses that are functionally equivalent to the trained target behavior. Write the corresponding type of generalization. [Answer: response generalization]
2. The extent to which a learner continues to perform the target behavior after a portion or all of the intervention responsible for the behavior's initial appearance in the learner's repertoire has been terminated. Write the corresponding type of generalization. [Answer: response maintenance]
3. The extent to which a leaner emits the target behavior in a setting or stimulus situation that is different from the instructional setting. Write the corresponding type of generalization. [Answer: setting/situation generalization]

Vignette-name (C'-A) relations:

4. A girl is taught to wash her hands before dinner at home. The girl then starts to use hand sanitizer before she eats dinner at home. Write the corresponding type of generalization. [Answer: response generalization]
5. A girl is taught to wash her hands before dinner at home. A year later the girl still washes her hands before she eats dinner. Write the corresponding type of generalization. [Answer: response maintenance]
6. A girl is taught to wash her hands before dinner at home. The girl then washes her hands before meals at school without prior training. Write the corresponding type of generalization. [Answer: setting/situation generalization]

Name-vignette (A-C) relations:

7. Provide an example of response generalization

[Answer: A grocery store clerk is taught how to use a cash register. When the cash register breaks, the grocery store clerk is able to make transactions using a ledger without any additional training.]

8. Provide an example of response maintenance

[Answer: A grocery store clerk is taught how to use a cash register. Without any further instruction, a year later the clerk still knows how to use the cash register.]

9. Provide an example of setting/situation generalization

[Answer: A boy is taught to wash the car counter clockwise in order to clean it. The boy is able to wash a different car the same way he was taught to wash the original car.]

Name-definition (A-B) relations:

10. Response Generalization. Write the definition for this type of generalization.

    [Answer: The extent to which a leaner emits untrained responses that are functionally equivalent to the trained target behavior.]

11. Response Maintenance. Write the definition for this type of generalization.

    [Answer: The extent to which a learner continues to perform the target behavior after a portion or all of the intervention responsible for the behavior's initial appearance in the learner's repertoire has been terminated.]

12. Setting/situation Generalization. Write the definition for this type of generalization.

[Answer: The extent to which a leaner emits the target behavior in a setting or stimulus situation that is different from the instructional setting.]

Novel vignette-name (Cg'-A) relations:

13. A grocery store clerk is taught how to use a cash register. Without any further instruction, a year later the clerk still knows how to use the cash register. Write the corresponding type of generalization. [Answer: response maintenance]
14. A grocery store clerk is taught how to use a cash register. The grocery store clerk is then able to use a different cash register without any additional training. Write the corresponding type of generalization. [Answer: setting/situation generalization]
15. A grocery store clerk is taught how to use a cash register. When the cash register breaks, the grocery store clerk is able to make transactions using a ledger without any additional training. Write the corresponding type of generalization. [Answer: response generalization]

Appendix E

Intraverbal Scoring Guide

	Stimulus A	Stimulus B	Stimulus C
Target Response	Response maintenance, response generalization, and setting and situation generalization	"The extent to which a leaner emits untrained responses that are functionally equivalent to the trained target behavior", "The extent to which a learner continues to perform the target behavior after the intervention responsible for the initial appearance of the behavior is removed", "The extent to which a leaner emits the target behavior in a setting or stimulus situation that is different from the instructional setting."	Vignettes for questions on response maintenance should illustrate a response emitted at another time after the original response was learned. Vignettes for questions on response generalization should illustrate that a response is being emitted in a different for to accomplish that same goal as the original response. Vignettes for questions on setting/situation generalization should illustrate that a response is being emitted in a new place or in a new situation.
Rule	No more than 4 letters incorrect in each word. Missing letters will be counted as incorrect.	An answer scored as correct will be either the exact definition of one that conveys the meaning of the definition.	An answer will be scored as correct if it illustrates the type of generalization asked.
Correct Examples	Response Maintenance	The extent to which a leaner emits the target behavior in a setting or stimulus situation that is different from the instructional setting	A boy is taught how to tie his shoe. A week later he is still able to tie his shoe.
Incorrect Examples	Response Maintenance	Response generalization means to generalize responses.	A boy is taught how to tie his shoe and he does a good job.

Appendix F

Social Validity Survey

Rate the degree to which you would prefer to be taught using this instructional method.

1	2	3	4	5	6	7
Not at all preferred			Somewhat preferred			Very preferred

How appropriate was the time commitment for this instructional method in relation to the amount you feel you have learned?

1	2	3	4	5	6	7
Not at all appropriate			Somewhat appropriate			Very appropriate

How do you feel about the length of this instructional method?

1	2	3	4	5	6	7
Not at all appropriate			Somewhat appropriate			Very appropriate

How confident are you in your computer skills?

1	2	3	4	5	6	7
Not at all confident			Somewhat Confident			Very confident

How much time do you spend on a computer per week?

1 2 3 4 5 6 7 8 9 10+

Povide 3 items you remember from the video that you watched.

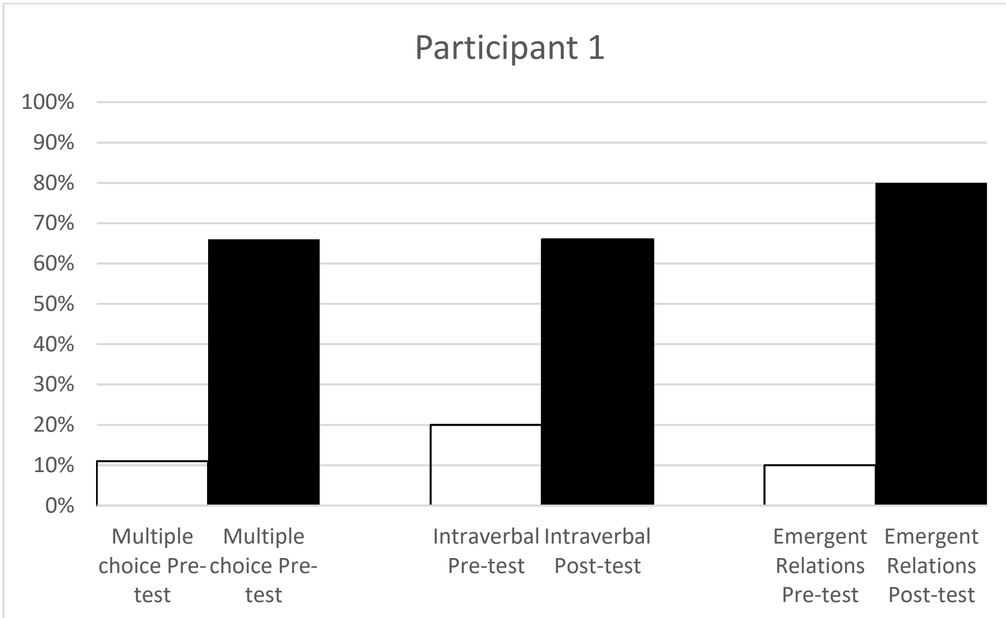
____

____

____

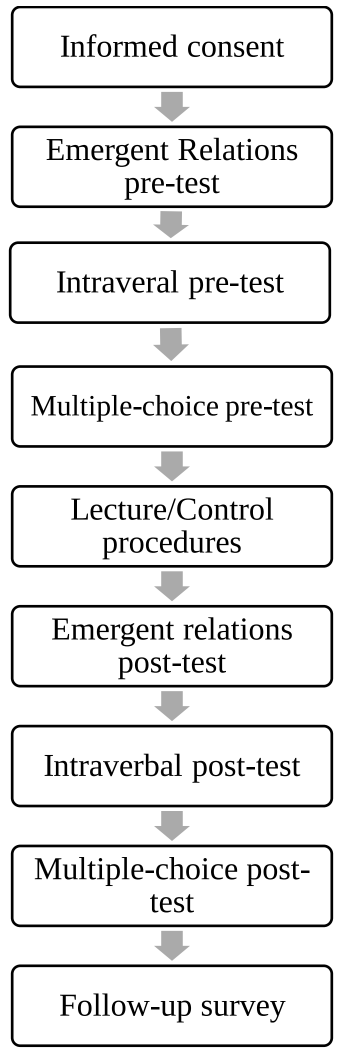
Appendix G

Data Analysis Example



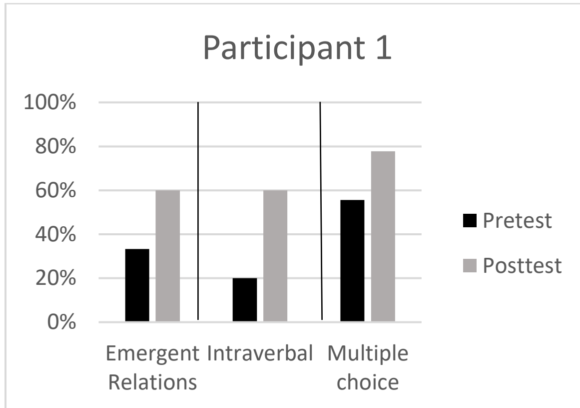
Appendix H

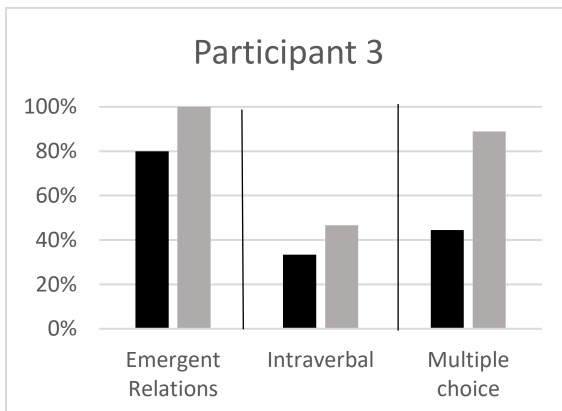
Flow Chart

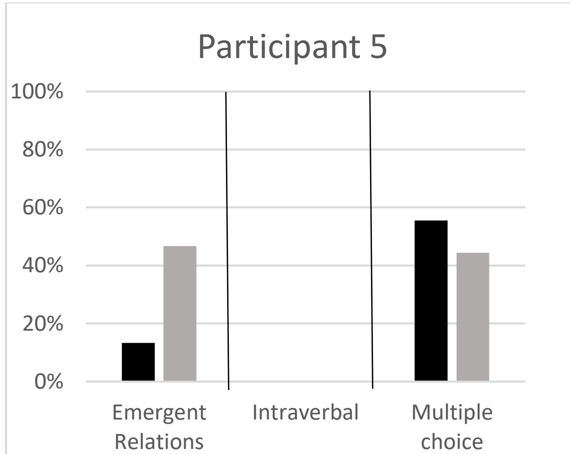


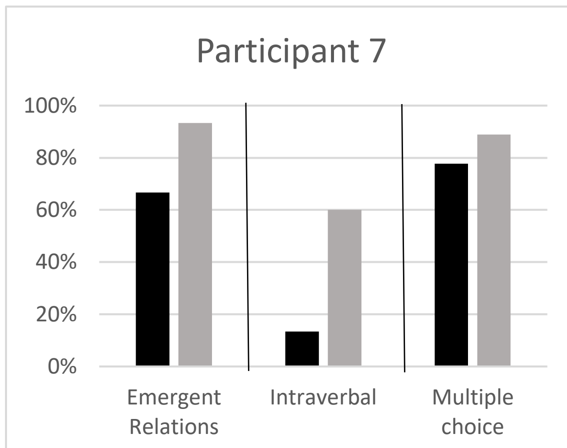
Appendix I

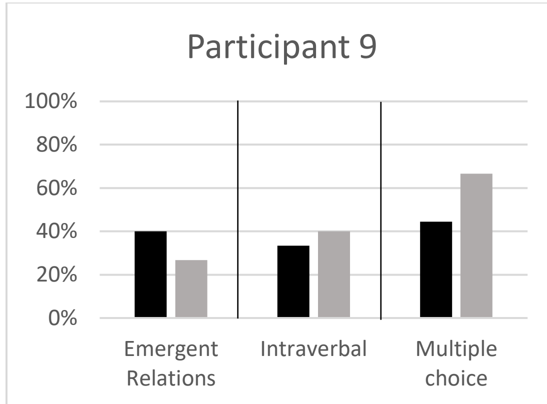
Pre- and Posttest Results for the Lecture Graphs Group

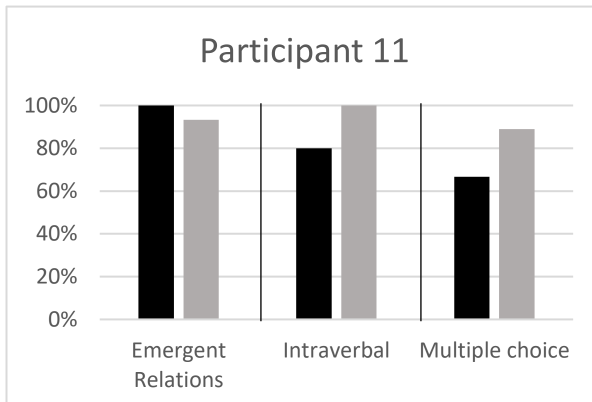


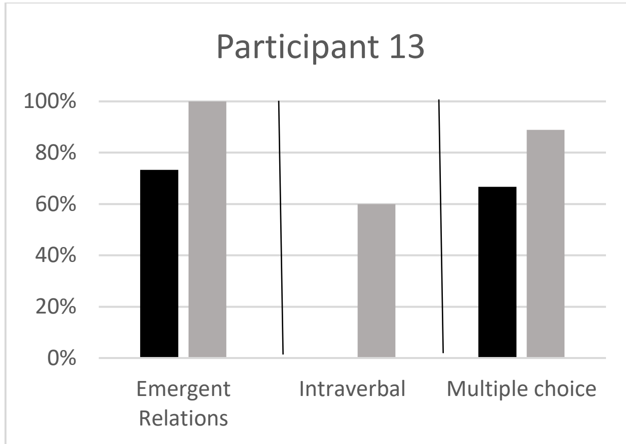


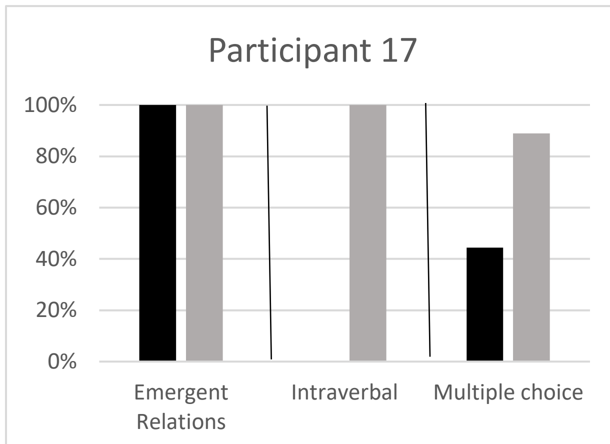

Participant 5
100%
80%
60%
40%
20%
0%
Emergent Relations
Intraverbal
Multiple choice

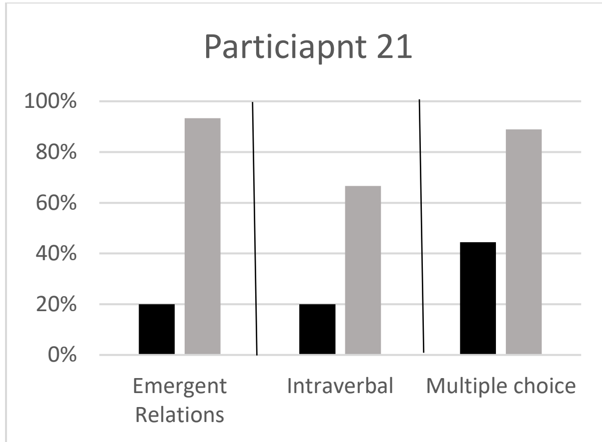

Participant 7
100%
80%
60%
40%
20%
0%
Emergent Relations
Intraverbal
Multiple choice

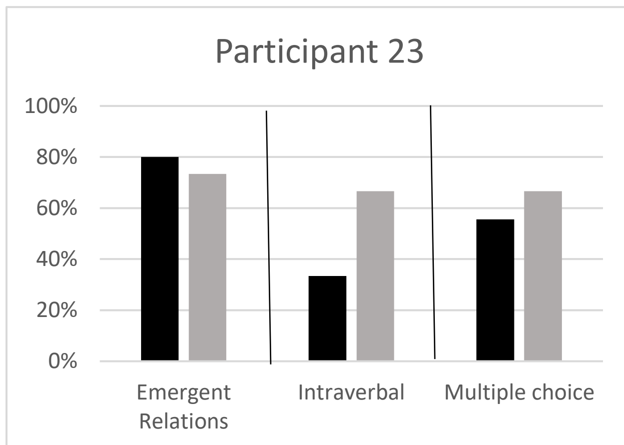

Participant 9
100%
80%
60%
40%
20%
0%
Emergent Relations
Intraverbal
Multiple choice


Participant 11
100%
80%
60%
40%
20%
0%
Emergent Relations
Intraverbal
Multiple choice


Participant 13
100%
80%
60%
40%
20%
0%
Emergent Relations
Intraverbal
Multiple choice

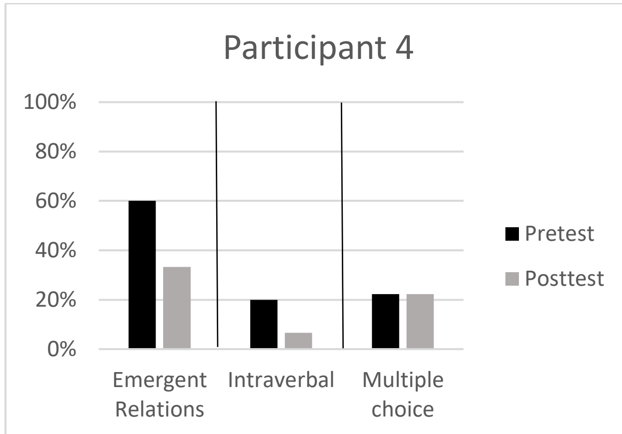

Participant 17
100%
80%
60%
40%
20%
0%
Emergent Relations
Intraverbal
Multiple choice

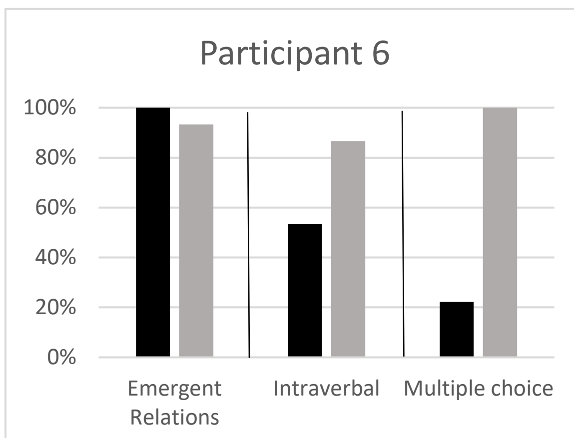

Particiapnt 21
100%
80%
60%
40%
20%
0%
Emergent Relations
Intraverbal
Multiple choice

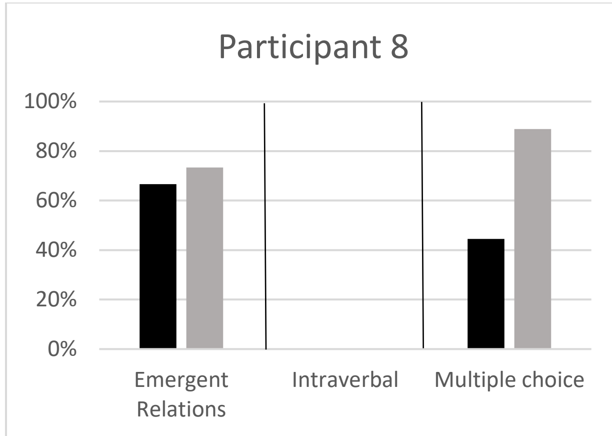

Participant 23
100%
80%
60%
40%
20%
0%
Emergent Relations
Intraverbal
Multiple choice

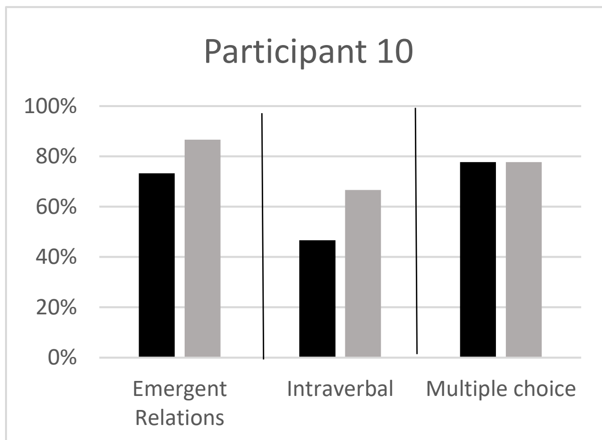
Appendix J

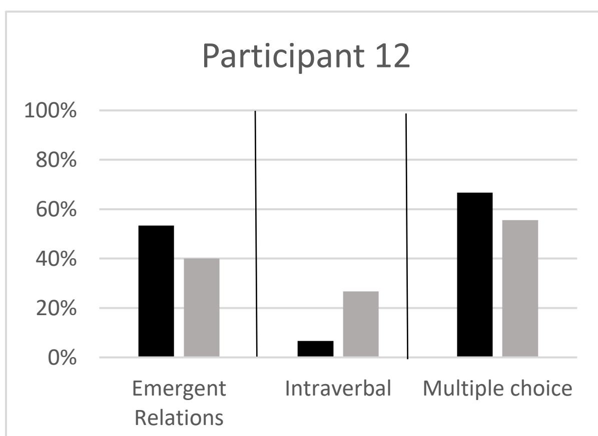
Pre- and Posttest Results for the Control Group

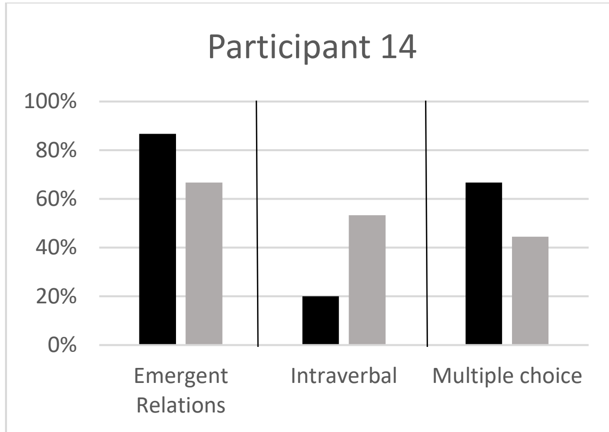




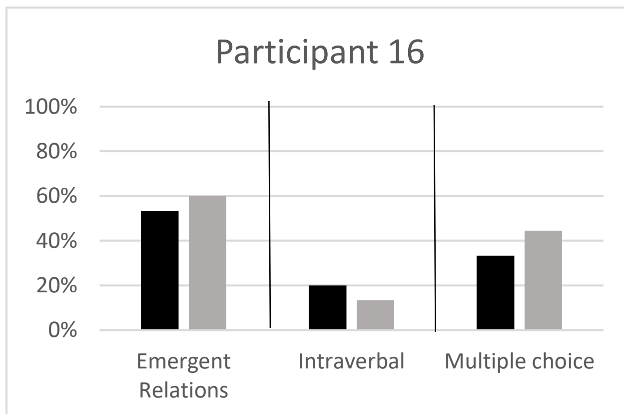

Participant 8
100%
80%
60%
40%
20%
0%
Emergent Relations
Intraverbal
Multiple choice


Participant 10
100%
80%
60%
40%
20%
0%
Emergent Relations
Intraverbal
Multiple choice

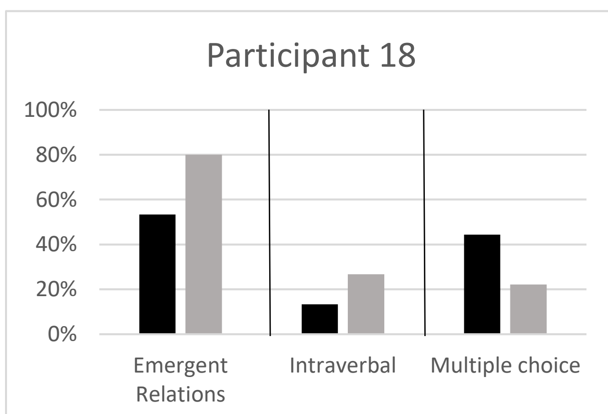

Participant 12
100%
80%
60%
40%
20%
0%
Emergent Relations
Intraverbal
Multiple choice



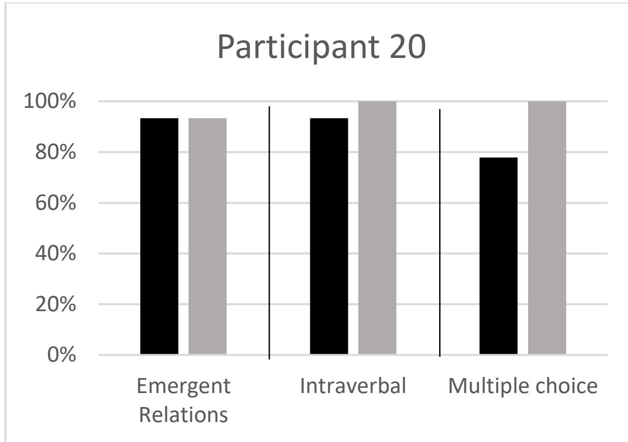
Participant 14
100%
80%
60%
40%
20%
0%
Emergent Relations
Intraverbal
Multiple choice

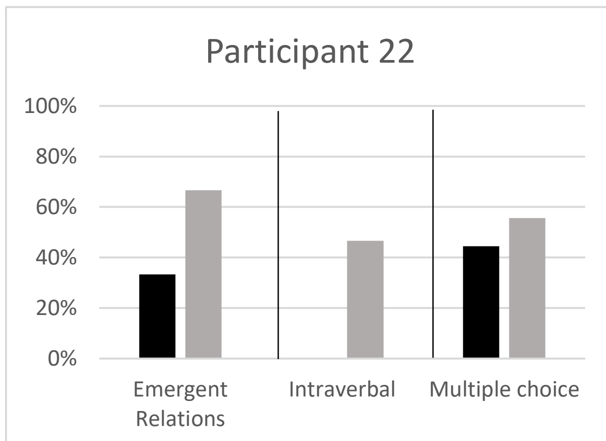


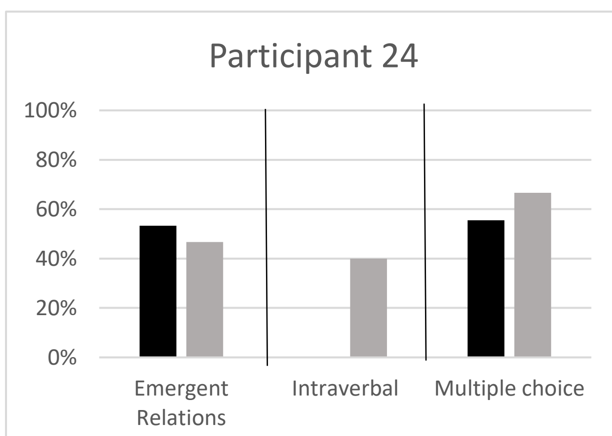
Participant 16
100%
80%
60%
40%
20%
0%
Emergent Relations
Intraverbal
Multiple choice



Participant 18
100%
80%
60%
40%
20%
0%
Emergent Relations
Intraverbal
Multiple choice

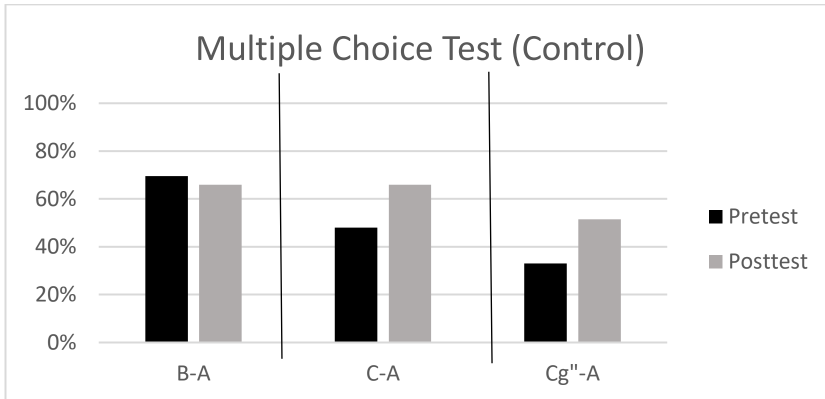

Participant 20
100%
80%
60%
40%
20%
0%
Emergent Relations
Intraverbal
Multiple choice

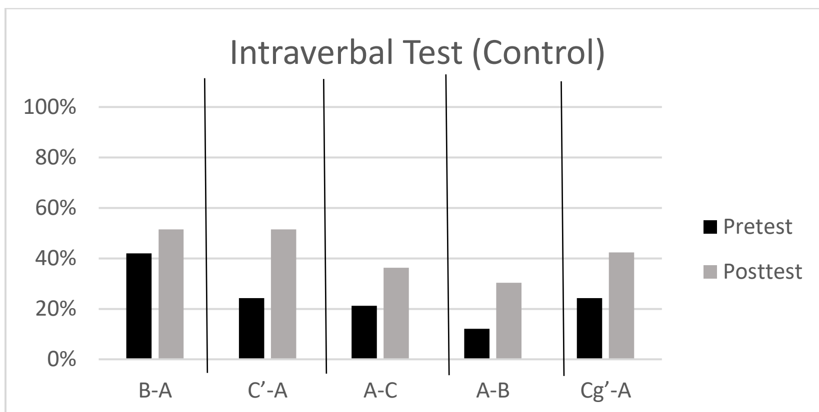

Participant 22
100%
80%
60%
40%
20%
0%
Emergent Relations
Intraverbal
Multiple choice

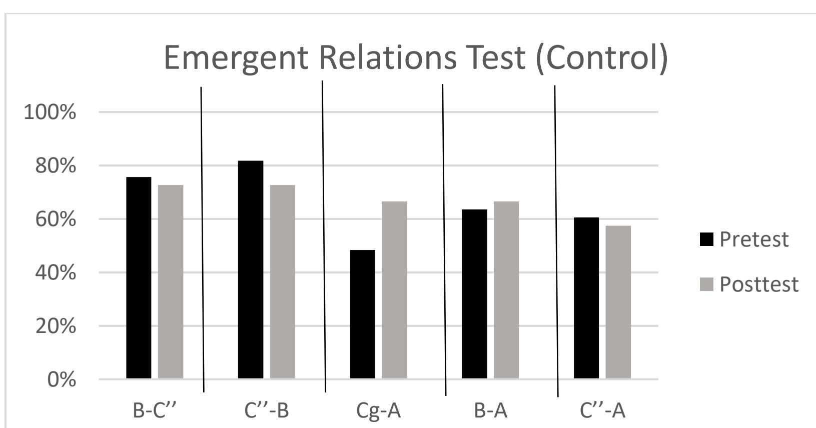

Participant 24
100%
80%
60%
40%
20%
0%
Emergent Relations
Intraverbal
Multiple choice

Appendix K

Scores on Individual Relations for the Control Group







Appendix L

Scores on Individual Relations for the Lecture Group