An Evaluation of Two Stimulus Equivalence Training Sequences on the Emergence of Novel Intraverbals

Brittany Zaring-Hinkle¹ · Charlotte Lynn Carp¹ ·
Tracy L. Lepper¹

Published online: 8 November 2016

© Association for Behavior Analysis International 2016

Abstract Researchers have begun to investigate the emergence of novel intraverbals using equivalence-based instruction (EBI) in typically developing children (Carp & Petursdottir, 2012; Pérez-González, Herszlikowicz, & Williams, 2008). We sought to replicate and extend the previous research by investigating two stimulus equivalence training sequences (e.g., linear series—LS and one to many—OTM) in the emergence of novel intraverbals in a two-part study with college students. Experiment 1 was designed to partially replicate the previous research by training intraverbals using an LS arrangement and then testing for the emergence of novel intraverbals. Novel intraverbals did not emerge after baseline training alone for the majority of participants. Experiment 2 investigated whether a different training sequence (i.e., OTM) would result in the emergence of novel intraverbals. Novel intraverbals did emerge following baseline training alone for the majority of participants. Overall, these results suggest that training intraverbals in an OTM training sequence may establish conditional discriminations during training, which may make it a more advantageous sequence, in that following training, more novel intraverbals emerge.

Keywords College students · Conditional discriminations · Emergence · Inferential statistics · Intraverbals · Linear series · Stimulus equivalence

Complex human behavior often requires persons to respond appropriately to novel stimuli on the basis of a previous history with other related stimuli (Spradlin, Cotter, & Baxley, 1973). One framework that has been suggested to explain these complex

This article is part of a special section on the intraverbal relation in *The Analysis of Verbal Behavior*.

✉ Charlotte Lynn Carp
charlottecarp@me.com

¹ Department of Psychology, McNeese State University, 4205 Ryan Street, Lake Charles, LA 70605, USA

networks of relations is stimulus equivalence (Sidman, 1994). Stimulus equivalence is a paradigm in which a person is taught a small set of stimulus-stimulus relations between stimuli, resulting in those stimuli becoming related in ways that have not been directly trained. Stimuli are said to enter into equivalence classes when the trained stimulus-stimulus relations demonstrate the three defining derived stimulus relations of reflexivity, symmetry, and transitivity (Sidman, 2000; Sidman & Tailby, 1982). Equivalence-based instruction (EBI) is suggested to be a very efficient teaching tool, in that you can teach several relations and get more relations “for free.” EBI has been used to create efficiencies in teaching in many different populations such as college-aged students, typically developing children, children and adults with and without developmental disabilities, children with visual impairments, and senior citizens with dementia (e.g., Carp & Petursdottir, 2012; Critchfield & Fienup, 2010; Hove, 2003; Lovett, Rehfeldt, Garcia, & Dunning, 2011; Pérez-González, Garcia-Asenjo, Williams, & Camerero, 2007; Pérez-González et al., 2008; Petursdottir, Carr, Lechago, & Almason, 2008; Saunders & McEntee, 2004; Saunders, Chaney, & Marquis, 2005; Saunders, Wachter, & Spradlin, 1988; Spradlin & Saunders, 1986; Toussaint & Tiger, 2010; Walker & Rehfeldt, 2010; Walker & Rehfeldt, 2012).

Recently, a new category of research has been developed that applies the methods used in EBI to the emergence of novel intraverbals (i.e., Carp & Petursdottir, 2012; Pérez-González et al., 2008). According to Skinner (1957), the intraverbal relation is a form of verbal behavior controlled by a verbal stimulus due to a history of generalized conditioned reinforcement in which there is no point-to-point correspondence between the stimulus and the response. This new line of research investigates the effects of training intraverbal relations on the emergence of other intraverbal relations. That is, all stimuli that form the equivalence class are vocal-verbal stimuli. This is different from other studies in the stimulus equivalence literature which typically involve visual stimuli. To date, only two studies have used three sets of stimuli (i.e., sets A, B, and C) in which were all intraverbal relations (i.e., Carp & Petursdottir, 2012; Pérez-González et al., 2008). Pérez-González et al. (2008) taught two types of intraverbals using Spanish countries, cities, and parks. The first intraverbal taught was the A-B country-city intraverbal. The antecedent stimuli consisted of either (a) “Name the city of Argentina” with the correct response being “Buenos Aires” or (b) “Name the city of Uruguay” with the correct response being “Montevideo.” The second type of intraverbal taught was the B-C city-park intraverbal. The antecedent stimuli consisted of (a) “Name the park of Buenos Aires” with the correct response being “el Botánico” or (b) “Name the park of Montevideo” with the correct response being “el Lecoc.” Pérez-González et al. (2008) investigated the emergence of novel intraverbals after the A-B and B-C intraverbals had been trained. Only one out of the five children tested showed all the emergent intraverbals. Pérez-González et al. (2008) also looked at the effects of categories and exemplars to see if these two simpler intraverbals would facilitate the emergence of novel intraverbals in an additional experiment. All of the children learned all of the intraverbals after being exposed to both category and exemplar training.

Carp and Petursdottir (2012) extended the Pérez-González et al. (2008) study with nine typically developing children using US states, cities, and parks. The same baseline intraverbal relations were trained (i.e., A-B, B-C). The first intraverbal (i.e., A-B state-city) consisted of either “Name a city in Florida” with the correct response being

“Branford” or “Name a city in Utah” with the correct response being “Midway.” The second intraverbal relation taught (i.e., B-C city-park) consisted of (a) “Name a park in Branford” with the correct response being “Troy Springs” or (b) “Name a park in Midway” with the correct response being “Deer Creek.” Following baseline training, all relations were tested. If novel intraverbals did not emerge, then some participants were exposed to category training, and then tested on the relations again. If novel intraverbals did not emerge, then participants were exposed to exemplar training, and then tested on the relations a final time. The other participants went through the same process, except that they were exposed to exemplar training and then category training. Three out of the nine participants demonstrated the emergence of novel intraverbals following baseline training alone. Furthermore, all of the other six participants required both category and exemplar training for novel intraverbals to emerge.

In the Carp and Petursdottir (2012) and Pérez-González et al. (2008) studies, the initial A-B and B-C training required participants only to make simple discriminations. That is, the participant’s response only needed to come under control of the discriminative stimulus (S^D). For example, in the Carp and Petursdottir (2012) study, during A-B training, the antecedent stimuli were “*Name a city in Florida*” and “*Name a city in Utah*.” For these two intraverbal relations, the participant only needed to discriminate between “*Florida*” and “*Utah*” to respond correctly on each trial. The same simple discrimination was also used in the B-C training in which the antecedent stimuli were “*Name a park in Branford*” and “*Name a park in Midway*.” The participant only needed to discriminate between “*Branford*” and “*Midway*” to respond correctly on each trial. In the testing phase, the participant was required to make conditional discriminations. That is, to respond correctly during testing trials, the participant’s response needed to come under control of both the S^D and the conditional stimulus. For example, in the A-C relation, the antecedent stimuli “*Name a park in Florida*,” the conditional stimulus would be “*park*” and the S^D would be “*Florida*.” However, during training, the child was never required to discriminate *park* from “*city*.” As a result, the child might continue responding to “*Florida*” by saying *Brandford*. Carp and Petursdottir (2012) conducted an error analysis to evaluate if this pattern might explain the lack of intraverbal emergence in the majority of participants following baseline intraverbal training alone. For at least three participants, the majority of errors were due to the baseline intraverbal training establishing simple control. That is, those participants’ responses during testing were under control of the S^D and not the S^D and the conditional stimulus. Therefore, they suggested that the relations involved in this task may be more complicated than the simple training A-B, B-C, and testing of B-A, C-B, A-C, C-A relations, and that a more accurate description might be training XA-B and YB-C and testing the emergence of ZB-A, XC-B, ZC-A, and YA-C (X is the verbal stimulus “*city*,” Y is the verbal stimulus “*park*,” and Z is the verbal stimulus “*country*,” and all three are constant, whereas A, B, and C vary across stimulus classes).

The Carp and Petursdottir (2012) and the Pérez-González et al. (2008) studies both implemented what is referred to as a linear series (LS) training structure. In the typical stimulus equivalence literature, there are three main training structures to establish equivalence classes: the LS (A-B and B-C), one-to-many or sample-as-node (OTM; A-B and A-C), and many-to-one or comparison-as-node (MTO; B-A and C-A). Overall, the experimental research on training structures regarding stimulus equivalence has suggested that the LS is the least successful in establishing equivalence classes

(Arntzen, 2012). Regarding the OTM and MTO, the literature suggests that the MTO is more effective in establishing equivalence classes in persons with developmental disabilities (e.g., Saunders et al., 2005; Spradlin & Saunders, 1986). For persons without developmental disabilities, the literature has been inconclusive in that some studies have found OTM to be more effective than MTO (e.g., Hove, 2003), and some studies have found MTO to be more effective than OTM (e.g., Arntzen & Holth, 2000). Regardless, Saunders and Green (1999) suggested that the OTM sequence actually establishes more of the component discriminations that in turn result in the equivalence class being more likely to form compared to the LS and the MTO. This may relate to the Carp and Petursdottir (2012) and the Pérez-González et al. (2008) studies in that, as discussed previously, the LS implemented did not establish the component discriminations needed for the participants to be successful in the equivalence test. Participants who did not pass the equivalence test following baseline training only passed after those component discriminations had been trained via category and exemplar training. Therefore, it may be that implementing a different training structure (e.g., OTM) could establish those component discriminations during baseline training. This procedure would increase the efficiency of the training by not needing the additional category and exemplar training for novel intraverbals to emerge.

The purpose of the present study was to replicate and extend the Pérez-González et al. (2008) and Carp and Petursdottir (2012) experiments with four goals. The first goal was to investigate whether the previously observed results are also evident with a different population (i.e., college students) besides typically developing children. The second goal was to investigate whether the previously observed results are also evident with completely new intraverbal relations (i.e., statistical relations) besides countries/states, cities, and parks. The third goal was to extend the two previous studies by investigating a different stimulus equivalence training sequence on the emergence of novel intraverbals. The final goal was to extend the previous research on equivalence classes and emergent intraverbals by including a visual-visual equivalence test.

Experiment 1

Experiment 1 sought to replicate and extend previous research by investigating the emergence of novel intraverbals using the LS training structure with inferential statistical relations in college students.

Method

Participants

Participants included 11 undergraduate students. Of the 11 participants, only eight proceeded to the intraverbal training and testing phase of the experiment. Two of the participants met the criterion needed to pass the initial visual-visual equivalence tests and were excluded from the study. The last participant completed the training phase but failed to reach criterion during the testing phase. Participants were required to be at least 18 years old; no other restrictions were placed on participants. Participants received course credit and up to \$10 for their participation.

Setting

Sessions were conducted in a classroom setting. The typical classroom contained approximately 40 student desks, an instructor desk in the front of the room, and a computer with a projector screen in the front of the room. The experimenter and the participant both sat in a computer chair next to each other. In all sessions, an observer sat approximately 2 m from the experimenter and the participant. The only people in the room during sessions were the experimenter, the participant, and one observer.

Stimuli and Apparatus

Table 1 shows the antecedent verbal stimuli presented in baseline and test trials, along with correct vocal responses. For consistency with the previous literature (i.e., Carp & Petursdottir, 2012; Pérez-González et al., 2008), alphanumeric designations were used to describe the trained and tested intraverbal relations. Thus, the labels A1 and A2 refer to the “ p values,” regardless of whether the names were presented as antecedent stimuli by the experimenter (as in A-B relations) or produced as correct responses by the participants (as in B-A relations). Similarly, B1 and B2 refer to the “research hypothesis,” and C1 and C2 refer to the “null.” During intraverbal training and intraverbal probes, all intraverbals were presented vocally by the experimenter.

Figure 1 shows the stimuli used during the visual-visual equivalence tests. Visual stimuli consisted of two p value stimuli (A1 [$p < .05$] and A2 [$p > .05$]), two research hypotheses stimuli (B1 [results are significant] and B2 [results are not significant]), and

Table 1 Antecedent stimuli and correct responses on baseline and test trials in experiment 1

	Antecedent stimuli	Correct response
A-B [P—RH (trained)]		
If p is [A1], then the results are what?	[A1] less than .05	[B1] significant
If p is [A2], then the results are what?	[A2] greater than .05	[B2] non-significant
B-C [RH-Null (trained)]		
If the results are [B1], then we do what to the null?	[B1] significant	[C1] reject
If the results are [B2], then we do what to the null?	[B2] non-significant	[C2] accept
B-A [RH—P]		
If the results are [B1], then p is what?	[B1] significant	[A1] less than .05
If the results are [B2], then p is what?	[B2] non-significant	[A2] greater than .05
C-B [Null-RH]		
If the null is [C1], then the results are what?	[C1] reject	[B1] significant
If the null is [C1], then the results are what?	[C2] accept	[B2] non-significant
A-C [P-Null]		
If p is [A1], then do what to the null?	[A1] less than .05	[C1] reject
If p is [A2], then do what to the null?	[A2] greater than .05 [	[C2] accept
C-A [Null—P]		
If the null is [C1], then p is what?	[C1] reject	[A1] less than .05
If the null is [C2], then p is what?	[C2] accept	[A2] greater than .05

Fig. 1 Stimuli presented during the visual-textual equivalence tests. The A stimuli were “*p* values,” B stimuli were “research hypotheses,” and C stimuli refer to decisions about “null”

two null stimuli (C1 [reject the null] and C2 [accept the null]). Each stimulus was contained within a black border measuring 5.72 by 5.92 cm. During the visual-visual equivalence test, stimulus presentation was controlled by Microsoft Office power point. The application ran on a laptop computer equipped with a Windows XP operating system and a 29.0 by 19.8 cm monitor. Visual stimuli were presented on the screen against a white background. Each trial began with the presentation of a visual sample stimulus in the horizontal center of the top third of the screen. Once the participant touched the sample stimulus, two comparison stimuli were presented equidistant from one another across the lowest third of the screen, with the two stimuli centered below the sample stimulus. The sample and comparison stimuli remained on the screen until the participant selected (i.e., touched) a comparison stimulus. Once a comparison stimulus was selected, a black screen was presented for 2 s, and then the next trial was presented.

Response Measurement and Interobserver Agreement

During intraverbal training, intraverbal probes, and the visual-visual equivalence tests, the experimenter recorded correct and incorrect responses on data sheets. Correct answers were defined as those that matched the experimenter-defined target answers, and incorrect answers were defined as any answers that did not match the defined target answers, including failure to answer. On all trials, the first response the participant emitted was scored as correct or incorrect. A second observer independently recorded data during at least 50 % of all intraverbal probes, intraverbal training, and visual equivalence testing for all participants. For each trial, an agreement was scored if both the experimenter and the second observer scored the trial as correct, or both scored it as incorrect. In order to calculate the interobserver agreement (IOA), the number of agreements was divided by the number of trials and then the score was multiplied by 100. IOA averaged 92.6 % (range, 77.8–100 %) for participant 1, 95.8 % (range, 91.7–100 %) for participant 2, 97.8 % (range, 96.3–100 %) for participant 3, 97.2 % (range, 91.7–100 %) for participant 4, 100 % for participant 5, 97.2 % (range, 91.7–100 %) for participant 6, 100 % for participant 7, and 90.2 % (range, 80.3–100 %) for participant 8.

Procedure

Experimental Design and Overview A non-concurrent multiple-probe design across participants was used to evaluate the effects of an LS training sequence on the emergence of novel intraverbal relations. Prior to training, participants took a visual-visual equivalence test over the to-be-trained and untrained relations. Following the initial visual-visual equivalence test, participants proceed to the baseline phase. During baseline, participants were exposed to intraverbal probes of all the relations to ensure that the intraverbal relations were not already in their repertoires. To control for additional baseline testing, some participants were exposed to one round of baseline probes, some participants were exposed to two rounds of baseline probes, and some participants were exposed to three rounds of baseline probes. Following baseline, intraverbal training commenced. Intraverbal training consisted of training the A-B and B-C intraverbals (see Table 1 for stimuli and Table 2 for training procedures). Following intraverbal training, intraverbal probes were conducted in the same manner as baseline intraverbal probes. Finally, participants were exposed to another visual-visual equivalence test.

Initial Visual-Visual Equivalence Test The initial visual-visual equivalence test was conducted on a laptop with the experimenter sitting next to the participant. Before the test, the experimenter informed the participant, “I am not going to tell you whether your responses are right or wrong, but when we are done you will get a bunch of nickels.” The experimenter would then start the power point presentation. Participants were provided with the following instructions on the first slide: In this part of the study, you will first be shown one box at the top of the screen. You will touch the box, and then you will see two boxes on the bottom of the screen. You will then touch a bottom box that you think best goes with the top box. Let us try it by clicking on the “Sample Trial” button! The participant was then provided with two sample trials consisting of familiar stimuli, in which feedback was given for each correct or incorrect response. Sample trials were conducted in the same manner as trials from the visual test (see below for an example of a trial from the visual test). Following the sample questions with feedback, participants were given the following instructions: Now this time, after you select a bottom box, you will not get feedback on if the answer was correct or incorrect. After you touch the bottom box, you will see a black screen, and then a new trial will start. Let us try it by touching the start button below! The participant was then provided with two sample trials consisting of familiar stimuli, in which no feedback was given for correct or incorrect responses. Following the sample questions with feedback, the participants were given the following instructions: If you have any questions, please ask the experimenter at this time. If you do not have any questions, then you are ready to start the study. You will not be told whether or not your responses are correct or incorrect. You will get some nickels after the test is finished! Press the start button when you are ready to begin. Once the participant pressed the start button, then the visual-visual equivalence test commenced.

The initial visual-visual equivalence test consisted of 12 trials that were presented in random order with one trial for each relation (i.e., A1-B1, A2-B2, B1-A1, B2-A2, A1-C1, A2-C2, C1-A1, C2-A2, B1-C1, B2-C2, C1-B1, C2-B2). Each trial commenced with a stimulus at the top of the screen. After the participant touched the stimulus, two

Table 2 Intraverbal (A-B and B-C) training procedures in experiment 1

Phase	Relations	Consequence of correct response	Consequence of incorrect response	Consequence of corrected response	Criterion for moving to next phase	Criterion for returning to a previous phase
1	A1B1	Praise + nickel	Vocal prompt and repeat trial	Praise	3 consecutive correct unprompted responses (phase 2)	16 consecutive trials with no more than 4 correct responses (phase 1)
2	A2B2	Praise + nickel	Vocal prompt and repeat trial	Praise	3 consecutive correct unprompted responses (phase 3)	
3	A1B1/A2B2	Praise + nickel	Vocal prompt and repeat trial	Praise	6 consecutive correct unprompted responses (phase 4)	
4	B1C1	Praise + nickel	Vocal prompt and repeat trial	Praise	3 consecutive correct unprompted responses (phase 5)	
5	B2C2	Praise + nickel	Vocal prompt and repeat trial	Praise	3 consecutive correct unprompted responses (phase 6)	
6	B1C1/B2C2	Praise + nickel	Vocal prompt and repeat trial	Praise	6 consecutive correct unprompted responses (phase 7)	16 consecutive trials with more than 4 incorrect responses (phase 4)
7	A1B1/A2B2/B1C1/B2C2	Praise + nickel	Vocal prompt and repeat trial	Praise	12 consecutive correct unprompted responses (phase 8)	32 trials with no upward trend or more than 8 incorrect responses (phase 1)
8	A1B1/A2B2/B1C1/B2C2	No feedback	Vocal prompt	–	12 consecutive correct unprompted responses (testing)	12 consecutive trials with 3 or more incorrect responses (phase 7)

stimuli were presented at the bottom of the screen. The participant then selected one of the two stimuli at the bottom of the screen. The test was conducted under extinction, so there was no feedback for correct or incorrect responses. Therefore, regardless of which stimulus the participant selected, a black screen was presented for 2 s, and then the next trial was presented. Nickels and praise were given to the participant after the test for completion of the test. The number of nickels delivered did not exactly correspond to the participants' performance during the 12-trial block; however, more or fewer nickels were awarded depending upon the number of correct or incorrect responses during the 12-trial block. For example, a participant would receive more nickels following a block in which 9 correct responses were emitted than a block in which 4 correct responses were emitted, which is consistent with previous research. The criterion for passing the visual-visual equivalence test was 10 out of 12 correct responses or approximately 83 % accuracy. Additionally, the 2 incorrect responses could not be from the same relation. If participants met the criterion on the initial visual-visual equivalence test, then they were excluded from the study.

Baseline Intraverbal Probes Each baseline intraverbal probe was one block, which consisted of 12 trials, with one trial for each relation shown in Table 1. All trials were presented in a random order, and all trials were conducted under extinction so no feedback was given for correct or incorrect responses. At the end of each probe block, the experimenter provided the participant with praise and nickels for participating. All tests were conducted with the experimenter sitting across the participant. Before each intraverbal probe, the participant was told that feedback would not be given and that he/she would receive nickels when the probe was finished. Each trial consisted of the experimenter vocally presenting one of the instructions and antecedent stimuli shown in Table 1. After the participant emitted a response or 5 s had elapsed, a new trial was presented. The criteria for passing a baseline intraverbal probe was 10 out of 12 correct responses or approximately 83 % accuracy and no 2 incorrect responses from the same relation. If a participant met the criteria in the baseline intraverbal probes, then that participant was excluded from the study. This did not occur for any of the participants.

Baseline Intraverbal Training Four relations were trained in baseline intraverbal training: A1-B1, A2-B2, B1-C1, and B2-C2. Training was conducted in eight phases. Table 3 shows the relations targeted in each phase, along with the consequences of correct and incorrect responses, and the performance criteria for advancing to the next phase or returning to a previous phase. A1-B1 and A2-B2 were introduced separately in phases 1 and 2, respectively. Phase 3 consisted of mixed A1-B1 and A2-B2 trials. Phases 4, 5, and 6 targeted B1-C1 and B2-C2 in a similar manner. In phase 7, all four relations were mixed and presented in four-trial blocks that contained one trial for each relation. Phase 8 was identical to phase 7, except that in preparation for testing, this phase was conducted under extinction. That is, no consequences were provided for correct or incorrect responses.

Each training trial began with the experimenter vocally presenting the appropriate antecedent instruction and stimulus (see Table 1). In phases 1 through 7, correct, unprompted responses were praised and the experimenter delivered a nickel to the participant. If the participant did not emit a response within 5 s or emitted an incorrect response, the experimenter provided the participant with the correct response, and then

Table 3 Trials to criterion in experiment 1

Participant	A-B	B-C	Mixed	Total
1	16	16	48	80
2	22	20	28	70
3	26	22	76	124
4	16	16	24	56
5	22	32	107	161
6	16	17	52(+24)	97
7	26	18	48	92
8	17	16	32	65

presented the next trial. In phases 1, 2, 4, and 5, in which new relations were introduced; an immediate vocal prompt was provided in the first two trials. The next two trials contained a 2-s prompt delay, and the remaining trials contained a 5-s prompt delay. In phases 3, 5, and 7, which contained mixed trials, all trials contained a 5-s prompt delay. Training in each phase continued until the participant either met the training criterion shown in Table 2 or met the criterion for returning to a previous phase.

Intraverbal Probes Intraverbal probes were conducted in the same manner as baseline intraverbal probes. Each intraverbal probe was one block, which consisted of 12 trials, with one trial for each relation (see Table 1). All trials were presented in a random order and conducted under extinction. If the trained relations (i.e., the A-B and B-C relations) were not passed, then the experimenter would go back to previous training phases (see Table 2 for criterion for return to specific phases). After training phases were passed, the intraverbal probe was once again conducted. This cycle continued until the trained relations were passed.

Final Visual-Visual Equivalence Test The final visual-visual equivalence test was conducted in the same manner as the initial visual-visual equivalence test, except that no sample questions were given at the beginning. The first slide on the power point provided the participants with the following instructions: “Welcome to the final phase in this study! In this part of the study, you will complete this task the same way as you completed the task in the beginning. You will first be shown one box at the top of the screen. You will touch the box, and then you will see two boxes on the bottom of the screen. You will then touch a bottom box that you think best goes with the top box. After you select a bottom box, you will not get feedback on if the answer was correct or incorrect. After you touch the bottom box, you will see a black screen, and then a new trial will start. If you have any questions, please ask the experimenter now. When you are ready to begin, touch the start button!” Once the participant pressed the start button, then the final visual-visual equivalence test commenced.

The final visual-visual equivalence test consisted of 12 trials that were presented in random order with one trial for each relation. The criteria for passing the visual-visual equivalence test was 10 out of 12 correct responses or approximately 83 % accuracy, and no two incorrect responses could be from the same relation.

Results

Figure 2 contains the individual data for the initial visual-visual equivalence test, the baseline intraverbal probes, the intraverbal probes following baseline intraverbal training, and the final visual-visual equivalence test for each participant. None of the participants met the criteria for the baseline intraverbal probes. Percentage correct during baseline intraverbal probes ranged from approximately 17 to 45 %, and baseline performance was not consistently increasing for any of the participants who were exposed to multiple rounds of baseline intraverbal probes. Following baseline intraverbal training, only two out of the eight participants (i.e., participants 2 and 4) met the criterion on the intraverbal probes. The other six participants did not meet the criterion on the intraverbal probes. Seven out of the eight participants maintained the baseline relations. Participant 6 did not maintain baseline relations; therefore, that

Fig. 2 The percentage of correct responses during probe sessions for experiment 1 for all probes except intraverbal probes following training in which baseline relations were not maintained. Participants with an “asterisk” next to their number designation did not meet the criterion during the intraverbal probes following intraverbal training. Performance on the trained relations (A-B, B-C) are depicted with *black color*, symmetry-like relations (B-A, C-B) are *gray*, and transitivity-like relations (A-C, C-A) are *white*. The *open circles* represent the visual-textual equivalence probes

participant received another round of baseline intraverbal training, followed by another intraverbal probe (see Table 3). Baseline relations were maintained during the second intraverbal probe test.

Figure 2 shows the outcomes of the visual-visual equivalence tests. During the initial test at the beginning of the experiment, none of the participants met the passing criteria. Percentage correct ranged between 17 and 50 %. On the final visual-visual equivalence test, only three out of the eight participants met the criteria (i.e., participants 1, 4, 7). Only one of the participants (participant 4) met the criteria on the intraverbal probe and the visual-visual equivalence test. One participant (participant 2) met the criteria on the intraverbal probe and did not meet the criteria on the visual-visual equivalence test. Two participants (participants 1 and 7) did not meet the criteria on the intraverbal probe and did meet the criteria on the visual-visual equivalence test. Finally, four participants (participants 3, 5, 6, and 8) did not meet the criteria on the intraverbal probe and the visual-visual equivalence test.

Table 3 shows the number of trials to criterion during intraverbal training. Total number of trials to criterion ranged from 56 to 161. Participant 6 was the only participant who did not meet the criteria on the baseline relations during the intraverbal probe, and therefore, required one additional round of mixed training.

Discussion

Overall, only two out of the eight participants demonstrated the emergence of novel intraverbals when baseline relations were trained in an LS training structure. These results were consistent with previous literature demonstrating that training the A-B and B-C intraverbals does not result in the emergence of novel intraverbals for the majority of participants (Carp & Petursdottir, 2012; Pérez-González et al., 2008). Additionally, an informal error analysis was conducted which demonstrated that for three out of the eight participants, the majority of errors were due the training contingencies establishing simple control. This pattern is consistent with the error analysis conducted in Carp and Petursdottir (2012); however, like Carp and Petursdottir, some errors made by other participants were not attributable to simple control, suggesting that other controlling variables may be in effect.

Finally, the visual-visual equivalence test showed that overall, only three out of the eight participants met the success criteria. In the stimulus equivalence literature, there is debate regarding the role of verbal behavior in the formation of equivalence classes (Hayes, Barnes-Holmes, & Roche, 2001; Horne & Lowe, 1996; Sidman, 1994). Horne and Lowe's (1996) naming hypothesis suggests that verbal behavior mediates the performance of equivalence classes. If that is the case, then we should see consistent results between the intraverbal probes and the visual-visual equivalence test. Results showed that for five out of the eight participants (i.e., participants 3, 4, 5, 6, and 8), the performance on the intraverbal probes corresponded to performance on the visual-visual equivalence test. This lends support to Horne and Lowe's (1996) naming hypothesis. For two participants (i.e., 1, 7), the criteria were not met on the intraverbal probe but were met on the visual-visual equivalence test. Although this seems to not support Horne and Lowe's naming hypothesis, it is possible that the participants who met the criteria on the visual-visual equivalence text but not the intraverbal probes had a sufficient number of statistical intraverbals in their repertoires to allow their verbal

behavior to mediate their performance. Furthermore, the textual visual stimuli could have prompted the intraverbals needed to mediate their performance. Finally, one participant (i.e., 2) met the criteria on the intraverbal probe but not on the visual-visual equivalence test, which may suggest that the participant was simply not attending to the visual stimuli and/or not engaging in the intraverbals when taking the visual-visual equivalence test.

One final interesting issue to address is that the participants who did not meet the criteria on both the intraverbal probes and the visual-visual equivalence test may still have had verbal behavior mediating performance on the visual-visual equivalence test. That is, the visual-visual equivalence test requires conditional discriminations, and if the LS only established intraverbals with simple control, then using those intraverbals to mediate performance on both the intraverbal probes and visual-visual equivalence tests would have been insufficient. Therefore, verbal behavior with faulty stimulus control could have mediated performance on the visual-visual equivalence test, resulting in poor performance.

Experiment 2

Due to the consistency between the results from experiment 1 and previous studies, the purpose of experiment 2 was to extend the previous research by investigating the emergence of novel intraverbals by using an OTM training sequence.

Methods

Participants

Participants included eight undergraduate students who were required to be at least 18 years old (or have parental consent before 18); no other restrictions were placed on participants. Participants could earn up to \$10 in addition for participation points in their undergraduate classes for their participation in the study.

Setting

Sessions were conducted in the same setting as described in experiment 1.

Stimuli and apparatus

Table 4 shows the antecedent verbal stimuli presented in baseline and test trials, along with correct vocal responses. Verbal stimuli were the same as in experiment 1, except that the A-B and A-C relations were trained instead of A-B and B-C relations. Stimuli for the visual-visual equivalence tests were the same as in experiment 1.

Measurement

Data collection and response measurement were conducted in the same manner as experiment 1. A second observer independently recorded data during at least 50 % of

Table 4 Antecedent stimuli and correct responses on baseline and test trials in experiment 2

	Antecedent stimuli	Correct response
A-B [P—RH (trained)]		
If <i>p</i> is [A1], then the results are what?	[A1] less than .05	[B1] significant
If <i>p</i> is [A2], then the results are what?	[A2] greater than .05	[B2] non-significant
A-C [P—Null (trained)]		
If <i>p</i> is [A1], then do what to the null?	[A1] less than .05	[C1] reject
If <i>p</i> is [A2], then do what to the null?	[A2] greater than .05 [	[C2] accept
B-A [RH—P]		
If the results are [B1], then <i>p</i> is what?	[B1] significant	[A1] less than .05
If the results are [B2], then <i>p</i> is what?	[B2] non-significant	[A2] greater than .05
C-A [Null—P]		
If the null is [C1], then <i>p</i> is what?	[C1] reject	[A1] less than .05
If the null is [C2], then <i>p</i> is what?	[C2] accept	[A2] greater than .05
B-C [RH—Null]		
If the results are [B1], then we do what to the null?	[B1] significant	[C1] reject
If the results are [B2], then we do what to the null?	[B2] non-significant	[C2] accept
C-B [Null—RH]		
If the null is [C1], then the results are what?	[C1] reject	[B1] significant
If the null is [C1], then the results are what?	[C2] accept	[B2] non-significant

all intraverbal probes, intraverbal training, and visual equivalence testing for all participants. IOA averaged 97.2 % (range, 94.4–100 %) for participant 9, 93.3 % (range, 88.2–100 %) for participant 10, 100 % for participant 11, 85.8 % (range, 80–91.7 %) for participant 12, 100 % for participant 13, 97.7 % (range, 95.8–100 %) for participant 14, 92.6 % (range, 93.8–100 %) for participant 15, and 97.2 % (range, 94.4–100 %) for participant 16.

Procedure

Experimental Design and Overview The experimental design and overview were the same as in experiment 1, except that instead of intraverbal training consisting of A-B and B-C training, it now consisted of A-B and A-C training (see Table 5). Therefore, participants were exposed to the initial visual-visual equivalence test, baseline intraverbal probe(s), intraverbal training (i.e., A-B, A-C), intraverbal probes, and then the final visual-visual equivalence test. All training, probe, and testing phases were conducted the same as in experiment 1.

Intraverbal Training Four relations were trained in baseline: A1-B1, A2-B2, A1-C1, and A2-C2. Training was conducted in a similar manner to experiment 1, except for the specific baseline relations trained (i.e., A-B and A-C instead of A-B and B-C). Table 5 shows the relations targeted in each phase, along with the consequences of correct and incorrect responses and the performance criteria for advancing to the next phase or returning to a previous phase.

Table 5 Intraverbal (A-B and A-C) training procedures in experiment 2

Phase	Relations	Consequence of correct response	Consequence of incorrect response	Consequence of corrected response	Criterion for moving to next phase	Criterion for returning to a previous phase
1	A1B1	Praise + nickel	Vocal prompt and repeat trial	Praise	3 consecutive correct unprompted responses (phase 2)	16 consecutive trials with no more than 4 correct responses (phase 1)
2	A2B2	Praise + nickel	Vocal prompt and repeat trial	Praise	3 consecutive correct unprompted responses (Phase 3)	
3	A1B1/A2B2	Praise + nickel	Vocal prompt and repeat trial	Praise	6 consecutive correct unprompted responses (phase 4)	
4	A1C1	Praise + nickel	Vocal prompt and repeat trial	Praise	3 consecutive correct unprompted responses (phase 5)	
5	A2C2	Praise + nickel	Vocal prompt and repeat trial	Praise	3 consecutive correct unprompted responses (phase 6)	
6	A1C1/A2C2	Praise + nickel	Vocal prompt and repeat trial	Praise	6 consecutive correct unprompted responses (phase 7)	16 consecutive trials with more than 4 incorrect responses (phase 4)
7	A1B1/A2B2 / A1C1/A2C2	Praise + nickel	Vocal prompt and repeat trial	Praise	12 consecutive correct unprompted responses (phase 8)	32 trials with no upward trend or more than 8 incorrect responses (phase 1)
8	A1B1/A2B2 /A1C1/A2C2	No feedback	Vocal prompt	–	12 consecutive correct unprompted responses (testing)	12 consecutive trials with 3 or more incorrect responses (phase 7)

Results

Figure 3 contains the individual data for the initial visual-visual equivalence test, the baseline intraverbal probes, the intraverbal probes following baseline intraverbal training, and the final visual-visual equivalence test for each participant in experiment 2. None of the participants met the criterion for the baseline intraverbal probes. Percentage correct during baseline intraverbal probes ranged from 18 to 60 %. Performance on baseline intraverbal probes increased consistently for participant 13, but not other participants who were exposed to multiple rounds of baseline intraverbal probes. Following baseline intraverbal training, six out of the eight participants (i.e.,

Fig. 3 The percentage of correct responses during probe sessions for experiment 2 for all probes except intraverbal probe sessions following training in which baseline relations were not maintained. Participants with *asterisk* next to the number designation did not meet the criterion during the intraverbal probes following intraverbal training. Performance on the trained relations (A-B, B-C) are depicted with *black color*, symmetry-like relations (B-A, C-B) are *gray*, and transitivity-like relations (A-C, C-A) are *white*. The *open circles* represent the visual-textual equivalence probes

participants 10, 11, 13, 14, 15, and 16) met the criterion on the intraverbal probes. The other two participants did not meet the criterion on the intraverbal probes. Five out of the eight participants maintained the baseline relations. Participants 10, 11, and 14 did not maintain baseline relations and subsequently received another round of baseline intraverbal training, followed by another intraverbal probe (see Table 6). Baseline relations were maintained during the second intraverbal probe test for all three participants.

Additionally, Fig. 3 shows the outcomes of the visual-visual equivalence tests. During the initial test at the beginning of the experiment, none of the participants met the criterion. Percentage correct ranged between approximately 22 and 65 %. On the final visual-visual equivalence test, all eight participants met the success criteria. In evaluating the correspondence between the intraverbal probes and the visual-visual equivalence test, six out of the eight participants (participants 10, 11, 13, 14, 15, and 16) met the criteria on the intraverbal probes and the visual-visual equivalence test. The other two participants (participants 9 and 12) did not meet the criteria on the intraverbal probes but did meet the criteria on the visual-visual equivalence test.

Table 6 shows the number of trials to criterion during intraverbal training. The total number of trials to criterion ranged from 73 to 180. Participants 10, 11, and 14 did not meet the criteria on the baseline relations during the intraverbal probe, and therefore, required one additional round of mixed training.

Discussion

Overall, six out of the eight participants demonstrated emergence of novel intraverbals. These results suggest that using a training sequence which requires conditional discriminations during training enhances the emergence of novel intraverbals for most participants. These results are consistent with the previous experimental research in stimulus equivalence consisting of visual stimuli which suggests that an OTM training structure is more effective than an LS training structure is in establishing equivalence classes (e.g., Hove, 2003).

Additionally, all eight participants met the criteria on the visual-visual equivalence tests, with six of those participants meeting the criteria on the intraverbal probes and two of those participants not meeting the criteria on the intraverbal probes. As stated previously, the six participants who met the criteria on both tests provides support for

Table 6 Trials to criterion in experiment 2

Participant	A-B	B-C	Mixed	Total
9	44	32	104	180
10	18	16	28(+40)	102
11	26	28	44(+40)	138
12	16	16	28	60
13	16	17	40	73
14	40	38	32(+28)	138
15	51	34	80	165
16	16	23	40	79

Horne and Lowe's (1996) naming hypothesis. Although two participants did not meet the criteria on the intraverbal probes but still met the criteria on the visual test, those participants could have had the sufficient intraverbals needed to pass the visual-visual equivalence test. Furthermore, the textual stimuli could have prompted the intraverbal behavior needed to have success on the test. The fact that all participants met the criteria on the visual test may demonstrate that appropriate conditional stimulus control was established during the baseline intraverbal probes, which in turn lead to success during the intraverbal probes and the visual-visual equivalence test.

General Discussion

The overall goal of the present study was to investigate the emergence of novel intraverbals by using EBI in which two training structures were evaluated (i.e., LS, OTM). Experiment 1 aimed to replicate previous literature by implementing an LS training structure. Only two of the eight participants demonstrated the emergence of novel intraverbals. These results were consistent with previous research in that when teaching intraverbals using an LS training structure, novel intraverbals do not emerge for most participants following baseline training alone (Carp & Petursdottir, 2012; Pérez-González et al., 2008). To date, there have only been two studies investigating EBI in which all sets of stimuli were intraverbals (Carp & Petursdottir, 2012; Pérez-González et al., 2008). Both of these studies tested typically developing children ages 4 to 7 and used intraverbals consisting of geographical locations such as countries/states, cities, and parks. Experiment 1 of the present study extended this line of research by (a) testing college students and (b) using intraverbals consisting of inferential statistical relations. Therefore, the results of experiment 1 replicate previous research with a different population and a different set of intraverbals. Experiment 2 aimed to extend previous research by implementing a different training sequence (i.e., OTM). Results showed that six of the eight participants who received the OTM training (A-B and A-C) demonstrated emergence of the novel intraverbals. Overall, the results of these two studies suggests that novel intraverbals are more likely to occur when baseline intraverbal relations are trained with an OTM stimulus equivalence training sequence compared to an LS. To our knowledge, this is the only study that has investigated (a) using different training structures to establish equivalence classes when all sets of stimuli (i.e., sets A, B, and C) consists of vocal-verbal stimuli and (b) teaching college students statistical relations in which all stimuli are vocal-verbal.

Although a comparison across experiments should be analyzed with caution, only two out of the eight participants' demonstrated emergence with the LS and six out of the eight participants demonstrated emergence with the OTM. The error analysis from the Carp and Petursdottir (2012) study demonstrated that using an LS training structure to train intraverbals may possibly result in only simple discriminations during training. In an informal error analysis for experiment 1, the majority of mistakes of at least three participants were consistent with training contingencies establishing simple discriminations. For example, during training, the participants were taught to respond "significant" to the antecedent "If p is less than .05, then the results are what?" Then, in testing, when the antecedent "If p is less than .05, then we do what to the null?" was presented, many participants responded with significant. The responses "significant/not

significant” were mostly controlled by “ $p > / < .05$ ” and the responses “reject/accept” were mostly controlled by significant/not significant. In experiment 2, six out of the eight participants passed the intraverbal probes following baseline intraverbal training alone. Therefore, it is possible that more novel intraverbals emerged due to the OTM training sequence establishing conditional discriminations in the baseline intraverbal relations. If the training contingencies established appropriate stimulus control, then the participants would be more likely to answer test questions involving conditional discriminations more accurately. Due to the low number of errors during the intraverbal probes in experiment 2, there was not a clear pattern of errors.

Overall, it took more trials to reach criterion in the OTM sequence than in the LS sequence. Additionally, only one participant in experiment 1 required additional baseline intraverbal training, whereas three participants in experiment 2 required additional baseline training. This finding is consistent with previous literature demonstrating that conditional discriminations involving some type of visual stimuli can be very hard to learn, especially in persons with developmental disabilities (Pérez-González & Williams, 2002; Saunders & Spradlin, 1989, 1990, 1993). Although it took more trials overall to establish baseline relations with conditional stimulus control compared to simple control, conditional control did result in more novel intraverbals. This study, however, did not address the efficiency of the two training sequences with regard to intraverbals. Because the research question was only to address the emergence of novel intraverbals following baseline training, it is not clear if the additional trials needed to train conditional discriminations during baseline training would be more efficient than training the LS with the category and exemplar training. Future research could investigate the efficiency of these two training sequences with regard to establishing novel intraverbals.

The visual-visual equivalence tests yielded interesting data. The criteria were met on the visual-visual equivalence test in only three out of the eight participants who were trained with the LS sequence and in all the eight participants who trained with the OTM sequence. This discrepancy may have occurred because the visual-visual equivalence test required conditional discriminations, and an OTM training sequence established those intraverbals involving conditional discriminations. That is, the OTM sequence may have possibly established intraverbals with appropriate stimulus control that is suggested to be needed to respond correctly during testing trials in the visual-visual equivalence test. The LS training sequence may have established intraverbals with simple control, and therefore, would not have established the appropriate verbal behavior to respond correctly during the visual-visual equivalence test.

The issue of whether verbal behavior is necessary for the emergence of equivalence classes has been debated in the literature (e.g., Horne & Lowe, 1996; Sidman, 2000). Horne and Lowe’s (1996) naming hypothesis suggests that verbal behavior is necessary for the formation of equivalence classes. If this is correct, then the participant’s performances on the final intraverbal probe and the final visual equivalence test would be predicted to correspond. That is, if the intraverbals were established, then any success criterion would be met on the visual equivalence test, and if the intraverbals were not established, then the criterion would not be met on the visual-visual equivalence test. If we take data from all participants across the two experiments, results showed that 11 out of the 16 participants demonstrated consistent performances between the intraverbal

probe and the visual-visual equivalence test, which would be consistent with Horne and Lowe's naming hypothesis. Out of the remaining five participants, four of them did not meet the criteria on the final intraverbal probe but did meet the criteria on the visual test. Although this may be inconsistent with Horne and Lowe's naming hypothesis, participants do not need all of the intraverbals in the intraverbal probe test to pass the visual-visual equivalence test. The intraverbals that were established during training may have been sufficient to guide responding during the visual test. Additionally, it is possible that the textual stimuli prompted the appropriate intraverbal behavior. Finally, one participant passed the intraverbal test but failed the visual test. It is possible that this participant did not attend to the visual stimuli due to fatigue or was selecting a visual stimulus due to visual properties of the stimuli as opposed to any type of verbal behavior.

Overall, the results of the previous study are consistent with the more basic research on stimulus equivalence training structures which have implemented at least one set of visual stimuli in the equivalence class. As stated previously, Arntzen (2012) found that the LS training structure (i.e., A-B and B-C training) was less effective in establishing equivalence classes compared to the OTM training structure (i.e., A-B and A-C training). Although more research needs to be conducted on the differences between visual and auditory stimuli in establishing equivalence classes, the present study is the first study to show that even when all sets of stimuli are auditory, the OTM is more effective with the emergence of novel intraverbals than the LS is. One might predict that the differences between the training structures could be more pronounced with auditory stimuli due to the fact that once they are presented, they "go away" in a sense. Visual stimuli remain present in the visual field as long as the experimenter presents the stimuli and the participants attends; therefore, it is possible that we might find more pronounced effects with specific types of stimuli. Future research could investigate this issue.

There is a growing body of applied literature which uses EBI to teach college students stimulus-stimulus relations in statistics and math, neuroanatomy, and single-subject designs (e.g., Albright, Reeve, Reeve, & Kisamore, 2015; Critchfield & Fienup, 2010; Fields et al., 2009; Fienup, Covey, & Critchfield, 2010; Fienup & Critchfield, 2010, 2011; Lovett et al., 2011; Ninness et al., 2005, 2006, 2009). Of the studies that have taught statistics and math to college students (i.e., Albright et al., 2015; Fienup & Critchfield, 2010, 2011; Ninness et al., 2005, 2006, 2009), all have used visual stimuli to present the statistical relations. The present study extends this line of research in college students by mainly investigating vocal-verbal inferential statistic stimuli. According to Skinner's Verbal Behavior (1957), intraverbals and textual are functionally independent in that the response emitted by the participant is under control of different variables due to a history of generalized conditioned reinforcement. Intraverbals are controlled by a verbal stimulus and textuials are controlled by a written stimulus. In a typical college classroom, lectures are a common means of instruction (Freeman et al., 2014). A traditional lecture involves the instructor speaking at the front of the room with visual stimuli (e.g., PowerPoint presentation) and the students sitting in their desks quietly listening. Schlinger (2008) suggests that "listening" refers to echoing what you hear to yourself. It is possible that students are not listening during lectures due to the fact that they are engaging in textual control by written

stimuli on the PowerPoint. Verbal behavior controlled by a written stimulus may be sufficient for written exams; however, it is also important for college students to be able to engage in vocal-verbal behavior about inferential statistics without any visual stimuli so that they can respond appropriately to another person's vocal-verbal behavior about statistics (e.g., a teacher's question). Due to the functional independence of verbal operants, teachers should not assume that textual behavior will result in the emergence of intraverbals. Future research could investigate the functional independence of these two verbal operants and the variables affecting that independence. Furthermore, students may only "listen" to certain vocal-verbal stimuli during a teacher's lecture, resulting in faulty stimulus control. Teachers may need to be careful in the way they word questions to ensure that students are listening to all relevant aspects of sentence thereby establishing appropriate stimulus control.

One limitation of the present study is the stimuli that were taught in that they were potentially already familiar stimuli. Socially relevant stimuli were chosen to (a) increase the ecological validity of the study and (b) connect the study to the previous research on teaching statistics to college students using EBI. Although these are beneficial reasons, choosing socially relevant stimuli increases the likelihood that participants could have had previous learning histories with the stimuli. Future research could determine participants' previous learning history with statistics or examine the effects of these two training procedures on the emergence of novel intraverbals with arbitrary stimuli. Additionally, the intent to focus on the different training structures necessitated keeping the intraverbal stimuli consistent with previous research and equal across the relations. Therefore, information pertaining to the p value being equal to .05 was not incorporated. Future applied research could incorporate all of the relevant information within the stimuli. Another limitation to the previous study is the limited number of baseline intraverbal probes conducted. In the present study, each intraverbal was only presented once during a probe. It is possible that not enough intraverbals were presented during a probe session to really assess the performance on the probes. In previous research (e.g., Carp & Petursdottir, 2012), a probe session consisted of 48 trials in which each intraverbal relation was presented 4 times. Previous research, however, was conducted with children in which the experimenter worked with them for months at a time rather than the 1-h session conducted with these college students. Regardless of the small number of trials in the probe sessions, stable baselines were evidence across participants with a difference in percentage correct on the intraverbal probes following the two different training sequences. Future research could replicate and extend these findings by conducting additional intraverbal probes following baseline intraverbal training.

In conclusion, the present study demonstrates that when using an EBI to investigate the emergence of novel intraverbals, the specific training structure used affects the emergence. Specifically, this study suggests that an OTM training structure results in the emergence of more novel intraverbals compared to an LS training structure. This study extends the previous literature on intraverbals and stimulus equivalence by testing different stimuli, participants, and training structures. Finally, this study extends the research on using stimulus equivalence to teach relations to college students by using all vocal-verbal stimuli.

Acknowledgments This experiment was conducted in partial fulfillment of the requirement of the MA degree by the first author. We thank Josh Charpentier for helping with data collection and Cam Melville who provided assistance and support that made this study possible.

Compliance with ethical standards All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

Funding This study was not funded by any grants.

Conflict of interests The authors declare that they have no conflict of interest.

References

- Albright, L., Reeve, K. F., Reeve, S. A., & Kisamore, A. N. (2015). Teaching statistical variability with equivalence-based instruction. *Journal of Applied Behavior Analysis*, 48, 883–894. doi:10.1002/jaba.249.
- Arntzen, E. (2012). Training and testing parameters in formation of stimulus equivalence: methodological issues. *European Journal of Behavior Analysis*, 13, 123–135.
- Arntzen, E., & Holth, P. (2000). Differential probabilities of equivalence outcome in individual subjects as a function of training structure. *The Psychological Record*, 50, 603–628.
- Carp, C. L., & Petursdottir, A. I. (2012). Effects of two training conditions on the emergence novel intraverbals: an extension of Pérez-González et al. (2008). *The Psychological Record*, 62, 187–206.
- Critchfield, T. S., & Fienup, D. M. (2010). Using stimulus equivalence technology to teach statistical inference in a group setting. *Journal of Applied Behavior Analysis*, 43, 763–768. doi:10.1901/jaba.2010.43-763.
- Fields, L., Travis, R., Roy, D., Yadlovker, E., Aguiar-Rocha, L., & Sturmey, P. (2009). Equivalence class formation: a method for teaching statistical interactions. *Journal of Applied Behavior Analysis*, 42, 575–593. doi:10.1901/jaba.2009.42-575.
- Fienup, D. M., & Critchfield, T. S. (2010). Efficiently establishing concepts of inferential statistics and hypothesis decision making through contextually controlled equivalence classes. *Journal of Applied Behavior Analysis*, 43, 437–462. doi:10.1901/jaba.2010.43-437.
- Fienup, D. M., & Critchfield, T. S. (2011). Transportability of equivalence based programmed instruction: efficacy and efficiency in a college classroom. *Journal of Applied Behavior Analysis*, 44, 435–450. doi:10.1901/jaba.2011.44-435.
- 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, 19–33. doi:10.1901/jaba.2010.43-763.
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performances in science, engineering, and mathematics. *PNAS*, 111, 8410–8415. doi:10.1073/pnas.1319030111.
- Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.) (2001). *Relational frame theory: a post-Skinnerian account of human language and cognition*. New York: Plenum.
- Horne, P. J., & Lowe, C. F. (1996). On the origins of naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior*, 65, 185–241. doi:10.1901/jeab.1991.56-139.
- Hove, O. (2003). Differential probability of equivalence class formation following a one-to-many versus a many-to-one training structure. *The Psychological Record*, 53, 617–634.
- Lovett, S., Rehfeldt, R. A., Garcia, Y., & Dunning, J. (2011). Comparison of stimulus equivalence protocol and traditional lecture for teaching single-subject designs. *Journal of Applied Behavior Analysis*, 44, 819–833. doi:10.1901/jaba.2011.44-819.
- 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, 1–22. doi:10.1901/jaba.2005.2-04.

- Ninness, C., Barnes-Holmes, D., Rumph, R., McCuller, G., Ford, A. M., Payne, R., Ninness, S. K., Smith, R. J., Ward, T. A., & Elliott, M. P. (2006). Transformations of mathematical and stimulus functions. *Journal of Applied Behavior Analysis*, 39, 299–321. doi:10.1901/jaba.2006.139-05.
- Ninness, C., Dixon, M., Barnes-Holmes, D., Rehfeldt, R. A., Rumph, R., McCuller, G., Holland, J., Smith, R., Ninness, S. K., & McGinty, J. (2009). Constructing and deriving reciprocal trigonometric relations: a functional analytic approach. *Journal of Applied Behavior Analysis*, 42, 191–208. doi:10.1901/jaba.2009.42-191.
- Pérez-González, L. A., & Williams, G. (2002). Multicomponent procedure to teach conditional discriminations to children with autism. *American Journal on Mental Retardation*, 107, 293–301. doi:10.1352/0895-8017.
- Pérez-González, L. A., García-Asenjo, L., Williams, G., & Carnerero, J. J. (2007). Emergence of intraverbal antonyms in children with pervasive developmental disorder. *Journal of Applied Behavioral Analysis*, 40, 697–703. doi:10.1901/jaba.2007.697-701.
- Pérez-González, L. A., Herszlikowicz, K., & Williams, G. (2008). Stimulus relations analysis and the emergence of novel intraverbals. *The Psychological Record*, 58, 95–129.
- Petursdottir, A. I., Carr, J. E., Lechago, S. A., & Almason, S. M. (2008). An evaluation of intraverbal training and listener training for teaching categorization skills. *Journal of Applied Behavior Analysis*, 41, 53–68. doi:10.1901/jaba.2008.41-53.
- Saunders, R. R., & Green, G. (1999). A discrimination analysis of training-structure effects on stimulus equivalence outcomes. *Journal of Experimental Analysis of Behavior*, 72, 117–137. doi:10.1901/jeab.1999.72-117.
- Saunders, R. R., & McEntee, J. E. (2004). Increasing the probability of stimulus equivalence with adults with mild mental retardation. *The Psychological Record*, 54, 423–435.
- Saunders, K. J., & Spradlin, J. E. (1989). Conditional discrimination in mentally retarded adults: the effect of training the component simple discriminations. *Journal of the Experimental Analysis of Behavior*, 52, 1–12. doi:10.1901/jeab.1989.52-1.
- Saunders, K. J., & Spradlin, J. E. (1990). Conditional discrimination in mentally retarded adults: the development of generalized skills. *Journal of the Experimental Analysis of Behavior*, 54, 239–250. doi:10.1901/jeab.1990.54-239.
- Saunders, K. J., & Spradlin, J. E. (1993). Conditional discrimination in mentally retarded subjects: programming acquisition and learning set. *Journal of the Experimental Analysis of Behavior*, 60, 571–585. doi:10.1901/jeab.1993.60-571.
- Saunders, R. R., Wachter, J. A., & Spradlin, J. E. (1988). Establishing auditory stimulus control over an eight-member equivalence class via conditional discrimination procedure. *Journal of the Experimental Analysis of Behavior*, 49, 95–115. doi:10.1901/jeab.1988.49-95.
- Saunders, R. R., Chaney, L., & Marquis, J. G. (2005). Equivalence class establishment with two-, three-, and four-choice matching to sample by senior citizens. *The Psychological Record*, 55, 539–559.
- Schlinger, H. D. (2008). Listening is behaving verbally. *The Behavior Analyst*, 31, 145–161.
- Sidman, M. (1994). *Equivalence relations and behavior: a research story*. Boston, MA: Authors Cooperative.
- Sidman, M. (2000). Equivalence relations and the reinforcement contingency. *Journal of Experimental Analysis of Behavior*, 74, 127–146. doi:10.1901/jeab.2000.74-127.
- Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: an expansion of the testing paradigm. *Journal of Experimental Analysis of Behavior*, 37, 5–22. doi:10.1901/jeab.1982.37-5.
- Skinner, B. F. (1957). *Verbal behavior*. New York, NY: Apple-Century-Crofts.
- Spradlin, J. E., & Saunders, R. R. (1986). The development of stimulus classes using match-to-sample procedures: sample classification versus comparison classification. *Analysis and Intervention in Developmental Disabilities*, 6, 41–58.
- Spradlin, J. R., Cotter, V. W., & Baxley, N. (1973). Establishing a conditional discrimination without direct training: a study of transfer with retarded adolescents. *American Journal of Mental Deficiency*, 77, 556–566.
- Toussaint, K. A., & Tiger, J. H. (2010). Teaching early braille literacy skills within a stimulus equivalence paradigm to children with degenerative visual impairments. *Journal of Applied Behavior Analysis*, 43, 181–194. doi:10.1901/jaba.2010.43-181.
- Walker, B. D., & Rehfeldt, R. A. (2010). Using the stimulus equivalence paradigm to teach course material in an undergraduate rehabilitation course. *Journal of Applied Behavior Analysis*, 43, 615–633. doi:10.1901/jaba.2010.43-615.
- Walker, B. D., & Rehfeldt, R. A. (2012). An evaluation of the stimulus equivalence paradigm to teach single-subject design to distance education students via blackboard. *Journal of Applied Behavior Analysis*, 45, 329–344. doi:10.1901/jaba.2012.45-329.