JOURNAL OF APPLIED BEHAVIOR ANALYSIS 2011, **44,** 819–833 NUMBER 4 (WINTER 2011)

COMPARISON OF A STIMULUS EQUIVALENCE PROTOCOL AND TRADITIONAL LECTURE FOR TEACHING SINGLE-SUBJECT DESIGNS

SADIE LOVETT, RUTH ANNE REHFELDT, YORS GARCIA, AND JOHNNA DUNNING

SOUTHERN ILLINOIS UNIVERSITY

This study compared the effects of a computer-based stimulus equivalence protocol to a traditional lecture format in teaching single-subject experimental design concepts to undergraduate students. Participants were assigned to either an equivalence or a lecture group, and performance on a paper-and-pencil test that targeted relations among the names of experimental designs, design definitions, design graphs, and clinical vignettes was compared. Generalization of responding to novel graphs and novel clinical vignettes, as well as the emergence of a topography-based tact response after selection-based training, were evaluated for the equivalence group. Performance on the paper-and-pencil test following teaching was comparable for participants in the equivalence and lecture groups. All participants in the equivalence group showed generalization to novel graphs, and 6 participants showed generalization to novel clinical vignettes. Three of the 4 participants demonstrated the emergence of a topography-based tact response following training on the stimulus equivalence protocol.

Key words: stimulus equivalence, college students, tact, verbal behavior, topography-based responding

Behaviorally based instructional protocols grounded in the principles of stimulus equivalence can provide instructors with an alternative method of communicating their subject matter of interest to students. The hallmark of stimulus equivalence protocols is that direct training on certain relations among instructional stimuli will result in the emergence of untrained relations among those stimuli (Sidman, 1994). Early research that examined this phenomenon focused on teaching reading comprehension and oral reading skills to individuals with intellectual disabilities (Sidman, 1971). For example, Sidman and Cresson (1973) used a stimulus equivalence protocol to teach an individual to read three-letter words, such as "cow." The learner was first trained to relate the spoken word "cow" to a picture of a cow and to relate the spoken word "cow" to the printed word *cow*. After training, the learner was then able to orally name the picture of the cow and the printed word *cow,* and he was able to relate the picture of the cow to the printed word *cow* without a direct history of reinforcement for relating those stimuli. According to the stimulus equivalence paradigm, the oral naming of the picture and printed word, which involves a reversal of the trained relation, is referred to as *symmetry*. The relation of the printed word and the picture that had never been previously paired is referred to as *transitivity*.

Instruction using the principles of stimulus equivalence has been applied successfully in a variety of situations. Cowley, Green, and Braunling-McMorrow (1992) taught adults with acquired brain injuries to relate the dictated names of their therapists to photographs of the therapists and to relate the dictated therapist names to the written therapist names. Participants then related the photographs to the written therapist names and orally named the therapists when shown the photographs without further training. An investigation by Lynch and Cuvo (1995) used a stimulus equivalence protocol to teach fifth- and sixth-grade students relations between pictorial representations of fractions and numerical fraction ratios and relations between

Address correspondence to Ruth Anne Rehfeldt, Behavior Analysis and Therapy Program, Rehabilitation Institute, Southern Illinois University, Carbondale, Illinois 62901 (e-mail: rehfeldt@siu.edu).

doi: 10.1901/jaba.2011.44-819

printed decimals and pictorial representations of fractions. Following training, students related the numerical fraction ratios to the appropriate printed decimals. A recent study by Toussaint and Tiger (2010) examined the use of a stimulus equivalence protocol to teach braille literacy to children with degenerative visual impairments. Participants who were able to relate a spoken letter name to the printed letter were trained to relate braille letters and the corresponding printed letters. Participants then were able to select the appropriate braille letter when provided with the spoken letter name and orally name the braille letters.

These examples of the application of stimulus equivalence in teaching various skills highlight its utility as an instructional method. Studies of stimulus equivalence frequently involve training and testing with a conditional discrimination procedure that relies on selection-based responding, or pointing to one stimulus that is presented in an array (Michael, 1985). The emergence of topography-based responding, or responding in a different topography than that trained, also has been documented (Michael, 1985). Following selection-based training, Cowley et al. (1992) demonstrated the emergence of a tact response, which is a verbal response under the control of a nonverbal stimulus (Skinner, 1957), by showing that participants were able to name the photographs of the therapists. Furthermore, Toussaint and Tiger (2010) documented a topography-based tact response by demonstrating the emergence of oral naming of the braille letters. The topography-based responding shown in these studies may reflect the acquisition of more meaningful verbal repertoires than selection-based responding produces (Sundberg & Sundberg, 1990), because skills commonly targeted in education, such as speaking and writing, require topography-based rather than selection-based responding. A successful educational curriculum will result in proficiency in both spoken and written topography-based response repertoires.

In recent years, stimulus equivalence protocols have been extended to the instruction of sophisticated learners (e.g., Fienup & Critchfield, 2010). Ninness et al. (2005, 2006) used computer-based stimulus equivalence protocols to teach algebraic and trigonometric mathematical functions. Participants were taught to relate standard mathematical formulas to factored formulas and to relate factored formulas to graphical representations of those formulas. After training, participants related the standard formulas and the graphs without direct training, and generalization to novel formulas and graphs was shown. Fields et al. (2009) taught statistical interaction concepts by training college students to relate line graphs and textual descriptions of interactions, textual descriptions of interactions and labels of the interactions, and interaction labels and definitions of each type of interaction. The emergence of untrained relations among the stimuli was shown, as was generalization of responding to novel variations of the trained stimuli. In addition, a study by Fienup, Covey, and Critchfield (2010) used stimulus equivalence to teach relations among regions of the brain, anatomical locations of brain regions, and psychological function of brain regions to college undergraduates. Finally, Walker, Rehfeldt, and Ninness (2010) taught college students to relate the names, definitions, causes, and treatments for disabilities. In addition to demonstrating the emergence of untrained relations, this study showed the emergence of written and vocal topography-based responding following selection-based training.

The aforementioned research indicates that instruction with stimulus equivalence protocols is successful in teaching relations among stimuli, including complex stimuli that are important in training advanced learners. However, there has been little effort to compare the efficacy of stimulus equivalence protocols to standard educational practices or to assess the social validity of the instructional method. One exception is Fields et al. (2009), who evaluated generalization to a paper-and-pencil posttest following computer-based training. The use of worksheets and paper tests, as opposed to computer-based procedures, more closely approximates the materials and procedures present in the average classroom. Fields et al. also included a questionnaire to assess the social validity of the instructional method (see also Fienup & Critchfield, 2011).

Further investigations of the efficacy and acceptability of stimulus equivalence protocols in instruction relative to standard educational practices would be beneficial in determining the utility of the method. The objectives of the present study were thus as follows: First, we compared the effectiveness of a stimulus equivalence protocol to that of a lecture in teaching undergraduate students concepts of single-subject experimental design. The protocol established relations among the names of designs, their definitions, representative graphs, and clinical vignettes in which the use of a design might be appropriate. Specifically, we compared performances for the two groups of participants on a paper-and-pencil test. Second, we evaluated the efficacy of the stimulus equivalence protocol by assessing generalization of the relations to novel graphs and clinical vignettes and the emergence of a topography-based repertoire. Finally, we investigated the social validity of the stimulus equivalence protocol and the lecture by administering a satisfaction questionnaire to participants in both conditions at the end of the experiment.

METHOD

Participants, Setting, and Apparatus

Twenty-four undergraduate students who were currently enrolled in a research methods course participated in this experiment. All participants received extra credit as compensation for participation. Sessions were conducted in an office (3 m by 5 m) that contained two desks, each with a chair and personal computer. The participant was seated at one desk, and the experimenter was seated at the second desk, throughout the session. All procedures were conducted on the computer, with the exception of the paper-and-pencil quiz and social validity survey. Sessions ranged in length from 75 to 140 min.

Equivalence Stimuli

Each of four stimulus classes contained four stimuli that were presented during training and two stimuli that were used to test for generalization. Stimuli were developed based on the definitions, graphical illustrations, and clinical examples of single-subject designs in an undergraduate textbook (Kennedy, 2005). The A stimuli were the names of each of the four basic single-subject designs. The B stimuli were definitions of the four corresponding designs. The C stimuli were graphs depicting the implementation of each of the four single-subject designs. Each graph had unique *y*-axis labels (e.g., percentage aggressive behavior) and unique intervention phase labels (e.g., reinforcement). The D stimuli were vignettes that described clinical situations in which the four corresponding designs would be appropriate for use. All vignettes were designed such that the length of each description was similar. The vignettes described clinical situations that were distinct from the behaviors and interventions illustrated on the graphical (C) stimuli. The B and D stimuli are presented in Figure 1 (see also Walker & Rehfeldt, in press).

There were four C′ generalization stimuli including one novel graphical representation of each of the single-subject designs. The novel graphs showed changes in behavior in the opposite direction from that of the stimulus used in training. For example, the C1 stimulus depicted a graph for a withdrawal design that showed an increase in the percentage correct responding, and the C′1 stimulus was a novel graph of a withdrawal design that showed a decrease in aggressive behavior. Novel graphs also included different *y*-axis and phase labels. The four D′ generalization stimuli were novel clinical vignettes that described situations in which each of the four designs would be

B	D
B1	D1
This design involves evaluating the effects of a treatment by implementing and then removing the treatment when its removal does not present risks to the client.	Teachers want to evaluate the effectiveness of noncontingent reinforcement on reducing Robin's occasional talking out of turn in her third-grade class.
B2	D2
This design involves evaluating the effects of a treatment via the staggered implementation of the treatment across two or more behaviors, clients, or settings, and is used when it is unfeasible or unethical to remove the treatment.	Staff want to evaluate the effectiveness of a DRA procedure on the reduction of Billy's elopement (running away) at school, the grocery store, and the shopping mall.
B3	D3
This design involves evaluating the effects of two or more treatments on the same behavior via the rapid alternation of treatments within sessions, across different times of the same day, or across different days.	Staff want to evaluate whether verbal praise or a token system is more effective at increasing the duration of Carla's time on task.
B4	D4
This design involves evaluating the effects of a treatment on the gradual, systematic increase or decrease of a single target behavior by changing, in a stepwise fashion, the criterion levels necessary for reinforcement.	Therapists want to evaluate the effectiveness of a smoking cessation program by allowing their client to quit smoking gradually, completing small goals along the way.

Figure 1. B and D stimuli for each of the four stimulus classes.

appropriate for use. Length of the descriptions in the vignettes was similar to that of the vignettes used in training. Like the C′ generalization stimuli, the vignettes described a change in behavior opposite of the direction of behavior change on the stimulus used in training. For example, the D1 stimulus described a situation in which a teacher wished to decrease talking out in class, and the D′1 stimulus described a situation in which self-monitoring was used to increase the number of math problems completed.

General Procedure

This experiment used a pretest–train–posttest sequence and included two groups of participants. Participants in the equivalence group were exposed to a computer-based stimulus equivalence protocol, and participants in the lecture group viewed a video of a lecture that provided an overview of the four basic single-subject designs. Participants were assigned randomly to either the equivalence or lecture group by the flip of a coin. Nine participants were included in the lecture group, and 15 participants were included in the equivalence group. The number of participants in the two groups was unequal because several participants assigned to the lecture group cancelled sessions. Additional participants could not be recruited for this group because course material had progressed to the point of covering the single-subject design topics targeted in this experiment by that time in the semester. The main dependent variable was performance on the paper-and-pencil quiz, and all participants completed a questionnaire to evaluate the social validity of the instructional methods.

Paper-and-pencil pretest and posttest evaluation. The paper-and-pencil quiz consisted of 15 multiple-choice questions on single-subject designs. Each question included four response options (a, b, c, and d). Questions included adaptations of the stimuli presented in the stimulus equivalence protocol and lecture. Four questions tested variations of the definition-to-design-name (B-A) relations by requiring participants to select the correct design name in the presence of a modified version of the design definition. Four questions required participants to select the appropriate design name in the presence of a novel graph (C-A relation). Four questions required participants to select the design name in the presence of a novel clinical vignette (D-A relation). Three questions required participants to select the appropriate definitional design feature in the presence of the design name (A-B relation). A sample quiz item that required the participant

1. The graph below represents an evaluation of an intervention using which design?

a. Multiple baseline
b. Changing criterion
c. Alternating treatments
d. Withdrawal

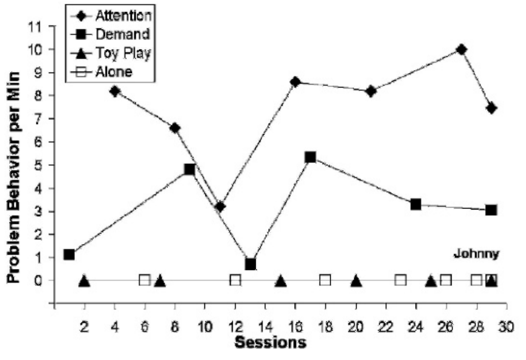

Figure 2. Sample question testing a C-A relation from the paper-and-pencil quiz.

to select the appropriate design name in the presence of a novel graph is depicted in Figure 2. The quiz was administered at the start of the experiment as a pretest, and an identical quiz was administered as a posttest after completion of the stimulus equivalence protocol or presentation of the lecture for the equivalence and lecture groups, respectively. Content of quiz questions was validated by a professor of behavior analysis with expertise in single-subject design methodology.

Interobserver agreement. Interobserver agreement on quiz scores was collected by an independent reviewer. Agreement was calculated for 33% of pretests and 33% of posttests for both the equivalence and lecture groups. Interobserver agreement scores were calculated by dividing the number of item-by-item agreements by the total number of agreements plus disagreements and multiplying that value by 100%. Item-by-item agreement for both pretests and posttests for the equivalence group was 100%. Mean agreement for pretests for the lecture group was 98% (range, 93% to 100%). Item-by-item agreement for posttests for the lecture group was 100%. Interobserver agreement was not conducted on responding during the stimulus equivalence protocol because all procedures and data collection were automated.

Social validity survey. A questionnaire that assessed participants' opinions of the respective instructional method was administered at the end of the experiment. The survey included four questions that participants rated on a 7-point Likert-type scale, with higher ratings indicating a more positive evaluation of the instructional method. Questions inquired as to the participant's confidence in his or her knowledge of single-subject designs, the degree to which he or she would prefer to be taught using the particular instructional method, and the participant's opinions on the time commitment for instruction. The survey is shown in Table 1.

Lecture Group

Following the paper-and-pencil pretest, participants were seated at desks and were told that they were going to view a video on single-subject designs. Up to two participants were present to view the video during a session; however, participants were seated at separate desks while completing the paper-and-pencil test. The video was presented on the computer and showed a 56-min lecture with an accompanying PowerPoint presentation that provided an overview of the four basic single-subject designs. Video content included an introduction to single-subject

Table 1
Social Validity Survey

How confident do you feel in your knowledge of single-subject designs?						
1	2	3	4	5	6	7
Not at all confident			Somewhat confident			Very confident
Rate the degree to which you would prefer to be taught using this instructional method.						
1	2	3	4	5	6	7
Don't prefer at all			Somewhat prefer			Strongly prefer
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

design terminology (e.g. independent variables, dependent variables, and functional relations) and an overview of the basic components of a graphical display (e.g. axes, phases, and data paths). The majority of the lecture was divided into four sections that corresponded to withdrawal, multiple baseline, alternating treatments, and changing criterion designs. Each section provided a definition of the design, showed a basic graphical display of the design, and presented an example of the application of the design in an applied setting with an accompanying graph. Lecture content was derived from the presentation of single-subject design material in two undergraduate textbooks (Kennedy, 2005; Richards, Taylor, Ramasamy, & Richards, 1999). After the video, the paper-and-pencil posttest was administered, followed by the social validity survey.

The paper-and-pencil pretest and posttest as well as the social validity survey were identical to those that were used with the equivalence group. These two measures were the only points of comparison between the two groups.

Equivalence Group

Training for this group consisted of a computer-based stimulus equivalence protocol programmed using Microsoft Visual Basic 2008 Express Edition. Training and test trials were presented in a match-to-sample format with one sample stimulus at the top of the screen and four comparison stimuli at the bottom of the screen on each trial. For example, on a trial that examined the A1-B1 relation, the A1 stimulus was presented at the top of the screen, and all four B stimuli appeared at the bottom of the screen. Trials in all training and testing phases were presented in random order, and the order in which the comparison stimuli were presented on the screen was randomized. During training, correct responses were followed by written and auditory feedback in the form of the word "correct" and the chime sound from the Windows operating system. Incorrect responses were followed by written and auditory feedback in the form of the word "incorrect" and the chord sound from the Windows operating system. No feedback was provided during testing. After completion of the paper-and-pencil pretest, participants were seated at the computer and read the following instructions on the screen:

> Thank you for participating in this experiment. Your job during this experiment is to do the best that you can at all times. One box will be presented at the top of the screen, and four boxes will be presented below

> it. Your job will be to choose one of the boxes at the bottom of the screen. During certain portions of the experiment you will receive feedback as to whether your choice was correct or not, but during other portions of the experiment you will not receive feedback. Please do the best you can at all times regardless of whether or not you receive feedback. Click on the button below to start.

Pretest for equivalence and generalization relations. This test consisted of 48 trials that evaluated equivalence (definition-to-vignette [B-D] and vignette-to-definition [D-B]) and generalization (design-name-to-novel-graph [A-C′] and design-name-to-novel-vignette [A-D′]) relations. There were 12 trials for each type of relation (e.g., B-D), and each individual relation (e.g., B1-D1) was presented three times. No feedback was provided following responses on this test.

Training and symmetry tests. Participants first were trained on the design-name-to-definition (A-B) relations. The design names appeared as the sample stimuli, and the design definitions appeared as comparison stimuli. A training block consisted of 12 trials, with each individual relation (e.g., A1-B1) presented three times. Responses were followed by written and auditory feedback. Participants repeated the training phase until they met the criterion of 11 of 12 correct (92%). After meeting the criterion for design-name-to-definition (A-B) training, participants advanced to the definition-to-design-name (B-A) symmetry test, which included 12 trials. No feedback was provided after responses during this test, and the criterion was 11 of 12 correct. If the criterion was not met on the definition-to-design-name (B-A) symmetry test, the participant returned to design-name-to-definition (A-B) training. When the criterion in training again was reached, the participant proceeded to the definition-to-design-name (B-A) symmetry test, and this process was repeated until he or she achieved the criterion on the test.

After passing the definition-to-design-name (B-A) symmetry test, training began on the design-name-to-graph (A-C) relations. The design names appeared as the sample stimuli, and graphs appeared as comparison stimuli. Training was conducted in the same manner as for the design-name-to-definition (A-B) relations, with a training block containing 12 trials and each relation (e.g., A1-C1) presented three times. After meeting the criterion of 11 of 12 trials correct, participants advanced to the graph-to-design-name (C-A) symmetry test. This test was similar to the definition-to-design-name (B-A) symmetry test and included 12 trials, with a mastery criterion of 11 of 12 correct. If the criterion was not attained on the graph-to-design-name (C-A) symmetry test, design-name-to-graph (A-C) training was repeated, and the graph-to-design-name (C-A) symmetry test again was administered until mastery was achieved.

Following the graph-to-design-name (C-A) symmetry test, the design-name-to-vignette (A-D) relations were trained in the same manner as the design-name-to-definition (A-B) relations. The design name was presented as the sample stimulus, and the clinical vignettes appeared as comparison stimuli. A training block contained 12 trials with each relation (e.g., A1-D1) presented three times, and the criterion for advancing was 11 of 12 correct. After achieving mastery on the design-name-to-vignette (A-D) relations, the vignette-to-design-name (D-A) symmetry test was presented in the same manner as the definition-to-design-name (B-A) symmetry test. The test included 12 trials with a mastery criterion of 11 of 12 correct. If criterion was not achieved, design-name-to-vignette (A-D) training was repeated, and the vignette-to-design-name (D-A) symmetry test was presented again until mastery was attained.

Mixed symmetry test. This test included 36 trials that evaluated the definition-to-design-name (B-A), graph-to-design-name (C-A), and vignette-to-design-name (D-A) symmetry relations. As in the previous tests, each relation (e.g., B1-A1) was presented three times. Trials were presented in random order, and the mastery criterion was 33 of 36 correct (91.7%). No feedback was provided following responses. Participants continued to the next test phase regardless of performance.

Transitivity test. This test consisted of 48 trials that evaluated the definition-to-graph (B-C), graph-to-definition (C-B), graph-to-vignette (C-D), and vignette-to-graph (D-C) transitive relations. Each relation (e.g., B1-C1) was presented three times, and the criterion was 44 of 48 trials correct (91.7%). Participants continued to the next test after completion of all 48 trials regardless of performance.

Equivalence test. This test included 24 trials that evaluated the definition-to-vignette (B-D) and vignette-to-definition (D-B) equivalence relations. Each relation (e.g., B1-D1) was presented three times, and the criterion was 22 of 24 trials correct (91.7%). Participants continued to the next test regardless of performance.

Generalization test. The first test for generalization included 12 trials that evaluated the design-name-to-novel-graph (A-C′) relations. The design name was presented as the sample stimulus, and the novel graphs appeared as comparisons. Each relation (e.g., A1-C′1) was presented three times, and the criterion was 11 of 12 trials correct (91.7%). No feedback was provided following responses, and participants continued to the subsequent generalization test regardless of performance.

The second test for generalization evaluated the design-name-to-novel-vignette (A-D′) relations and was presented in the same manner as the test for design-name-to-novel-graph (A-C′) relations. The design name appeared as the sample stimulus, and the novel vignettes were presented as comparisons. This test included 12 trials with each relation (e.g., A1-D′1) presented three times, and no feedback was provided after responses. Following completion of the design-name-to-novel-vignette (A-D′) generalization test, the computer program ended, and the paper-and-pencil posttest was administered for 11 of the 15 participants in this group. The remaining four participants continued to the tact test.

Tact test. The tact test was conducted in two blocks using flash cards presented by the experimenter. Before beginning the tact test, the experimenter read the following instructions to the participant: "I'm going to show you some cards, and I'd like you to tell me which experimental design the picture or description on the card represents. I won't tell you whether your responses are correct or incorrect, but try your best." An individual trial consisted of the experimenter showing the participant one card and asking, "What design is this?" The participant was allowed 10 s to review the stimulus and respond. No feedback was provided after responses, and if the participant failed to tact the stimulus within 10 s, the next trial was presented.

The first tact-test block included 24 trials that evaluated the graph-to-design-name (C-A) and vignette-to-design-name (D-A) relations. The graph (C) and vignette (D) stimuli used during training on the stimulus equivalence protocol were individually presented on cards (7 cm by 10 cm). Trials were presented in a predetermined random sequence, and each relation (e.g., C1-A1) was presented three times. Criterion for mastery was 22 of 24 correct (91.7%), and the second test block immediately followed regardless of performance. The second tact-test block consisted of 24 trials that tested the novel-graph-to-design-name (C′-A) and novel-vignette-to-design-name (D′-A) generalization relations, and it was conducted in the same manner as the tact test for the trained stimuli. The novel graph (C′) and novel clinical vignette (D′) stimuli were presented on flash cards. Criterion for mastery was 22 of 24 correct (91.7%). Regardless of performance, the paper-and-pencil posttest was administered after the tact test.

Interobserver agreement on the tact test was collected by an independent observer for 50% of sessions. Interobserver agreement scores were calculated by dividing the number of trial-by-trial agreements by the total number of agreements plus disagreements and multiplying that value by 100%. Trial-by-trial agreement was 100% for all sessions.

RESULTS

Session Length

Participants in the equivalence group spent an average of 85 min completing the stimulus equivalence protocol (range, 62 to 120 min). All participants in the lecture group spent 56 min in the instructional portion of the experiment viewing the recorded lecture.

Paper-and-Pencil Evaluation

Results for the paper-and-pencil quiz for both the equivalence and lecture groups are presented in Table 2. For the equivalence group, mean quiz scores were 7.4 points (SD = 1.5) on the pretest and 10.4 points (SD = 2.3) on the posttest. For the lecture group, mean quiz scores were 7.0 points (SD = 1.3) on the pretest and 9.4 points (SD = 2.7) on the posttest. The average increase in scores from pretest to posttest was 2.9 points (SD = 2.3) for the equivalence group and 2.4 points (SD = 2.4) for the lecture group. Bonferroni adjusted alpha levels of .017 per test were used to conduct statistical analyses. An independent groups t test revealed no significant difference between the posttest means of the lecture and equivalence groups, $t(21) = 0.9$, p =.392. Within-group t tests revealed significant differences between pretest and posttest for both the equivalence group, $t(13) = 4.8$, $p < .001$, and the lecture group, $t(8) = -3.1$, $p = .016$.

Table 2
Quiz Scores

Group	Participant	Pretest	Posttest	Pretest to posttest change
Equivalence	1	8	13	5
	2	7	12	5
	3	7	13	6
	4	7	12	5
	5	8	9	1
	6	6	11	5
	7	8	10	2
	8	9	8	−1
	9	6	8	2
	10	5	5	0
	12	8	9	1
	13	7	12	5
	14	7	11	4
	15	11	12	1
Lecture	16	7	13	6
	17	6	10	4
	18	6	10	4
	19	6	10	4
	20	9	12	3
	21	8	10	2
	22	8	9	1
	23	5	4	−1
	24	8	7	−1

Note. The highest possible score was 15.

Social Validity Survey

The average rating regarding participants' confidence in knowledge of single-subject designs was 4.3 (range, 3 to 6) and 4.6 (range, 3 to 6) for the equivalence and lecture groups, respectively. The degree to which participants would prefer to receive instruction like that of the experimental procedures received an average rating of 4.3 (range, 2 to 7) for the equivalence group and 4.8 (range, 1 to 7) for the lecture group. The appropriateness of the time commitment for the amount that was learned received an average rating of 5.1 (range, 3 to 7) for both equivalence and lecture groups. The appropriateness of the length of the instructional method received an average rating of 4.9 (range, 2 to 7) for the equivalence group and 5.3 (range, 3 to 7) for the lecture group. The average overall sum score for all items on the rating scale was 18.7 (range, 12 to 25) and 19 (range, 13 to 26) for the equivalence and lecture groups, respectively, with higher ratings signifying a more positive review of the instructional method.

Equivalence Class Formation

The number of training and symmetry test blocks to criterion is depicted in Table 3. Fourteen participants in the equivalence group met criterion for all training sessions and successfully passed the symmetry tests for the individual relations. Participant 11 did not complete design-name-to-graph (A-C) training, and her participation ended after 54 trial blocks. Therefore, there are no data for that participant for the remaining analyses.

Table 3
Number of Training and Symmetry Testing Blocks to Criterion

Participant	A-B Train	B-A Test	A-C Train	C-A Test	A-D Train	D-A Test
1	3	1	2	1	2	1
2	3	1	2	1	1	1
3	1	1	3	1	2	1
5	3	1	3	1	3	1
4	4	1	4	1	3	1
6	3	2	3	1	3	1
7	8	1	4	2	3	1
8	4	1	3	1	2	1
9	3	3	4	1	3	1
10	4	1	4	1	4	1
11	16	1	54[a]			
12	2	2	2	1	3	1
13	1	1	1	1	1	1
14	3	1	3	1	4	1
15	2	1	2	1	3	1

[a] The participant did not meet criterion.

Results for the stimulus equivalence protocol pretests and posttests for trained and generalization stimuli are presented in Figure 3. No participants met criterion on the equivalence or generalization pretests. All participants reached criterion on the mixed test for symmetry (definition-to-design-name [B-A], graph-to-design-name [C-A], and vignette-to-design-name [D-A] relations) after training. Except for Participants 4, 9, and 10, all participants reached criterion on the test for transitivity (definition-to-graph [B-C], graph-to-definition [C-B], graph-to-vignette [C-D], and vignette-to-graph [D-C] relations). Participants 4 and 9 were below criterion by one trial on the test for transitivity, and Participant 10 was below criterion by two trials. Eleven of the 14 participants achieved mastery criterion on the equivalence posttest (definition-to-vignette [B-D] and vignette-to-definition [D-B] relations). All 14 participants met criterion on the design-name-to-novel-graph (A-C′) generalization test. Only 6 of the 14 participants (Participants 1, 2, 3, 4, 12, and 13) achieved criterion on the design-name-to-novel-vignette (A-D′) generalization test.

Tact test. Results for the tact test are depicted in Figure 4. Three of the four participants met criterion on the tact test for the graph-to-design-name (C-A) relations using stimuli that were presented during training. Two participants achieved mastery criterion on the test for the vignette-to-design-name (D-A) relations.

Three of the four participants attained criterion on the novel-graph-to-design-name (C′-A) generalization relations. Two participants met criterion on the novel-vignette-to-design-name (D′-A) generalization relations. Participant 15 was the only one of the four participants exposed to this condition who did not meet criterion on the tact test using trained or generalization stimuli.

DISCUSSION

A stimulus equivalence protocol was designed to teach undergraduate students concepts of single-subject experimental designs. Results of the current study extend previous findings on the use of equivalence-based instruction with advanced learners by demonstrating the emergence of relations among stimuli that were not directly trained (Fields et al., 2009; Fienup et al., 2010; Ninness et al., 2005, 2006; Walker et al., 2010). As a result of this experiment, participants in the equivalence group were able to inspect graphical stimuli and make inferences about single-subject designs. This study also included more complex stimuli than the stimuli

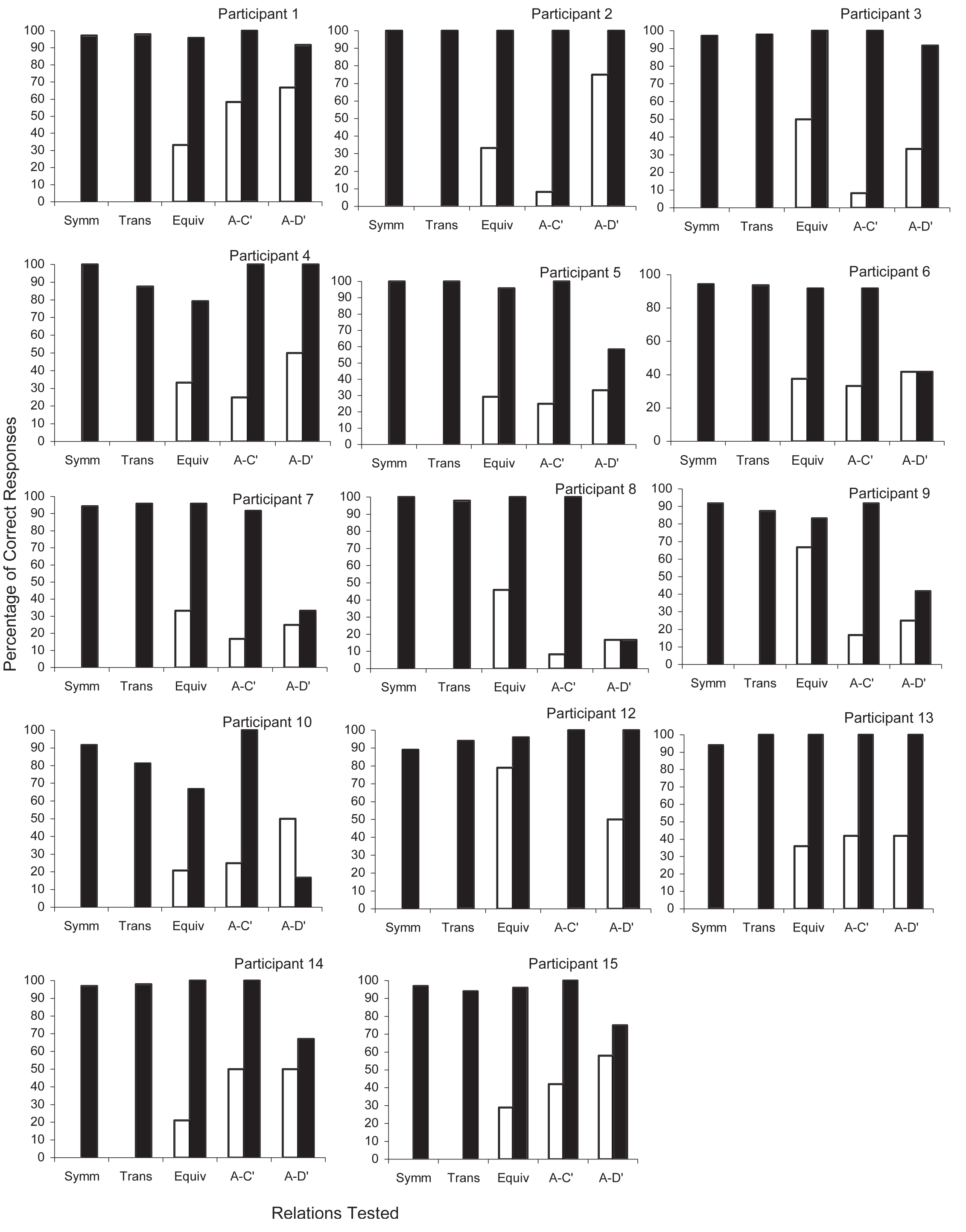

Figure 3. Posttest scores for symmetry and transitivity relations and pretest and posttest scores on equivalence and generalization relations for the equivalence group.

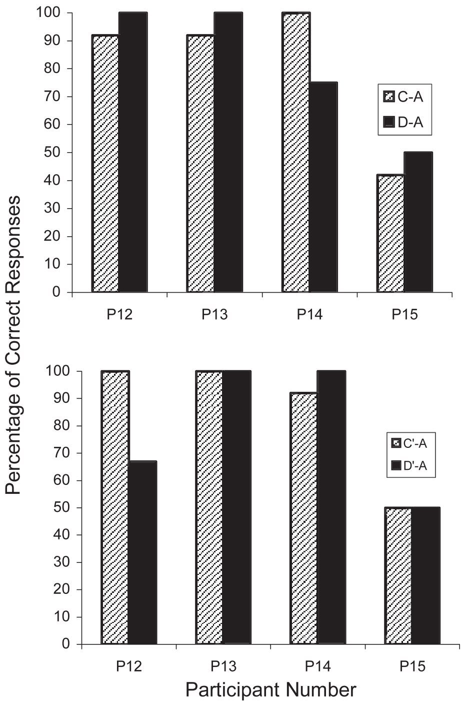

Figure 4. Tact test results for Participants 12, 13, 14, and 15. Relations that involved trained stimuli are depicted in the top graph, and relations that involved generalization stimuli are depicted in the bottom graph.

used in stimulus equivalence research with special populations. Previous research has examined relations that involved simpler stimuli, such as pictures, letters, numbers, or single words (Cowley et al., 1992; Lynch & Cuvo, 1995; Toussaint & Tiger, 2010). The lengthier descriptions provided in the design definitions and clinical vignettes in this study are more appropriate for a college-level student and more reflective of complex relational skills. All participants showed generalization of responding to novel graphical stimuli, and six of the 14 participants in the equivalence group showed generalization to novel clinical vignettes. Three of four participants were able to tact both training and generalization stimuli at the conclusion of the experiment.

Comparisons of the effectiveness of the stimulus equivalence protocol to a traditional, lecture-teaching format revealed similar performance on paper-and-pencil quizzes and similar ratings regarding participants' opinions of the two instructional methods, despite the fact that there was a 30-min difference between the two procedures. Only two prior studies addressed participants' satisfaction with instruction using stimulus equivalence. Fields et al. (2009) found that participants exposed to a stimulus equivalence protocol provided higher ratings of both their understanding of statistical interactions and their satisfaction with the instructional method compared to a control group that received no instruction (see also Fienup & Critchfield, 2011). Surprisingly, the present findings failed to show significantly higher ratings for equivalence protocol instruction over lecture instruction. *The Technology of Teaching* (Skinner, 1968) leads us to expect that participants would prefer instruction that involves an active response component. According to Skinner (1968), learning does not take place when a student is merely shown or told information, as is the case in a traditional lecture (p. 103). Results from this study indicate that, from the student's perspective, passive learning (i.e., being shown or told information in a lecture) is equally as desirable as active learning (i.e., the stimulus equivalence protocol). A potential explanation for this finding is that each participant experienced only one instructional method. A lecture format is the typical teaching procedure used in a college setting. Without an alternative with which to compare it, a student may rate that method as desirable because it is the standard practice to which they are accustomed. Skinner also maintained that recall of information must be directly reinforced for learning to occur. Responding for the equivalence group was directly reinforced during training, yet performance on the paper-and-pencil quiz did not surpass that of the lecture group. It is possible that a college student with a lengthy learning history of exposure to lecture teaching is better able to attend to the important aspects of the lecture (i.e., noting the repetition of the lecturer or the words in boldface on the slides).

Also of consideration are the possible limitations of the lecture condition regarding the degree to which it approached a truly traditional lecture. In the typical research methods course from which participants were recruited, approximately 50 students are present for lecture. With a class this size, potentially distracting stimuli are present (e.g., laptops, comments from peers, less relevant comments from the instructor) that can divert student attention from lecture material. In this study, lecture sessions included only one or two students, and this may have resulted in enhanced effects over the traditional lecture that is presented to a large group of students. However, online distance learning courses often include recorded lectures that students watch individually on the computer, and the lecture condition in this study closely resembles the conditions under which a student enrolled in a distance education course might view a lecture. The lecture used in this study also differed from a traditional lecture because it was specifically designed for this experiment and was tailored to present the relations targeted in the equivalence protocol. Modeling the lecture material on the equivalence protocol may have increased the likelihood that participants would learn the relations among the stimuli from lecture alone. Considering these limitations of the lecture condition, a comparison of performance following the equivalence protocol with a more naturalistic lecture delivered to a large class warrants investigation (see also Critchfield & Fienup, 2010). Gains in performance after training on the equivalence protocol may exceed performance gains after this form of traditional lecture.

One of the key contributions of this study is the demonstration of the emergence of a topography-based repertoire after selection-based training. To date, few studies have examined the emergence of topography-based responding more complex than naming pictures or reading sight words in equivalence instruction (Cowley et al., 1992; Toussaint & Tiger, 2010; Walker et al., 2010). The topography-based tact response required in this study was of greater complexity than that seen in previous research, in that participants were required to inspect and interpret the graphs and read and interpret each clinical vignette. Furthermore, the emergence of a topography-based repertoire was shown for the generalization stimuli as well as the trained stimuli. These findings suggest that this method of instruction may result in the emergence of a form of responding that is more functional in daily life than emergent selection-based responses alone (Michael, 1985; Sundberg & Sundberg, 1990). Selection-based responses, such as pointing at or selecting stimuli, are not functional for a college student who must be able to read, describe, and discuss material to attain the competency to practice in his or her future profession. This study is a small step in the direction of providing scientific instructional methods that will produce such complex repertoires.

Conclusions that can be drawn from the results of the tact test in this study are limited, because no pretest for the tact responses was included. However, all participants in the equivalence group received low pretest scores for the equivalence and generalization relations, and it is unlikely that they would have named the experimental design correctly when shown a graph or clinical vignette if they were not able to select the appropriate design name during the selection-based pretest. Moreover, the topography-based repertoire was not evaluated for the lecture group, and it remains unclear if this procedure would have resulted in the emergence of a tact repertoire.

Another interesting aspect of our results concerns the tests for generalization. All participants in the equivalence group showed generalization to novel graphs, but only six participants showed generalization to novel clinical vignettes. This finding is not surprising, given the physical similarity between the graphical stimuli. The differences among the appearances of phase-change lines, the presence of several tiers in the multiple baseline designs, the intervention phase with multiple data paths in the alternating treatments design, and the gradually changing criteria of the changing criterion design provide more salient visual cues than the descriptions of the clinical vignettes. In addition to containing less salient visual cues, the vignettes required the participant to attend to multiple aspects of the description. Choice of the appropriate design was dependent on attention to the number of behaviors, participants, and settings in the description; the manner in which behavior was changed (i.e., gradually or not); and the number of interventions. Due to the complexity of the vignettes, additional training may have been necessary for participants to show generalization to novel stimuli. Training with multiple exemplars of stimuli previously has been shown to be effective in promoting generalization. Ninness et al. (2005) included multiple exemplars of graphs depicting mathematical functions during training, and responding generalized to over 40 novel graphs. A similar use of training with multiple exemplars could enhance discrimination of the relevant aspects of the clinical vignettes used in this study.

Another difficulty with the vignettes is the fact that participants could successfully complete training and testing for equivalence relations without reading the full descriptions. For example, several vignettes included a name for the client that differed across vignettes. A participant could respond accurately by attending to the client name only. Training with multiple exemplars of clinical vignettes could help to overcome this difficulty. By requiring the participant to respond appropriately to several different examples of vignettes, the features important to choosing the appropriate design may become apparent. An alternative method of ensuring attendance to the relevant aspects of the stimuli is to include verbal rules that describe the stimuli and the relations among them (e.g., Ninness et al., 2005).

The paper-and-pencil quiz used as a pre- and posttest also can be viewed as a test for generalization (Fields et al., 2009). Each quiz question included either a variation of the design definition, a novel graph, or a novel clinical vignette. Four of the 15 quiz questions included a novel clinical vignette and required the participant to select the appropriate design name. For the equivalence group, the lack of generalization to novel vignettes in the stimulus equivalence protocol likely was reflected in quiz scores and affected the differentiation in posttest scores for the equivalence and lecture groups. Given the generalization failures and lack of group differences on the paper-and-pencil quiz, one might question the rationale for continued investigations of stimulus equivalence protocols in higher education if this instructional method does not surpass that of standard educational practices. This issue is particularly important given that the lecture condition presented more information and required less time for participants to complete. Nonetheless, the lecture may well have required more preparation time on part of the instructor than did the equivalence condition. In fact, the additional preparation time required for this lecture may not differ significantly from the time required to design a lecture that incorporates an active student response component, such as response cards (Marmolejo, Wilder, & Bradley, 2004) or guided notes (Neef, McCord, & Ferreri, 2006). However, the likelihood of an instructor adopting a lecture format that requires greater response effort remains a question for future research. Furthermore, the lack of difference between the lecture and equivalence groups need not devalue the pursuit of research on stimulus equivalence in education. A multitude of different forms of teaching are likely effective, and instructors can benefit from having a variety of teaching methods available. Because each student has a unique learning history, it also may be the case that some students will benefit more from lecture teaching and others will succeed using an equivalence protocol. In addition, the two approaches used in the present study easily could be used as complementary teaching strategies.

REFERENCES

Cowley, B. J., Green, G., & Braunling-McMorrow, D. (1992). Using stimulus equivalence procedures to teach name-face matching to adults with brain injuries. *Journal of Applied Behavior Analysis*, *25*, 461–475.

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.

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*, 575–593.

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.

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.

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.

Kennedy, C. H. (2005). *Single-case designs for educational research*. Boston: Allyn & Bacon.

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

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

Michael, J. (1985). Two kinds of verbal behavior plus a possible third. *The Analysis of Verbal Behavior*, *3*, 1–4.

Neef, N. A., McCord, B. E., & Ferreri, S. J. (2006). Effects of guided notes versus completed notes during lectures on students' quiz performance. *Journal of Applied Behavior Analysis*, *39*, 123–130.

Ninness, C., Barnes-Holmes, D., Rumph, R., McCuller, G., Ford, A. M., Payne, R., et al. (2006). Transformations of mathematical and stimulus functions. *Journal of Applied Behavior Analysis*, *39*, 299–321.

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.

Richards, S. B., Taylor, R. L., Ramasamy, R., & Richards, R. Y. (1999). *Single subject research: Applications in educational and clinical settings*. Belmont, CA: Wadsworth Cengage Learning.

Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech and Hearing Research*, *14*, 5–13.

Sidman, M. (1994). *Equivalence relations and behavior: A research story*. Boston: Authors Cooperative.

Sidman, M., & Cresson, O. (1973). Reading and crossmodal transfer of stimulus equivalence in severe retardation. *American Journal of Mental Deficiency*, *77*, 515–523.

Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.

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

Sundberg, C. T., & Sundberg, M. L. (1990). Comparing topography-based verbal behavior with stimulus selection-based verbal behavior. *The Analysis of Verbal Behavior*, *8*, 31–41.

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.

Walker, B. D., & Rehfeldt, R. A. (in press). Using the stimulus equivalence paradigm to teach single-subject designs to college students via Blackboard. *Journal of Applied Behavior Analysis*.

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*, 615–633.

Received July 26, 2010
Final acceptance March 28, 2011
Action Editor, Joel Ringdahl