USING THE STIMULUS EQUIVALENCE PARADIGM TO TEACH COURSE MATERIAL IN AN UNDERGRADUATE REHABILITATION COURSE

BROOKE D. WALKER AND RUTH ANNE REHFELDT
SOUTHERN ILLINOIS UNIVERSITY

AND

CHRIS NINNESS
STEPHEN F. AUSTIN STATE UNIVERSITY

In 2 experiments, we examined whether the stimulus equivalence instructional paradigm could be used to teach relations among names, definitions, causes, and common treatments for disabilities using a selection-based intraverbal training format. Participants were pre- and posttested on vocal intraverbal relations and were trained using multiple-choice worksheets in which selection-based intraverbal relations were taught and feedback was delivered until mastery. Most participants in Experiment 1 showed the emergence of vocal intraverbal relations, but responding did not generalize to final written intraverbal tests. Participants in Experiment 2 showed the emergence of not only vocal intraverbal relations but also written intraverbal relations on final tests. Results suggest that the stimulus equivalence paradigm can be effectively implemented using a selection-based intraverbal training format, the protocol may be an effective means of emphasizing important concepts in a college course, and emergent skills may generalize to novel response topographies.

Key words: derived stimulus relations, generalization, higher education, stimulus equivalence, verbal behavior
words with movable letters. Following this instruction, the children read 51 words and showed improvement in spelling, and some children read novel words that were recombinations of the within-syllable units of the training words (de Rose et al.). Rosales and Rehfeldt (2007) established derived mands in adults with severe intellectual disabilities by first teaching individuals to mand for items needed to complete chained tasks by exchanging pictures. Participants were then taught to relate conditionally the pictures to their dictated names and the dictated names to the corresponding printed words. As a result, participants exchanged text for the items needed to complete the chained task in the absence of explicit instruction. One month later, some of the participants even vocally requested the needed items.

This body of work underscores the role that a curricular approach based on derived stimulus relations might have in a variety of educational and habilitative settings. In many studies on derived stimulus relations, the ultimate performance inferred to indicate the emergence of new stimulus relations is match to sample, which involves selection-based responding or the selection of stimuli from an array (see Michael, 1985). However, these studies also demonstrated the emergence of untrained skills of relevance to participants’ everyday lives (i.e., naming therapists, spelling, manding; see Cowley et al., 1992). Moreover, in the case of Cowley et al. and de Rose et al. (1996), the untrained skills were also topography based or had topographies that were unique from one another and differed from the response topography established during training (Michael). Topography-based responding has been argued to better reflect verbal behavior according to the Skinner (1957) conceptualization (see C. T. Sundberg & Sundberg, 1990), as well as to occur more frequently in everyday life than selection-based responding (e.g., Perez-Gonzalez, Herszlikowicz, & Williams, 2008).

Recently, pioneering efforts by several researchers have highlighted a possible role for the derived stimulus relations paradigm in teaching course material in the college classroom (e.g., Fields et al., 2009; Fienup, Covey, & Critchfield, 2010; Ninness et al., 2005, 2006, 2009). Studies by Ninness et al. served as the impetus for this line of research; Ninness et al. (2005) used a computer-based procedure to enhance knowledge of algebraic and trigonometric transformations in participants who lacked training in advanced mathematical functions. Specifically, participants were taught to relate standard to factored formulas and factored formulas to graphs. Most participants demonstrated the emergence of untrained relations among the stimuli, and responding generalized to include 40 different and complex variations of the original training formulas and graphs (Ninness et al., 2005; see also Ninness et al., 2006, 2009). Another study in this vein is that of Fields et al., who applied the derived stimulus relations protocol to the concept of statistical interaction effects in two-way analyses of variance. Participants in this study were undergraduate introductory psychology students, and stimuli included graphical representations of the types of interactions (e.g., crossover, synergistic, divergent, and no interaction; the A stimuli), written descriptions of the types of interactions (the B stimuli), the names of the interactions (the C stimuli), and definitions of the interactions (the D stimuli). After participants completed a computerized match-to-sample procedure that consisted of matching the A and B stimuli, followed by the B and C stimuli, followed by the C and D stimuli, participants not only showed the emergence of four four-member equivalence classes among the stimuli but also responded relationally with novel examples of original training stimuli. Participants also showed improved scores on a multiple-choice paper-and-pencil test relative to their pretest scores and scores for a control group of participants.
who did not participate in the experiment. The study by Fienup et al. (2010) also used equivalence-based instruction to establish relations between the anatomical location of various brain functions and the actual functions of those locations with introductory psychology students. These authors administered the instructional protocols as a series of computerized lessons, ultimately showing that the application of instructional protocols based on derived stimulus relations can be applied to college instruction as a form of programmed instruction, with numerous opportunities for students to respond and receive feedback. Thus, continued investigation of the efficacy of derived relations instructional protocols in the college classroom seems warranted, particularly protocols that arrange for the emergence of skills in response topographies such as speaking and writing, which have relevance for students beyond their completion of the experiment. It may also be worthwhile to explore other delivery modes for situations in which the use of computerized protocols is infeasible. For example, the procedure developed by Fienup et al. may well lend itself to completion in a large laboratory in an undergraduate psychology course in which students complete many computerized exercises over the course of a semester. Unfortunately, this arrangement, although ideal, may not be feasible or even available for other courses.

An apparatus that has received little attention in the research literature, but which may have high face validity in the college classroom, is the paper-and-pencil format used by Eikeseth, Rosales-Ruiz, Duarte, and Baer (1997) to investigate the effects of written instructions on equivalence class formation. In this study, sample and comparison stimuli were presented on training worksheets, and participants were instructed to make matches by circling the comparison stimulus that went with each sample stimulus, much like a multiple-choice exam. Although this study used Roman letters as stimuli in lieu of socially relevant stimuli, the authors accentuated the advantages of this training format by describing match to sample as an “elaborate and rigorous training protocol” (Eikeseth et al., p. 276) that was actually originally developed for research with pigeons, then later modified for use with persons with mental retardation or young children. The paper-and-pencil format used by these authors can be conceptualized as a protocol for teaching selection-based intraverbal relations or relations between verbal stimuli and verbal responses in which the responses share no point-to-point correspondence with the stimuli that evoke them (Skinner, 1957). The response of circling the correct comparison stimulus is selection based (Michael, 1985), as is a match-to-sample response. This training format may be cost effective in situations in which computers are not readily available: Numerous participants can complete a protocol simultaneously using inexpensive materials (Eikeseth et al.), and the training format has high face validity in that it is one that is used in many educational settings. An equivalence worksheet is similar in form to the many worksheets, homework assignments, and exercises students complete in and out of class. To date, the use of a paper-and-pencil training format in establishing derived stimulus relations in an educational curriculum has not been evaluated (see also Smeets, Dymond, & Barnes-Holmes, 2000).

The purpose of the present investigation was to extend the use of instructional protocols that program for the emergence of derived stimulus relations in the college classroom. In this study, three protocols were implemented that corresponded with the three units of an undergraduate course on disabilities. For each unit’s protocol, three worksheets were used to establish stimulus equivalence relations that included the names of a disability, the definition of a disability, the disability’s primary cause, and an effective treatment or service for the particular disability. Training thus established selection-
based intraverbal relations. Vocal intraverbal posttests evaluated participants' naming of a particular disability when presented with a flash card stating its definition, cause, or an effective service or treatment. Students enrolled in the course served as participants for each instructional protocol in Experiment 1. For these students, final tests were administered at the conclusion of the semester. Final tests were shortanswer and fill-in-the-blank formats and required students to supply either the name of a disability when given its definition, cause or an effective treatment or service or the definition, cause, and an effective treatment or service when given the disability name. Like vocal intraverbal posttests, final tests evaluated the emergence of intraverbal responses, but in this case the topography of the target responses was written rather than vocal (see also Chase, Johnson, & Sulzer-Azaroff, 1985). Participants in Experiment 2 were not students in the course, and for these students, the final written intraverbal tests were administered immediately following the completion of vocal intraverbal posttests. All emergent performances targeted in this experiment were thus topography based.

EXPERIMENT 1

Method

Participants. A total of 22 undergraduate students enrolled in an introductory course on disabilities at Southern Illinois University participated in the study during the Fall 2008 semester. All students in the course were at different stages of their undergraduate academic careers and varied in the extent of their prior knowledge of various disabilities. A total of 13 students participated in each of the three unit protocols (Table 1), but the same 13 students did not necessarily participate in each unit protocol. Students were compensated with course credit for participating.

Setting and stimulus materials. Sessions lasted approximately 30 to 45 min and took place in a small laboratory (2.7 m by 3.3 m) that contained two chairs and a desk. Each participant sat in one chair at the desk, and
the experimenter sat directly across from him or her.

Training materials included multiple-choice questionnaires or worksheets (three for each unit), printed on two pieces of white paper (21.6 cm by 27.9 cm) in black Calabri font (Size 11). Each worksheet consisted of 12 items. The question presented in each item was conceptualized as the sample stimulus, and the four response options were conceptualized as comparison stimuli. Each question was presented three times on each worksheet, with the order of questions randomly determined before the worksheets were constructed. The order of response options for a given question was also randomly determined. The order in which questions and response options were presented was derived via a random sequence generator (http://www.random.org/sequences/). Test materials for each unit consisted of three sets of 12 white index cards (10.2 cm by 15.2 cm) with definitions, causes, and treatments for disabilities printed in black Calibri font (Size 20) on the front of the cards. Disability names were printed on the back of each card. The four questions that were presented on each of the three worksheets for Units 1, 2, and 3, respectively, are shown in Appendixes A, B, and C.

Training and testing materials for each unit included four stimuli each. Stimuli were designated with the following alphanumeric symbols: names of disabilities (A stimuli), definitions of disabilities (B stimuli), primary causes of the disabilities (C stimuli), and common treatments or services for the disabilities (D stimuli). The combination of the A, B, C, and D stimuli constituted a class. Each unit’s protocol was thus intended to result in some (but not all) of the emergent relations (B-A, C-A, and D-A) indicative of the formation of four four-member stimulus classes, with each class pertaining to a particular disability. The primary textbook, which was used throughout all three units of the course and from which stimulus sets were obtained, was Bowe’s (2000) *Physical, Sensory, and Health Disabilities: An Introduction*. Stimulus sets incorporated into each unit protocol were as follows (Table 2): The Unit 1 protocol included cerebral palsy, spina bifida, multiple sclerosis, and muscular dystrophy (physical and developmental disabilities); the Unit 2 protocol included cystic fibrosis, sickle-cell disease, hypertension, and diabetes (health disabilities); and the Unit 3 protocol included apraxia, schizophrenia, Alzheimer’s disease, and dementia (psychiatric and age-related disabilities).

**Design.** A pretest-train/posttest-maintenance test sequence was used in all three unit protocols. Participants completed one unit protocol in its entirety in one session. In each session, participants first completed B-A (naming a disability when given its definition), C-A (naming a disability when given its primary cause), and D-A (naming a disability when given a common treatment or service) vocal intraverbal pretests and then received worksheet training on the A-B (name to definition) relations. The B-A vocal intraverbal posttest was administered immediately following mastery of the A-B relations. The A-C (name to primary cause) relations were then trained using the worksheets, and immediately following their mastery, the C-A vocal intraverbal posttest was administered. The C-D (primary cause to common treatment or service) relations were then trained using the worksheets, and immediately following their mastery, the D-A vocal intraverbal posttest was administered. Training was thus conceptualized as a variation of the simple-to-complex protocol (Fields et al., 2009). The session corresponding to each unit’s protocol was conducted prior to coverage of the particular unit in the course. Final written intraverbal tests were presented 114 days after the completion of Unit 1, 87 days after the completion of Unit 2, and 52 days after the completion of Unit 3.

**Dependent measure and interobserver agreement.** The primary dependent measure was the percentage of correct responses during vocal
intraverbal pre- and posttest probes for each of the three unit protocols. Vocal intraverbals were inferred to have emerged on posttests if a participant performed correctly on 11 of the 12 test trials (92% correct) for each particular vocal intraverbal relation, including the B-A (naming a disability when given its definition), C-A (naming a disability when given its primary cause), and D-A relations. A secondary measure was the number of trial blocks required to attain mastery criterion during worksheet training. An additional secondary measure was scores on final written intraverbal tests.

Interobserver agreement was recorded by an independent observer for 33% of all participants’ pre- and posttest probe trials and 100% of all participants’ training worksheets for all three units. The second observer viewed pre- and posttest probe sessions via video. For test probes, agreement as to whether a correct response occurred was calculated by dividing agreements by agreement plus disagreements and converting the ratio to a percentage. Mean interobserver agreement was 94% for test probes. The same formula was used to evaluate whether or not participants responded correctly for each item on each training and final worksheet, with resulting interobserver agreement equaling 100% for both training and follow-up worksheets.

**General procedure.** This study included three instructional protocols (Table 2) that corresponded to the three units of the class: Unit 1 protocol, developmental and physical disabilities; Unit 2 protocol, health-related disorders; and Unit 3 protocol, psychiatric and age-related disabilities. The Unit 2 protocol was implemented 27 days after completion of the Unit 1

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Class</th>
<th>A Name</th>
<th>B Definition</th>
<th>C Primary cause</th>
<th>D Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit 1</td>
<td>1</td>
<td>Cerebral palsy</td>
<td>Incomplete control of body’s muscles</td>
<td>Oxygen deprivation in the brain</td>
<td>Positioning aids, environmental control</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Spina bifida</td>
<td>Failure of the spinal cord to close completely</td>
<td>Folic acid insufficiencies during pregnancy</td>
<td>Supplementation, prenatal surgery</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Multiple sclerosis</td>
<td>Condition in which the body attacks the central nervous system</td>
<td>Causes unknown</td>
<td>Medication, air conditioning</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Muscular dystrophy</td>
<td>Progressive weakness and deterioration of muscles</td>
<td>Insufficient supply of a required protein</td>
<td>Experimental research</td>
</tr>
<tr>
<td>Unit 2</td>
<td>1</td>
<td>Cystic fibrosis</td>
<td>Abnormal secretions leading to mucus accumulation in lungs</td>
<td>A defective recessive gene</td>
<td>Anti-inflammatory treatments, drugs, diet</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Sickle-cell disease</td>
<td>Disorder of the oxygen-transporting red blood cells</td>
<td>Genetic inheritance from the parents</td>
<td>Vaccinations, penicillin</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Hypertension</td>
<td>Condition in which blood pressure is constantly elevated</td>
<td>Narrow arteries/fat build-up</td>
<td>Drugs, cardiac rehabilitation, surgery</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Diabetes</td>
<td>Condition in which the body is unable to make or use insulin</td>
<td>Destruction of beta cells in the pancreas</td>
<td>Physical exercise, healthy eating, injections</td>
</tr>
<tr>
<td>Unit 3</td>
<td>1</td>
<td>Apraxia</td>
<td>Loss of ability to carry out learned movements</td>
<td>Neurological damage or diseases</td>
<td>Repetitive motor drills, behavioral therapy</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Schizophrenia</td>
<td>Chronic, severe, and disabling mental disorder</td>
<td>Genetic predisposition or environmental stressors</td>
<td>Antipsychotic medication, psycho-social treatment</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Alzheimer’s disease</td>
<td>A progressive neurological disease of the brain</td>
<td>Build-up of proteins in the brain</td>
<td>Emotional and medical support</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Dementia</td>
<td>Severe loss of cognitive ability</td>
<td>Diseases in the nervous system or blood vessels</td>
<td>Healthy diet, memory aids</td>
</tr>
</tbody>
</table>
protocol, and the Unit 3 protocol was implemented 35 days after completion of the Unit 2 protocol. Participants were given the following instructions before their completion of a particular unit:

You will be asked to complete a series of pretests, training sessions, and posttests on the definitions, causes, and treatments for various disabilities. Test conditions will be conducted in a flash-card-style fashion, and no feedback will be delivered. Training conditions will be conducted using paper-and-pencil questionnaires, and feedback will be provided after completion of each. There will be a total of three training sessions, each followed by a test condition. In order to move on to the test conditions, you must first answer 11 of 12 questions correctly on the worksheets. If you answer fewer than 11 correct, you will be asked to retake the previous worksheet.

Vocal intraverbal pretests. Each unit protocol began with a pretest. Pretests assessed whether participants could name the correct disability when presented with its definition (B-A relations), cause (C-A relations), and treatment (D-A relations) on a flash card. The B-A intraverbal relations were evaluated first, followed by the C-A and then D-A relations. The vocal intraverbal B-A relations consisted of 12 consecutive test trials, as did the vocal intraverbal C-A and D-A relations. Within each 12-trial block of trials, each flashcard was presented three times each. Participants were given the following instructions before the vocal intraverbal pretest: “At this time, I am going to show you different definitions, causes, and treatments for various disabilities; do your best to try and tell me the name of the disability for each.”

The presentation of a flash card on which either the B (definitions for disabilities), C (common causes), or D (common treatments or services) stimuli were printed marked the onset of each trial. Flash cards were presented by the experimenter one at a time and were positioned approximately 72 cm from the participant. Participants were given 5 s to provide the name of the correct disability (the A stimuli). The experimenter scored the trial as incorrect if a response did not occur, or if a participant responded with the name of an incorrect disability or said “I don’t know.” Correct and incorrect responses did not result in feedback from the experimenter. The experimenter recorded all responses on a computerized data sheet, which faced away from the participant.

The order of stimulus presentations was determined randomly within each 12-trial block. Prior to the onset of each 12-trial block, the experimenter shuffled the flash cards text side down. Participants advanced to selection-based intraverbal training for a particular unit if they scored below 50% on that unit’s vocal intraverbal pretest.

Selection-based intraverbal (worksheet) training. The following selection-based intraverbal relations were trained using a paper-and-pencil protocol: names of disabilities to their definitions (A-B), names of disabilities to their primary causes (A-C), and primary causes to common treatment or services (C-D) for the particular disabilities in each unit protocol. Worksheets for the A-B relations (names of disabilities to their definitions) were completed to mastery first, after which a posttest for the vocal intraverbal B-A relations (naming a disability when given its definition) was presented, followed by the completion of the A-C relations (names of disabilities to their primary cause) worksheets to mastery, after which posttests for the vocal intraverbal C-A relations (naming a disability when given its primary cause) were presented, followed by the completion of the C-D relations (primary cause to common treatment or service) worksheets to mastery, after which posttests for the vocal intraverbal D-A relations (naming a disability when given a common treatment or service) were presented.

For each unit, participants were taught four selection-based A-B (A1-B1, A2-B2, A3-B3, and A4-B4; disability names to definitions), A-C (A1-C1, A2-C2, A3-C3, and A4-C4; names to primary causes), and C-D (C1-D1, C2-D2, C3-D3, and C4-D4; primary cause to effective
treatments or services) relations each. During training of the selection-based A-B and A-C relations, the A stimuli or disability names were conceptualized as the sample stimuli (i.e., cerebral palsy), and the B or C stimuli or the definitions or causes were conceptualized as the comparison stimuli (i.e., incomplete control of the body’s muscles). During selection-based C-D training, the C stimuli or causes were conceptualized as sample stimuli, and the D stimuli or common treatments or services were conceptualized as comparison stimuli. Each item on the worksheet was conceptualized as a trial. A selection-based response of circling the correct response option was required for a trial to be scored as correct. A mastery criterion of 11 of 12 questions (92%) was in effect for each worksheet. If a participant failed to meet this criterion, he or she was given written feedback, allowed time to look over the answers, and then was asked to repeat an identical worksheet until mastery criterion was met.

Participants were given the following instructions before training:

Now, I am going to ask you to complete a multiple choice worksheet. Once you have completed the worksheet, I will score your responses and give you some time to look over any items that you may have missed. You must answer at least 11 out of 12 correct before moving on to the next phase.

Participants were then given as much time as necessary to complete the worksheet. The experimenter remained present in case participants had questions, although none did. When participants were finished, the experimenter immediately scored the worksheets using an answer key. Incorrect responses were marked with an X, and the correct answer for that item was circled in red pen. Participants were given as much time as they needed to look over their graded worksheets, after which they were given an identical worksheet to complete if they scored below mastery on the first one.

Vocal intraverbal posttests. Vocal intraverbal posttests were identical to vocal intraverbal pretests, except that posttest trials for all of the emergent vocal intraverbal relations were not conducted at one time. Rather, posttests for the vocal intraverbal B-A relations (naming a disability when given its definition) were conducted following mastery of the selection-based A-B training worksheets, posttests for the vocal intraverbal C-A relations (naming a disability when given its primary cause) were conducted following mastery of the selection-based A-C training worksheets, and posttests for the vocal intraverbal D-A relations (naming a disability when given a common treatment or service) were conducted following mastery of the selection-based A-D training worksheets.

Written intraverbal final tests. On the last day of the course, all students were asked to complete two final tests for extra points toward their class grade. Final tests were conducted to evaluate the transfer of the intraverbal relations established during selection-based training and shown to have emerged during vocal intraverbal posttests across response repertoires. Specifically, posttests evaluated the emergence of written intraverbal relations. Both Final Tests 1 and 2 were in a paper-and-pencil format and were presented to students consecutively. Final Test 1 evaluated the maintenance of the emergent B-A (naming a disability when given its definition), C-A (naming a disability when given its primary cause), and D-A (naming a disability when given a common treatment or service) relations. The format for each question on Final Test 1 (Appendix D) was as follows: Given the combination of B (definition), C (cause), and D (treatment) shown simultaneously, participants were asked to provide the name (A) of the disability. Final Test 1 included 12 items, each addressing a different stimulus class among the three unit protocols. Scores were obtained for each unit separately. There were four items for Unit 1 (Questions 1 to 4), four for Unit 2 (Questions 5 to 8), and four for Unit 3 (Questions 9 to 12). Thus, for each unit protocol, there were a total of four possible points that could have been achieved on Final
Test 1. The number of items correct was divided by the number of points possible and converted to a percentage.

Final Test 2 evaluated the maintenance of the directly trained relations for each unit protocol, including names of disabilities to their definitions (A-B), names of disabilities to their primary causes (A-C), and primary causes to common treatment or services (C-D) for the particular disabilities in each respective unit protocol. As shown in Appendix E, participants were provided with the name of a disability (A) and were asked to provide the definition (B), primary cause (C), and common treatment or service (D) for the disability. Final Test 2 included 12 items, each also addressing a different stimulus class among the three unit protocols. Each item on Final Test 2 was worth three points (1 point for definition, 1 point for cause, and 1 point for treatment). Scores were obtained for each of the three units separately. Thus, there were 12 points possible for Unit 1, 12 for Unit 2, and 12 for Unit 3. The number of points achieved was divided by the number of points possible for each unit and converted to a percentage.

Results

Vocal intraverbal pretests. Figure 1 shows that all participants (13 in Unit 1, 13 in Unit 2, and 13 in Unit 3) who completed the Unit 1, 2, or 3 protocols scored in the range of 0% to 42% correct on pretest probes for all vocal intraverbal relations.

Selection-based intraverbal training. Table 1 indicates the number of training trial blocks required for each participant to demonstrate mastery criterion during training of names of disabilities to their definitions (A-B), names of disabilities to their primary causes (A-C), and primary causes to common treatments or services (C-D) relations for Units 1, 2, and 3. Most participants required only one or two trial blocks to master all relations for all three unit protocols.

Vocal intraverbal posttests. The B-A (naming a disability when given its definition), C-A (naming a disability when given its primary cause), and D-A (naming a disability when given a common treatment or service) vocal intraverbal relations were inferred to have emerged if a participant performed with 92% accuracy on test trials for the particular relation. Figure 1 shows the posttest scores for all participants on the B-A, C-A, and D-A vocal intraverbal relations in all three unit protocols. Of the 13 participants who completed Unit 1, nine (Participants 1, 3, 9, 12, 15, 16, 17, 19, and 20) demonstrated the emergence of the definition-to-name (B-A) relations, 12 (Participants 1, 3, 5, 8, 9, 12, 15, 16, 17, 18, 19, and 20) demonstrated the emergence of the cause-to-name (C-A) relations, and eight demonstrated the emergence of the treatment-to-name (D-A) relations. Of the 13 participants who completed Unit 2, 10 (Participants 3, 5, 9, 13, 14, 15, 18, 19, 24, and 25) demonstrated the emergence of the definition-to-name (B-A) relations, 11 (Participants 3, 5, 9, 12, 13, 14, 15, 18, 19, 23, and 25) demonstrated emergence of the cause-to-name (C-A) relations, and nine demonstrated the emergence of the treatment-to-name (D-A) relations. Of the 13 participants who completed Unit 3, nine (Participants 3, 4, 5, 9, 11, 14, 15, 16, and 23) demonstrated the emergence of the definition-to-name (B-A) relations, nine (Participants 1, 4, 5, 9, 13, 14, 15, 16, and 19) demonstrated emergence of the cause-to-name (C-A) relations, and 10 demonstrated the emergence of the treatment-to-name (D-A) relations. Means and standard deviations for accuracy on test trials assessing the emergence of the B-A, C-A, and D-A relations for each of the three units were as follows: Unit 1: B-A ($M = 89\%$, $SD = 13.3$), C-A ($M = 98\%$, $SD = 5$), and D-A ($M = 86\%$, $SD = 18$); Unit 2: B-A ($M = 94\%$, $SD = 14.5$), C-A ($M = 94\%$, $SD = 14.2$), and D-A ($M = 93\%$, $SD = 10.7$); Unit 3: B-A ($M = 86\%$, $SD = 22.7$), C-A ($M = 91\%$, $SD = 11$), and D-A ($M = 94\%$, $SD = 14.5$). Means and standard deviations for overall accuracy for
the three protocols were as follows: Unit 1 ($M = 91\%$, $SD = 13.9$), Unit 2 ($M = 94\%$, $SD = 12.9$), and Unit 3 ($M = 90\%$, $SD = 16.7$).

**Written intraverbal final tests.** Scores for Final Tests 1 and 2 are shown for each of the three unit protocols in Table 3. Only one participant scored 100% accuracy on Final Test 1 for the Unit 1 protocol (Participant 3), and only one participant scored 100% accuracy for the Unit 2 protocol (Participant 3). For Final Test 2, most participants scored no higher than 50% correct. The highest score of 83% correct on Final Test 2 was displayed by Participant 19 for the Unit 2 protocol. Means and standard

---

**Figure 1.** Percentage of correct responses before and after test probes for Units 1, 2, and 3 in Experiment 1. The dotted line represents accuracy level of 92%.

---

**Relations**
deviations for overall accuracy for the three protocols were as follows: Unit 1 protocol: Final Test 1 \( (M = 60\%, SD = 24.2) \) and Final Test 2 \( (M = 34\%, SD = 12) \), Unit 2 protocol: Final Test 1 \( (M = 63\%, SD = 23.1) \) and Final Test 2 \( (M = 54\%, SD = 19.4) \), Unit 3 protocol: Final Test 1 \( (M = 31\%, SD = 29.1) \) and Final Test 2 \( (M = 23\%, SD = 15.3) \).

**Summary of Experiment 1**

These results indicate that all participants acquired the directly trained selection-based intraverbal relations for all three unit protocols using the paper-and-pencil format with ease. Most participants required only one or two training sessions with the worksheets to attain mastery. In addition, test probes revealed that nearly all the participants for each protocol demonstrated the emergence of untrained topography-based relations, specifically, vocal intraverbal. These results are encouraging regarding the role that stimulus equivalence technology might play in higher education learning. Not encouraging, however, were participants’ performances on the written intraverbal final tests. A number of studies have suggested that equivalence relations may be remarkably stable over time, enduring for as long as 2 to 5 months in participants with intellectual disabilities (Saunders, Wachter, & Spradlin, 1988). Whether participants’ poor performances on the written intraverbal final tests reflect a failure of derived responding to generalize across novel test formats or a failure of the emergent relations to endure over time cannot be determined from the present results. For this reason, Experiment 2 was conducted with four participants. These four participants completed the two written intraverbal final tests immediately after completion of their final test probes.

**EXPERIMENT 2**

**Method**

**Participants.** Four graduate students at Southern Illinois University, who had little formal education in types of disabilities, participated. All were compensated with a $20 gift card to a local electronics store for their participation.

**Interobserver agreement.** Interobserver agreement was calculated in the same manner described for Experiment 1 for 0% of pre- and posttest probes, 100% of training worksheets, and 100% of written intraverbal final test scores. Agreement was 100% for the training worksheets and 100% for the final tests.

**General procedure.** The same procedure used in Experiment 1 was used for all participants, with the exception that the written intraverbal final tests were completed immediately after completion of Unit 3 protocol test probes. Due to high pretest scores on Unit 2 probes, Participants 28 and 30 completed only the Unit 1 and 3 protocols. Participants 27 and 29 completed all three protocols. As was the case in Experiment 1, participants completed one unit protocol in its entirety in one session.

**Results**

**Vocal intraverbal pretests.** Figure 2 shows that all participants who completed each of the unit protocols scored in the range of 0% to 42% correct on pretest probes for all of the vocal intraverbal relations.
Selection-based intraverbal training. Table 4 shows the number of training trial blocks required for each participant to demonstrate mastery criterion during training of names of disabilities to their definitions (A-B), names of disabilities to their primary causes (A-C), and primary causes to common treatments or services (C-D) relations. Most participants required only one or two trial blocks to attain mastery criterion in all relations for the unit protocols in which they participated.

Vocal intraverbal posttests. The B-A, C-A, and D-A vocal intraverbal relations were inferred to have emerged if a participant performed with 92% accuracy. Figure 2 shows the posttest scores for all participants on the B-A (naming a disability when given its definition), C-A (naming a disability when given its primary cause), and D-A (naming a disability when given a common treatment or service) vocal intraverbal relations in all three unit protocols. Of the four participants who completed the Unit 1 protocol, four demonstrated the emergence of the B-A relations, three demonstrated the emergence of the C-A relations, and two demonstrated the emergence of the D-A relations. Of the two participants who completed the Unit 2 protocol, both demonstrated the emergence of the B-A and C-A relations, and one participant demonstrated the emergence of the D-A relations. Of the four participants who completed the Unit 3 protocol, four demonstrated the emergence of the B-A relations, two demonstrated emergence of the C-A relations, and four demonstrated the emergence of the D-A relations.

Written intraverbal final tests. Scores for Final Tests 1 and 2 are shown for each of the three

![Diagram of percentage correct responses before and after test probes for Units 1, 2, and 3 in Experiment 2. The dotted line represents accuracy level of 92%.

Figure 2. Percentage of correct responses before and after test probes for Units 1, 2, and 3 in Experiment 2. The dotted line represents accuracy level of 92%.

Selection-based intraverbal training. Table 4 shows the number of training trial blocks required for each participant to demonstrate mastery criterion during training of names of disabilities to their definitions (A-B), names of disabilities to their primary causes (A-C), and primary causes to common treatments or services (C-D) relations. Most participants required only one or two trial blocks to attain mastery criterion in all relations for the unit protocols in which they participated.

Vocal intraverbal posttests. The B-A, C-A, and D-A vocal intraverbal relations were inferred to have emerged if a participant performed with 92% accuracy. Figure 2 shows the posttest scores for all participants on the B-A (naming a disability when given its definition), C-A (naming a disability when given its primary cause), and D-A (naming a disability when given a common treatment or service) vocal intraverbal relations in all three unit protocols. Of the four participants who completed the Unit 1 protocol, four demonstrated the emergence of the B-A relations, three demonstrated the emergence of the C-A relations, and two demonstrated the emergence of the D-A relations. Of the two participants who completed the Unit 2 protocol, both demonstrated the emergence of the B-A and C-A relations, and one participant demonstrated the emergence of the D-A relations. Of the four participants who completed the Unit 3 protocol, four demonstrated the emergence of the B-A relations, two demonstrated emergence of the C-A relations, and four demonstrated the emergence of the D-A relations.

Written intraverbal final tests. Scores for Final Tests 1 and 2 are shown for each of the three
unit protocols in Table 5. For Unit 1, three of the four participants scored 100% accuracy on Final Test 1, and all participants scored at or above 67% accuracy on Final Test 2. For Unit 2, both participants scored 100% accuracy on Final Test 1; Participant 29 scored 92% accuracy on Final Test 2, and Participant 27 scored 75% accuracy on Final Test 2. For Unit 3, three of the four participants scored 100% accuracy on Final Test 1, and all participants scored at or above 75% accuracy on Final Test 2.

Summary of Experiment 2

These results replicate the findings of Experiment 1 in that (a) most participants acquired the selection-based intraverbal relations for all three unit protocols within one or two training sessions using the paper-and-pencil format, and (b) almost all participants demonstrated the emergence of untrained topography-based vocal intraverbal relations following mastery of the worksheets. Further, participants in Experiment 2 derived untrained written intraverbal relations for all three unit protocols immediately after completion of the Unit 3 vocal intraverbal test probes. In Experiment 2, scores on final written intraverbal tests were much higher than those of Experiment 1, indicating that poor final test performances in Experiment 1 were a retention, not generalization, failure. The results also indicate that training selection-based intraverbal relations may facilitate the emergence of both vocal and written topography-based intraverbal relations. It should be noted that the participants in Experiment 2 were graduate students, whereas the participants in Experiment 1 were undergraduate students. It is possible that the participants in Experiment 2 had richer histories of relating instructional stimuli and were thus better prepared to be successful on posttest probes. However, given similarities in the pre- and posttest performances of participants in both experiments, time of testing is a plausible explanation for the differences in final test performance.

DISCUSSION

The present results contribute to the growing body of literature that underscores the utility of the stimulus equivalence paradigm in higher education. The protocol seems to be useful in illustrating course concepts, emphasizing relations, and expanding repertoires to a variety of novel topographies following just a small amount of teaching. Although the paradigm has been valuable in teaching a variety of basic language and academic skills in learners with rudimentary repertoires, the present results join those of Ninness et al. (2005, 2006, 2009), Fienup et al. (2010), and Fields et al. (2009) in accentuating the equivalence protocol as a teaching tool with verbally sophisticated learn-
ers. The protocols used in this and other experiments seem to reflect the best practices in teaching articulated by Skinner (2003), who argued for the provision of numerous opportunities for student feedback and teaching until mastery.

Not only do the results support the role of stimulus equivalence protocols in higher education, but the paper-and-pencil training format used in this study is a particularly noteworthy innovation. Superficially, paper-and-pencil protocols are no different from the myriad of homework assignments, worksheets, and evaluation tools already present in many classrooms (Eikeseth et al., 1997). Moreover, the protocol was efficient from both learning and instructional standpoints: Most participants mastered the selection-based intraverbal relations in few training blocks, and little time or effort was required on the part of the experimenter, who only needed to read the instructions to participants and grade the worksheets. For this reason, the protocol might easily be implemented in small-group learning arrangements that are common in many classroom settings. Finally, given the widespread availability of materials, the effects of the paper-and-pencil protocol in one classroom could be easily replicated in another classroom. Because of these benefits, future research should continue to explore the ease with which this protocol can be implemented in educational settings of any level.

All of the emergent skills demonstrated by participants were topography based rather than selection based. Untrained skills revealed on vocal intraverbal posttests and written intraverbal final tests all emerged from the simple selection-based intraverbal training conducted using the worksheets. Although successful performance on multiple-choice exams, which requires selection-based responding (Michael, 1985), is emphasized in many college courses, this repertoire is arguably not a good representation of the desirable verbal skills that should be acquired in college. Higher education emphasizes the development of skills in both written and oral expression. The present results suggest that vocal and written intraverbal behavior can emerge following the establishment of selection-based intraverbal relations (see also Chase et al., 1985; Perez-Gonzalez et al., 2008). In addition to reflecting skills in written and oral expression, intraverbal responding may also be said to reflect comprehension of the material (see M. L. Sundberg, 2008). Thus, the most noteworthy findings may well be the performance of the four participants in Experiment 2 on the written intraverbal final tests. Critchfield and Fienup (2008, p. 363) describe a learning experience as generative if it “spawns novel abilities.” Both final tests required considerably more generative behavior on the part of participants than was required during posttests or training; moreover, responding on final tests required a different response topography than that which was required at any other time during the experiment. These results suggest that with the stimulus equivalence paradigm, relatively inexperienced but verbally skilled students are capable of going well beyond basic multiple-choice items to master intricate and comprehensive relations with impressive competence. It should be noted that all participants in this study had sophisticated repertoires of textual responding or covert or overt responding under the control of text (Skinner, 1957), so a proficient textual repertoire is an apparent prerequisite for participants to demonstrate desired outcomes with similar instructional protocols.

Several limitations merit mention. First, performances on the written intraverbal final tests for Experiment 1 were not encouraging. Prior research has suggested that derived stimulus relations may be considerably stable over time (Rehfeldt & Hayes, 2000; Saunders, Wachten, & Spradlin, 1988), even in participants with limited cognitive abilities. That participants in Experiment 1 performed so poorly on the written intraverbal tests is
somewhat surprising. However, we did not examine the long-term stability of the directly trained selection-based relations or the emergent vocal intraverbal relations demonstrated on posttests, so it is uncertain whether these performances would have been maintained on the final tests as well. In addition, participants who did not show the emergence of the vocal intraverbal relations during posttests did complete written intraverbal final tests, so there is no reason to expect that those participants would have showed retention of the relations. Because most instructors hope that students will retain the information learned in a course well beyond the completion of the course, future research should examine the long-term stability of relations established via the stimulus equivalence protocol in classroom settings. A second limitation is the fact that participants’ responding during training could have been controlled by the first few words of each response option only. In other words, because the configuration of words and letters was different for the different response options, it is possible that a participant could have responded correctly by reading only the first part of each response option. Similarly, the order of questions on the training worksheets did not differ from one training worksheet to the next, so responding could have also come under the control of the question numbers. It is doubtful that participants would have performed with such high accuracy on vocal intraverbal posttests and written intraverbal final tests (Experiment 2) had this been the case, but future studies should strive for structural similarity when employing phrases as stimuli as well as employ different versions of worksheets targeting the same relations in which the questions are presented in different orders. Third, one might argue that because a single-subject design was not used, the pretest-train/posttest sequence used in this study did not control for plausible threats to internal validity given that participants in Experiment 1 were completing an undergraduate course on disabilities at the time of their participation. However, all of the participants completed the unit protocols prior to any lectures or assigned readings on the particular disabilities in the course. In addition, the fact that pretests, training, and posttests were all completed in one session makes threats to internal validity (e.g., history and maturation) highly implausible. Fourth, we cannot discern from the present results whether the protocols used in this study would be more effective than more conventional educational approaches, including the participation in lecture or completion of study questions. Future research should compare the performance of students who participate in such protocols to those of students who participate in more conventional classroom practices (see Fields et al., 2009).

The construction of the final written intraverbal tests merits mention. On Final Test 1, students were required to supply the name of a disability (A stimulus) when provided with its definition (B stimulus), primary cause (C stimulus), and a common treatment or service (D stimulus). Given the presence of the B, C, and D stimuli on Final Test 1, it is possible that these stimuli controlled participants’ responding on Final Test 2, on which they had to supply the definition (B stimulus), primary cause (C stimulus), and a common treatment or service (D stimulus) when provided with the name of a disability (the A stimulus). Thus, performance on Final Test 2 may have been influenced by the stimuli provided to participants on Final Test 1.

Future research should further examine the efficacy of paper-and-pencil methods for implementing stimulus equivalence methods in higher education by continuing to explore the long-term stability or retention of verbal relations established during experimental sessions. The stability of both emergent and directly trained performances should be examined. In addition, future research should continue to explore the degree to which
instruction in one repertoire facilitates the emergence of skills in another repertoire. Finally, this study focused on the establishment of relations of equivalence or sameness only. Relational frame theory (i.e., Hayes, Barnes-Holmes, & Roche, 2001) posits that any variety of contextually controlled relations may emerge via a history of relating multiple exemplars, including relations of sameness, opposition, and comparison (e.g., greater than and less than; see Berens & Hayes, 2007; Ninness et al., 2009). Students are often asked to “compare and contrast” different aspects of course material, a complex form of relating that likely involves the derivation of comparative relations. Future research should examine the role of more complex stimulus networks in college courses.

REFERENCES


APPENDIX A

Multiple-Choice Training Worksheet for A-B Relations in the Unit 1 Protocol

Each question was presented three times on a worksheet in a randomly determined order. The order of the response options was also randomly determined and differed from one question to the next for a given relation.

1.) Define cerebral palsy
   a. Failure of the spinal cord to close completely
   b. Incomplete control of the body’s muscles
   c. Progressive weakness and deterioration of muscles
   d. Condition in which the body attacks the central nervous system

2.) Define muscular dystrophy
   a. Incomplete control of the body’s muscles
   b. Failure of the spinal cord to close completely
   c. Condition in which the body attacks the central nervous system
   d. Progressive weakness and deterioration of muscles

3.) Define spina bifida
   a. Failure of the spinal cord to close completely
   b. Progressive weakness and deterioration of muscles
   c. Condition in which the body attacks the central nervous system
   d. Incomplete control of the body’s muscles

4.) Define multiple sclerosis
   a. Condition in which the body attacks the central nervous system

APPENDIX B

Multiple-Choice Training Worksheet for A-C Relations in the Unit 2 Protocol

Each question was presented three times on a worksheet in a randomly determined order. The order of the response options was also randomly determined and differed from one question to the next for a given relation.

1.) What causes muscular dystrophy?
   a. Oxygen deprivation in the brain
   b. Insufficient supply of a required protein
   c. Causes unknown
   d. Folic acid insufficiencies during pregnancy

2.) What causes cerebral palsy?
   a. Causes unknown
   b. Oxygen deprivation in the brain
   c. Insufficient supply of a required protein
   d. Folic acid insufficiencies during pregnancy

3.) What causes spina bifida?
   a. Insufficient supply of a required protein
   b. Causes unknown
   c. Oxygen deprivation in the brain
   d. Folic acid insufficiencies during pregnancy

4.) What causes multiple sclerosis?
   a. Folic acid insufficiencies during pregnancy
   b. Causes unknown
   c. Oxygen deprivation in the brain
   d. Insufficient supply of a required protein
Multiple-Choice Training Worksheet for C-D Relations in the Unit 3 Protocol

Each question was presented three times on a worksheet in a randomly determined order. The order of the response options was also randomly determined and differed from one question to the next for a given relation.

1.) How do you treat a disability for which causes are unknown?
   a. Positioning aids, environmental controls
   b. Diet supplementation, prenatal surgery
   c. Medication, air conditioning
   d. Experimental research

2.) How do you treat a disability caused by oxygen deprivation in the brain?
   a. Diet supplementation, prenatal surgery
   b. Positioning aids, environmental controls
   c. Medication, air conditioning
   d. Experimental research

3.) How do you treat a disability caused by folic acid insufficiencies during pregnancy?
   a. Positioning aids, environmental controls
   b. Experimental research
   c. Diet supplementation, prenatal surgery
   d. Medication, air conditioning

4.) How do you treat a disability caused by insufficient supply of a required protein?
   a. Medication, air conditioning
   b. Positioning aids, environmental controls
   c. Diet supplementation, prenatal surgery
   d. Experimental research

5.) Samantha’s spinal cord is not closed completely due to folic acid insufficiencies. To decrease the chances of her children being born with the same disability, she is taking dietary supplements. What disorder does she have?

3.) Thomas has a condition where his body attacks his CNS, and professionals are unsure of what causes this condition. However, his doctors recommend medication and air conditioning for treatment. What disorder does he have?

4.) Samuel has a condition where he experiences progressive weakness and deterioration of his muscles caused by an insufficient supply of a required protein. His mother had considered gene therapy during pregnancy, however decided against it due to it being highly experimental. What disorder does he have?

5.) Nicholas has a condition where his body is unable to produce a certain type of protein, which was caused by a defective recessive gene and is treating his condition through the use of anti-inflammatory treatments, drugs, and a specialized diet. What disorder does he have?

6.) Leah has a disorder of the oxygen-transporting cells, which she genetically inherited from her parents and can often be prevented through prenatal screening and vaccinations. What disorder does she have?

7.) Bianca’s blood pressure is constantly elevated due to the narrowing of her arteries and fat build-up. Her doctor has prescribed her medication and recommended she exercise regularly to treat her condition or else she may need surgery if her condition worsens. What disorder does she have?

8.) Dennis is unable to produce insulin due to destruction of beta cells in his pancreas. He now uses injections to treat his condition and closely monitors his
blood sugar. Dennis encourages his son to eat a healthy diet and exercise routinely to prevent him from acquiring this condition. What disorder does he have?

9.) Sabrina has lost her ability to carry out learned movement due to neurological damage. She now practices repetitive motor drills and receives behavior therapy to treat her condition. What disorder does she have?

10.) Tyler has a chronic, severe, and disabling mental disorder caused by a combination of genetic predisposition and environmental stressors. To treat this condition, he takes antipsychotic medication and receives psychosocial treatment. What disorder does he have?

11.) Carson has a progressive neurological disease of the brain caused by build-up of proteins in the brain and now receives emotional and medical support. What disorder does she have?

12.) Mimi experienced a severe loss of her cognitive ability due to a disease in her nervous system or blood vessels. Her doctor now encourages her to eat a healthy diet and use memory aids. What disorder does she have?

APPENDIX E

Final Test 2
List and describe the (a) definition, (b) primary cause, and (c) treatment for the following disabilities:

1.) Cerebral palsy
2.) Spina bifida
3.) Multiple sclerosis
4.) Muscular dystrophy
5.) Cystic fibrosis
6.) Sickle-cell disease
7.) Hypertension
8.) Diabetes
9.) Apraxia
10.) Schizophrenia
11.) Alzheimer’s disease
12.) Dementia