Behavioral Interventions
Behav. Intervent. **18**: 279–289 (2003)
Published online in Wiley InterScience (www.interscience.wiley.com). **DOI**: 10.1002/bin.144

THE EFFECTS OF THREE STIMULUS-EQUIVALENCE TESTING CONDITIONS ON EMERGENT US GEOGRAPHY RELATIONS OF CHILDREN DIAGNOSED WITH AUTISM

Linda A. LeBlanc*, Caio F. Miguel, Anne R. Cummings, Tina R. Goldsmith, and James E. Carr
Western Michigan University, USA

Two children with autism were trained in US geography using a match to sample procedure. Different testing procedures commonly used in stimulus equivalence research were compared, including some conditions with reinforcement for responding to enhance motivation. Both children were able to master the trained geography relations and emergent stimulus relations were also noted. All three testing procedures produced similar effects, suggesting that incorporation of procedures to enhance motivation to respond does not invalidate testing in a stimulus equivalence preparation. Copyright © 2003 John Wiley & Sons, Ltd.

INTRODUCTION

One way in which stimuli that share no similarities can enter into and subsequently comprise stimulus classes is through equivalence training (Sidman & Tailby, 1982). When the mathematical properties of reflexivity (A = A), symmetry (if A = B, then B = A), and transitivity (if A = C and B = C, then A = C) are reliably shown between stimuli, then they are said to be part of the same equivalence class. Although stimulus equivalence has been primarily demonstrated with typically developing children and adults, there have been several applied demonstrations with individuals who have mental retardation. For example, equivalence training has been used to teach money (see, e.g., Trace, Cuvo, & Criswell, 1977) and reading skills (e.g., Mackay, 1985; Lane & Critchfield, 1998) to individuals with mental retardation. Although studies have been conducted with children with mental retardation, Cautilli, Hancock, Thomas, and Tillman (2002) recently commented that stimulus equivalence technology has not been well integrated into early intervention curricula for children with autism, identifying it as a promising area for future investigation. There is a need for more *research* on the benefits of equivalence training (Sidman, 1994) across

*Correspondence to: Linda A. LeBlanc, Department of Psychology, Western Michigan University, 1903 West Michigan Avenue, Kalamazoo, MI 49008-5439, USA. E-mail: linda.leblanc@wmich.edu

instructional areas and on possible methodological modifications required to demonstrate such relations with individuals with autism.

The train-and-test method is the most common preparation for establishing and demonstrating stimulus equivalence relations. Individuals may be given pretests for key relations (e.g. transitivity), then experience subsequent training of several relations, and finally receive a post-test for the same untrained, emergent relations (Green & Saunders, 1998). The administration of pre-tests may not be necessary for basic preparations when the stimuli are arbitrarily generated and unlikely to be known until training occurs. However, in applied studies naturally occurring stimuli such as numbers, letters, and whole words are more likely to be used, necessitating demonstration of poor pre-test performance prior to training (see, e.g., Lynch & Cuvo, 1995).

Although the modal testing method is to assess relations under an extinction condition (see, e.g., Kennedy, Itkonen, & Lindquist, 1994), several modifications have been reported in the literature (Green & Saunders, 1998). For example, Saunders, Wachter, and Spradlin (1988) interspersed reinforced trials of trained relations with testing trials of untrained relations. Such modifications to equivalence testing procedures are potentially beneficial for individuals with disabilities because of the likelihood of poor performance and/or problem behavior when reinforcement is scarce or nonexistent (i.e. extinction). Task interspersal procedures are designed to counteract such negative outcomes (Neef, Iwata, & Page, 1980). Reinforcement could also be provided for continuing effort regardless of testing performance, though this procedure has not yet been used in a stimulus equivalence study. The use of testing procedures that include some level of reinforcement may be necessary for children with autism but no research has compared the utility of different procedural variations. Without experimental comparison of these different procedures, researchers cannot be certain that reinforcement for responding does not invalidate or falsely inflate or deflate performance on tests of equivalence.

The purpose of the current study was twofold. The primary purpose was to directly compare three testing procedures to assess emergent performances of children diagnosed with autism. Two of these procedures included reinforcement for responding, while the third did not. The second purpose of the study was to extend applied research on stimulus equivalence with individuals who have autism to a common academic topic, US geography.

METHOD

Participants, Setting, and Materials

Two male children diagnosed with autism participated in this study. Parental permission was obtained and other participant protections were taken in accordance

with the University's Human Subjects Review Board. Nick was a 6-year-old male educated partially in a self-contained classroom for children with autism and partially in a regular first-grade classroom. Nick could speak clearly in one or two word phrases, but most speech was prompted or occurred in the form of echolalia. Marty was 13 years old and was educated in a self-contained special education classroom for children with autism. He could speak in short, full sentences, but most speech was prompted. Nick's sessions were conducted in a separate classroom with only the therapist, participant, and data collectors present. Marty's sessions occurred in a separate area of his classroom.

All matching-to-sample trials were presented manually using a 29 cm by 30 cm three-ring binder. Materials for each trial were presented on a transparent sheet protector with a cardboard insert. Each page included a sample stimulus at the top and three comparison stimuli at the bottom for selection. The sample stimulus was covered with a small cardboard flap that the participant uncovered to reveal the stimulus (i.e. an observing response). As each trial was completed, the trial page was flipped by the experimenter and the next page became visible. The stimuli were 7.5 cm by 10.2 cm. Each stimulus was encased in a transparent hard plastic cover and attached to one of the locations on the page by Velcro. Participants were taught to remove their selected comparison stimulus from the array and hand it to the experimenter. The three-choice matching procedure was chosen to prevent conditional discriminations from coming under the control of an incorrect comparison (i.e. reject relation), which typically results in failure on tests of transitivity (de Rose, 1996).

Three groups of US geography stimuli were created. Each group consisted of information about three states from a contiguous geographic region. Sets A, B, and C were comprised of printed state names, maps of corresponding state shapes, and printed state capitals, respectively. Maps of state shapes included each of the three contiguous states with the target state colored red and other states colored light gray on a white background (see Figure 1). The Midwest group consisted of the state names, shapes, and capitals of Minnesota, Illinois, and Wisconsin. The Southeast

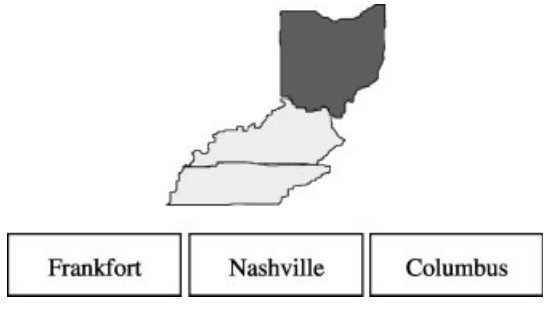

Figure 1. Sample stimulus board for B–C relation.

group consisted of Florida, Georgia, and Alabama for Nick and North Carolina, South Carolina, and Virginia for Marty, who performed well on the three initial states at pretest. The Ohio Valley region consisted of Ohio, Kentucky, and Tennessee for Nick and Kentucky, Tennessee, and West Virginia for Marty, who performed well for Ohio only.

Procedures

Vocal Tests

A vocal pretest was administered for each region to determine whether participants could identify the target states, shapes, and capitals when asked in the form of a standard geography quiz (e.g. 'What is the capital of Ohio?'). The vocal post-test was re-administered after completion of all other phases of the study.

Stimulus Preference Assessment

A multiple-stimulus without replacement preference assessment (DeLeon & Iwata, 1996) was conducted at the beginning of the study with preferred food items (e.g. mini-cookies, M & Ms) generated by the participant's teacher and parent. An array of food items was presented and the participant was asked to select one. When an item was selected it was given to the participant for consumption and the remaining items were randomly rearranged. The participant was then instructed to select another one. The order of selection was translated into a preference ranking and average rankings were calculated based on three presentations of the stimulus array. Before each session, a stimulus preference assessment was repeated with the top three to five stimuli from the initial assessment. The selected item was present during the session and made available as indicated below in description of phases and experimental conditions.

Testing and Training Phases

A summary of all testing and training phases is presented in Table 1. Sessions occurred two to five days per week with no more than two sessions per day, depending on the participant's schedule. Sessions were approximately 20–30 min in duration and consisted of one to seven ten-trial blocks.

Pre-Training (Phases 1 and 2)

These phases were designed to teach the participant the selection and observing responses. During phase 1, the participant was instructed to lift a cardboard flap over

Table 1. Sequence of training and testing phases

Phase	*Task*	*Skill/relation*	*Trials*	*Verbal instruction*
1	Training	Observing response	10[a]	'Lift flap'
2	Training	Selection response	10[a]	'Lift flap' 'Find the same and pull it off'
3	Pre-test	C–A (capital to state name)	12	'Lift flap' '___is the capital for which state?'
4	Pre-test	A–C (state name to capital)	12	'Lift flap' '___is the state for which capital?'
5	Pre-test	B–A (state shape to state name)	12	'Lift flap' 'What is the name of the red state?'
6	Pre-test	C–B (capital to state shape)	12	'Lift flap' '__ is the capital of which red state?'
7	Training	A–B (state name to state shape)	10-trial blocks[b]	'Lift flap' 'What red state is___?'
8	Training	B–C (state shape to capital)	10-trial blocks[b]	'Lift flap' 'What is the capital of the red state?'
9	Mixed training	A–B & B–C	10-trial blocks[b]	Alternating questions
10	Post-test	C–A	12	Same as phase 3
11	Post-test	A–C	12	Same as phase 4
12	Post-test	B–A	12	Same as phase 5
13	Post-test	C–B	12	Same as phase 6

[a]Progression criterion: three consecutive correct and independent trials.
[b]Progression criterion: three consecutive ten-trial blocks with 90% or greater independent and accurate responses.

the sample (top) stimulus. Phase 2 was designed to teach the selection response. The participant was presented with a picture (e.g. turtle) and instructed to find the same among the comparison pictures and pull it off the page. In each phase, correct responses were differentially reinforced and incorrect responses lead to a gestural model and then hand-over-hand prompting. The progression criterion was three consecutive trials that were independent and correct.

Pretests (Phases 3–6)

These phases served as pretests for equivalence relations. The relations were symmetry (B–A and C–B), transitivity (A–C), and a simultaneous test of symmetry and transitivity referred to in the literature as the direct test of equivalence (C–A) (Green & Saunders, 1998). Each test phase consisted of 12 trials (four trials per state). Children had to perform at or below 50% on each phase to be included in the study. Stimulus presentation procedures were identical but consequences varied according to testing condition as described below.

Training A–B, B–C, and Review (Phases 7–9)

Each phase consisted of a series of ten-trial training blocks to teach the A–B and B–C conditional relations. During each phase, a question was presented corresponding to the sample stimulus at the top of the page and three stimuli at the bottom of the page (e.g. 'What is the capital of the red state?' for B–C). The experimenter differentially reinforced accurate stimulus selection by delivering the chosen edible and using descriptive praise (e.g. 'Great, the capital of Minnesota is St. Paul'). When the child selected an incorrect stimulus, the experimenter reissued the instruction and provided a gestural prompt by directly pointing to the correct stimulus and, if necessary, by using a hand-over-hand prompt. Therefore, the participant selected the correct card at the end of each trial independently or with a prompt before moving on to the next trial. Gestural prompts were sufficient in almost every correction procedure. A series of ten trial training blocks of the same relation was conducted until the participant performed at 90% or greater accuracy for three consecutive blocks.

Post-tests

Stimulus presentation procedures were identical to pre-test conditions and were designed to determine whether untrained relations had emerged. Consequences varied with testing condition as described below.

Testing Conditions

We evaluated the effects of three testing conditions on the identification of emergent relations. The conditions were (i) extinction during both pre-test and post-test (EXT–EXT), (ii) extinction during pre-test and interspersal of training trials during post-test (EXT–INT), and (iii) continuous stimulus delivery (regardless of accuracy) during both pre-test and post-test (FR1–FR1). Inter-trial intervals were kept constant at 4–5 s for all conditions. At the session's end during all conditions, the experimenter delivered one preferred edible and praised the participant's effort for that session. Each testing condition was paired with a different state set and delivered in a different serial position for each participant resulting in a partial counterbalancing across order and state set. Nick encountered the following conditions in order: (i) EXT–EXT/Midwest, (ii) FR1–FR1/Southeast, (iii) EXT–INT/Ohio Valley. Marty encountered the following conditions in order: (i) FR1–FR1/Ohio Valley, (ii) EXT–INT/Midwest, (iii) EXT–EXT/Southeast.

EXT–EXT

During both pre-test and post-test, each trial was presented and no differential response was provided by the experimenter. Completion of one trial, regardless of

performance, resulted in the immediate presentation of the next trial. The item selected in the preference assessment was present but was not delivered until the end of the session. Similar procedures have been previously used in the equivalence literature (Eikeseth & Smith, 1992).

FR1–FR1

During both pre-test and post-test, each response was reinforced regardless of accuracy. Upon completion of a trial, the experimenter delivered the preferred edible selected in the preference assessment without any statement about performance. After a brief delay (2–5 s), the experimenter presented the next trial. Participants could simultaneously consume the edible and complete the next trial. All items were immediately consumed. This condition was included to determine whether reinforcement for 'participation' would affect responding.

EXT–INT

During the pretest, each trial was presented and no differential response was provided by the experimenter. Completion of one trial, regardless of performance, resulted in the immediate presentation of the next trial. During the post-test, test trials met with no differential response by the experimenter. However, after every third test trial, a training trial was conducted for one of the previously learned relations and each correct response was differentially reinforced with a preferred edible. No training trials were presented in any other testing conditions. Similar testing conditions have been used in the equivalence literature (Saunders et al., 1988; Spradlin, Cotter, & Baxley, 1973).

Interobserver Agreement

Observers scored each training and testing trial as correct or incorrect. A second independent observer collected data for 65% of training trials and 100% of testing trials for Nick and 72% of training trials and 96% of testing trials for Marty. Interobserver agreement was calculated by dividing the number of agreements by the total number of agreements plus disagreements and multiplying by 100. Agreement for Nick was 100% for training and 100% for testing. Agreement for Marty was 100% for training and 99.6% for testing.

RESULTS AND DISCUSSION

Figure 2 depicts each child's performance. On average, Nick mastered the trained relations (phases 7–9) in approximately equal numbers of trials (180, 210, 160) with

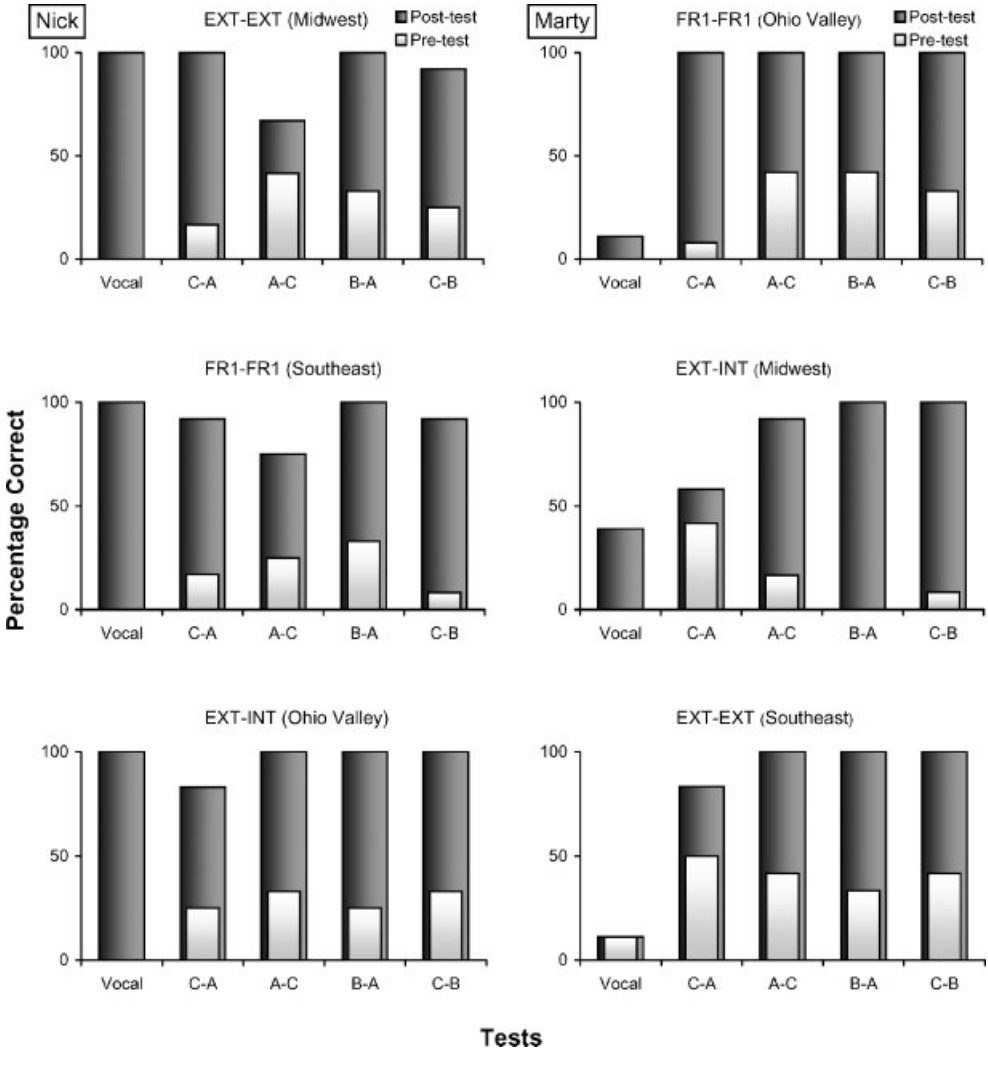

Figure 2. Pre-test and post-test scores for vocal test and C–A, A–C, B–A, and C–B relations presented in the order of administration with the first condition in the top panel and the third condition in the bottom panel. Data for Nick are presented in the left-hand column and data for Marty are presented in the right-hand column.

the highest number of trials in the FR1–FR1 state set. Marty also mastered the trained relations in approximately equal numbers of trials (120, 100, 140) with the highest number of trials in the EXT–EXT state set. Nick's test performances were similar across all conditions and untrained relations reliably emerged. Pretest performance was 8–41% correct for all pre-tests (i.e., chance responding) and increased to an

average of 90% correct for post-tests in the EXT–EXT and FR1–FR1 conditions and 96% correct in the EXT–INT condition. Nick was unable to answer any questions on the vocal pre-test ($M = 0\%$) and was able to answer all questions accurately on the vocal post-test, though no vocal responses were required during any testing or training conditions.

Marty's pre-test performance was 0–50% for all pre-tests (i.e. chance responding). All untrained relations reliably emerged during post-test. Marty's test performance was slightly better under the FR1–FR1 condition ($M = 100\%$), which was his first experimental condition, followed by the EXT–EXT condition ($M = 96\%$) and the EXT–INT condition ($M = 87\%$). The weakest emergence occurred under the EXT–INT condition for the C–A relation (58%). Marty performed poorly on the vocal pre-test (3.7%) and only slightly better on the vocal post-test (20%).

One purpose of the study was direct comparison of the three testing procedures. All conditions required approximately the same amount of time to complete. No substantial differences in performance were observed across testing conditions for these two children, but additional studies will need to be conducted to determine the generality of these findings with children with autism. These results may suggest that our concerns about procedural variations were unwarranted because these reinforcement-based procedural differences did not affect responding. Alternatively, the results may indicate that we neglected to include dependent measures that might capture meaningful differences between conditions such as response latency, attention, or problem behavior. Anecdotal observations suggest that at least one of the participants was less attentive during the EXT post-testing conditions, but no direct measures of attention were made. Future studies should assess other aspects of performance that might also be affected such as attention and problem behavior to determine whether certain testing conditions produce optimal results. Additionally, children with more extreme problem behavior might show a differential performance in certain conditions. Both of these children were reported to display some level of problem behavior prior to entry into the study, although this was not a specific inclusionary criterion. Future studies might examine differences with participants specifically selected for elevated levels of problem behavior.

The second purpose of the study was to extend the equivalence literature to US geography. These children mastered the trained relations among the state maps, the state names, and the corresponding capitals and demonstrated emergent stimulus relations. Additionally, one student was able to respond to vocal tests similar to a standard classroom geography quiz after completion of the protocol.

It is known that the emergence of equivalence classes may be facilitated when stimuli are named (see, e.g., Eikeseth & Smith, 1992). In the current procedure, the printed stimuli were labeled aloud by the experimenter during testing (e.g., 'What red state is Ohio?') and as part of the praise statements during training (e.g. 'Columbus is

the capital of Ohio'), which may have enhanced development of equivalence class formation. It is also possible that the children named covertly. If participant speaker and listener behaviors were interdependent, the training could have facilitated the acquisition of tacts, though this was unlikely given the limited verbal skills of both participants. Regardless of the mechanism, the procedure seemed successful in generating the intended stimulus relations and in producing a socially valid form of responding for the children (e.g. accurate responding on oral geography quizzes).

There are at least three potential limitations of the current study. First, only two participants were included, resulting in a partial counterbalance rather than full; however, the similarity of results across participants suggests minimal or no sequence effect. Second, our protocol standardized the order of relation testing with the most difficult relation (equivalence, C–A) always presented first. This procedure was used because the extinction post-test condition might result in progressively lower motivation across consecutive tests while other conditions should not. The equivalence test was presented first to ensure that the most difficult tests did not occur as motivation waned the most. Also, ordering tests from complex to simple minimizes the likelihood of the development of stimulus classes as a function of testing rather than a product of conditional discrimination training (Green & Saunders, 1998). However, the results might have differed if easier tests had occurred prior to the more difficult tests. Additionally, the brevity of the test conditions (12 trials per condition) may have minimized the potential for extinction of responding in the EXT condition. If longer test conditions were run, this condition might compare less favorably with the others. Finally, since no procedural integrity data were collected it is possible that inadvertent cuing occurred; however, all experimenters were graduate students experienced in conducting tabletop procedures who were instructed to avoid such cues.

In summary, these preliminary results suggest that procedural variations that include some form of reinforcement for responding are appropriate for use with children with special needs and do not invalidate testing procedures, particularly when testing conditions are brief. Additionally, the results demonstrate the utility of matching to sample procedures for teaching US geography to children with autism.

ACKNOWLEDGEMENTS

We thank Stacey Cherup and Jennifer Joseph for their assistance with data collection and the Kalamazoo Regional Special Education Agency for their on-site support.

REFERENCES

Cautilli, J. D., Hancock, M. A., Thomas, C. A., & Tillman, C. (2002). Behavior therapy and autism: Issues in diagnosis and treatment. *Behavior Analyst Today, 3*, 230–242.

DeLeon, I. G., & Iwata, B. A. (1996). Evaluation of a multiple-stimulus presentation format for assessing reinforcer preferences. *Journal of Applied Behavior Analyis, 29*, 519–533.

de Rose, J. C. (1996). Controlling factors in conditional discriminations and tests of equivalence. In T. R. Zentall, & P. M. Smeets (Eds.), *Stimulus class formation in humans and animals* (pp. 253–277). Amsterdam: Elsevier.

Eikeseth, S., & Smith, T. (1992). The development of functional and equivalence classes in high-functioning autistic children: The role of naming. *Journal of the Experimental Analysis of Behavior, 58*, 123–133.

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

Kennedy, C. H., Itkonen, T., & Lindquist, K. (1994). Nodality effects during equivalence class formation: An extension of sight-word reading and concept development. *Journal of Applied Behavior Analysis, 27*, 673–683.

Lane, S. D., & Critchfield, T. S. (1998). Classification of vowels and consonants by individuals with moderate mental retardation: Development of arbitrary relations via match-to-sample training with compound stimuli. *Journal of Applied Behavior Analysis, 31*, 21–41.

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

Mackay, H. A. (1985). Stimulus equivalence in rudimentary reading and spelling. *Analysis and Intervention in Developmental Disabilities, 5*, 373–387.

Neef, N. A., Iwata, B. A., & Page, T. J. (1980). The effects of interspersal training versus high-density reinforcement on spelling acquisition and retention. *Journal of Applied Behavior Analysis, 13*, 153–158.

Saunders, R. R., Wachter, J. A., & Spradlin, J. E. (1988). Establishing auditory stimulus control over an eight-member equivalence class via conditional discrimination procedures. *Journal of the Experimental Analysis of Behavior, 49*, 95–115.

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

Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching-to-sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior, 37*, 5–22.

Spradlin, J. E., Cotter, V. W., & Baxley, N. (1973). Establishing a conditional discrimination without direct training: A study of transfer with retarded adolescents. *American Journal of Mental Deficiency, 77*, 556–566.

Trace, M. W., Cuvo, A. J., & Criswell, J. L. (1977). Teaching coin equivalence to the mentally retarded. *Journal of Applied Behavior Analysis, 10*, 85–92.

