The Psychological Record, 2005, 55, 539-559

EQUIVALENCE CLASS ESTABLISHMENT WITH TWO-, THREE-, AND FOUR-CHOICE MATCHING TO SAMPLE BY SENIOR CITIZENS

RICHARD R. SAUNDERS, LISA CHANEY, and JANET G. MARQUIS
Schiefelbusch Institute for Life Span Studies
University of Kansas

In Experiment 1, 12 senior citizens from the community were trained with 18 sets of conditional discriminations. Training included 2-, 3-, and 4-choice matching-to-sample (MTS) configurations in linear series (LS), many-to-one (MTO), and one-to-many (OTM) training structures. Training structure order was counterbalanced across participants. The design permitted tests for class establishment ranging from 2 classes of 3 stimuli each to 4 classes of 4 stimuli each in the LS, MTO, and OTM structures. The experiment tested the hypothesis that 3- and 4-choice MTS would increase the probability of class establishment, relative to 2-choice MTS, by reducing the potential for sample/S- control to arise during training. Results showed, however, that training with 3- and 4-choice MTS did not significantly increase equivalence class establishment and unequivocal evidence of sample/S- control was found in only 1 instance of a 2-choice training and testing structure. Experiment 2 systematically replicated Experiment 1 with 6 additional senior citizens in a 0-s delayed MTS paradigm. As in Experiment 1, equivalence class establishment was not related to number of choice stimuli. The delayed MTS paradigm, however, required fewer training trials to establish the conditional relations and led to more class establishment overall. The results are compared to data from previous studies with younger and older participants.

Stimulus equivalence refers to the substitutability of one stimulus for another in a given context. Stimulus equivalence is a particularly interesting phenomenon because the equivalence of previously unrelated stimuli can arise without direct training (e.g., Sidman, 1971; Spradlin, Cotter, & Baxley, 1973) and the equivalence of numerous stimuli can be established by training only a few stimulus-stimulus relations (e.g., R. R. Saunders, Wachter, & Spradlin, 1988). In most experiments on stimulus

Preparation of the manuscript was supported in part by National Institute of Child Health and Human Development Grants HD02528 and HD18955 to the Schiefelbusch Institute for Life Span Studies, University of Kansas. Requests for reprints should be made to Richard R. Saunders, Institute for Life Span Studies, 1052 Dole, University of Kansas, 1000 Sunnyside Ave., Lawrence, KS 66045. (E-mail: rrsaun@ku.edu).

equivalence, stimuli are presented in a series of conditional discriminations in a matching-to-sample (MTS) format. For example, Stimulus A1 is presented as a sample and following a response to the sample, two or more choice stimuli, such as B1 and B2, are presented simultaneously. A response to B1 and not B2 in the presence of A1 is reinforced. On other trials, a response to B2 and not B1 is reinforced in the presence of sample Stimulus A2. Subsequently, another conditional discrimination is arranged and taught with a combination of additional stimuli (e.g., C1 and C2) and some of the stimuli from the first conditional discrimination (e.g., B1 and B2), as shown in the three panels of Figure 1.

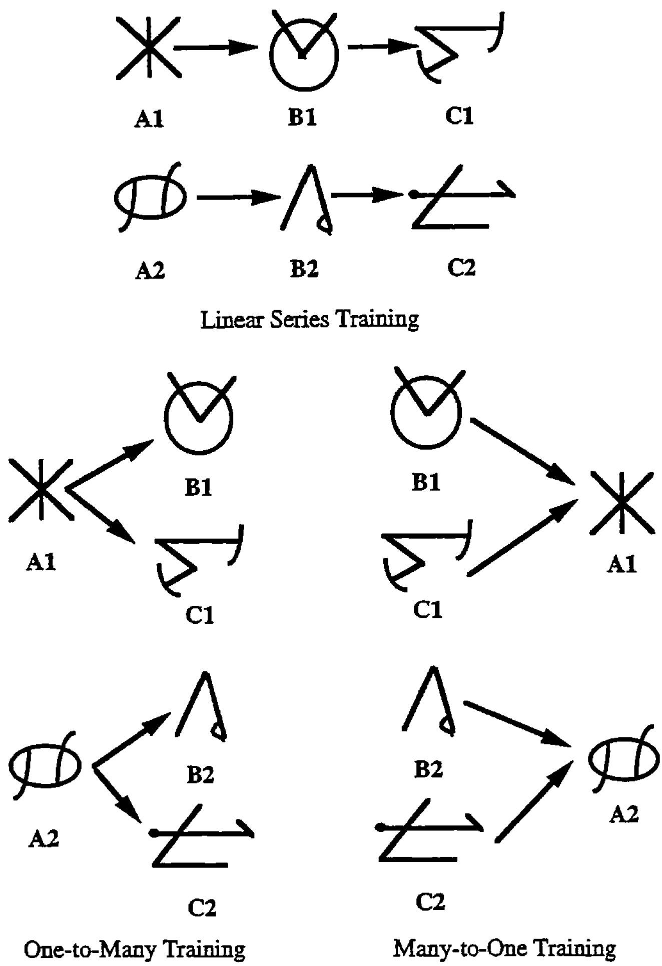

Figure 1. Schematics of three permutations of two conditional discriminations with stimuli in common, leading potentially to two equivalence classes of three stimuli each. Arrows point from samples to choice stimuli and represent trained relations.

The training arrangement in the upper panel of Figure 1 has been referred to as a "linear series" (LS) training structure (K. J. Saunders, Saunders, Williams, & Spradlin, 1993). In this arrangement, relations between the A and C stimuli may develop because the B stimuli participate in both conditional discriminations. Tests for emergent relations between the A and C stimuli are a subset of the complete tests for stimulus equivalence; that is, tests to determine whether the trained conditional relations, AB and BC, are equivalence relations. In alternative training structures, the A stimuli could be employed as sample stimuli in two conditional discriminations, as shown in the lower left panel of Figure 1, or as choice stimuli, as shown in the lower right panel of Figure 1. Most authors now refer to these latter structures as one-to-many (OTM) (Urcuioli & Zentall, 1993) and many-to-one (MTO) (Urcuioli, Zentall, Jackson-Smith, & Steirn, 1989), respectively.

Variations in training structure have interested researchers because structural variables have been implicated as causal in differences in outcomes in equivalence tests. Until recently, nearly all equivalence research showing such differences was conducted with two-choice MTS methods (Fields, Hobbie-Reeve, Adams, & Reeve, 1999; K. J. Saunders et al., 1993; R. R. Saunders, Drake, & Spradlin, 1999; R. R. Saunders et al., 1988; Spradlin & Saunders, 1986). Several researchers have noted that with two-choice MTS procedures, responding during training and testing might induce sample/S- control rather than sample/S+ control (Carrigan & Sidman, 1992; Johnson & Sidman, 1993; Sidman, 1994). Sample/S- control refers to the inference that "correct" responding is a function of "rejecting" the incorrect choice stimulus in the trial as opposed to "selecting" the correct choice (Carrigan & Sidman, 1992). Consistent sample/S- control during training can lead to performances on tests for equivalence that suggest that equivalence classes have been established, but not the classes intended by the experimenter. Hypothetically, in a two-choice LS structure, intended to produce classes consisting of A1, B1, C1 and A2, B2, C2, sample/S- control could produce class compositions of A1, B2, C1 and A2, B1, C2.

According to Carrigan and Sidman (1992), with a shift to additional choice stimuli per trial, sample/S- responding should be reduced, thereby increasing the likelihood that the experimenter-intended classes will be established. MTS preparations with three choice stimuli per trial, for example, would present more than one incorrect stimulus per trial. Carrigan and Sidman reasoned that rejecting just one of several incorrect stimuli would not reliably lead to a response to the correct stimulus and reinforcement. Thus, stimulus control that leads more reliably to reinforcement (i.e., sample/S+ control) should be more likely to arise. In summary, MTS preparations with more than two choice stimuli should foster sample/S+ control and consistent sample/S+ control should lead to test performances that show establishment of the experimenter-intended equivalence classes (Sidman, 1994).

Sidman's predictions regarding class establishment were not upheld in a recent series of experiments (Arntzen & Holth, 1997, 2000; Holth & Arntzen, 1998, 2000). Arntzen and Holth's test protocol differed, however, from that of most other investigators (e.g., Fields et al., 1999; K. J. Saunders et al., 1993; R. R. Saunders et al., 1988). Specifically, most investigators test for equivalence classes in the OTM and MTO structures in Figure 1 by presenting trials with C1 or C2 as the sample stimulus and Stimuli B1 and B2 as the choices (for CB relations), as well as trials with the B stimuli as samples and the C stimuli as choices (for BC relations). Arntzen and Holth's testing in MTO and OTM was limited to CB tests only. In Arntzen and Holth's LS tests, the CA relations were tested, but not the AC relations. Other investigators have tested both CA and AC.

Arntzen and Holth, in the four studies cited, consistently reported low percentages of equivalence class establishment with a LS structure intended to produce three classes of three stimuli each. Adams, Fields, and Verhave (1993) reported research with a LS structure intended to produce two classes of three stimuli each. Adams et al. demonstrated that testing for the AC relations before conducting tests of the CA relations produced more positive results overall than concurrent AC and CA testing. Thus, one purpose of the present study was to determine whether more class establishment would be shown in the LS structure with three intended classes of three stimuli each if AC test trials were included in the test series. Arntzen and Holth's studies also collectively showed more class establishment with OTM structures than MTO structures in certain MTS configurations. A second purpose of the present study was to determine whether these OTM-MTO differences would be replicated if a more typical testing protocol were employed (e.g., BC and CB test trials). A third purpose was to evaluate Sidman's predictions (Carrigan & Sidman, 1992; Sidman, 1994) with a wide array of structure-MTS configurations, overlapping and extending beyond those evaluated by Arntzen and Holth. When the present study was conducted, we did not expect older normal adult participants to perform differently than college students and other younger adults. Data reported subsequently (Wilson & Milan, 1995) challenge this assumption, and the present data are discussed accordingly.

Experiment 1

Method

In Experiment 1, participants received training with six unique sets of stimuli in each of three training structures (OTM, MTO, and LS). Order of training was counterbalanced across participants. The training in the three training structures was also counterbalanced across number of stimuli per intended class, number of intended classes, and number of choice stimuli presented on each trial. Thus, the primary dependent variable was equivalence class establishment. A secondary dependent variable was number of training trials required to meet the criterion for testing for equivalence. That is, does increasing class size, number of classes, or number of choice stimuli increase task acquisition complexity or difficulty?

Participants

Twelve adult volunteers from the community were recruited. Potential participants responded to flyers inviting healthy adults to participate

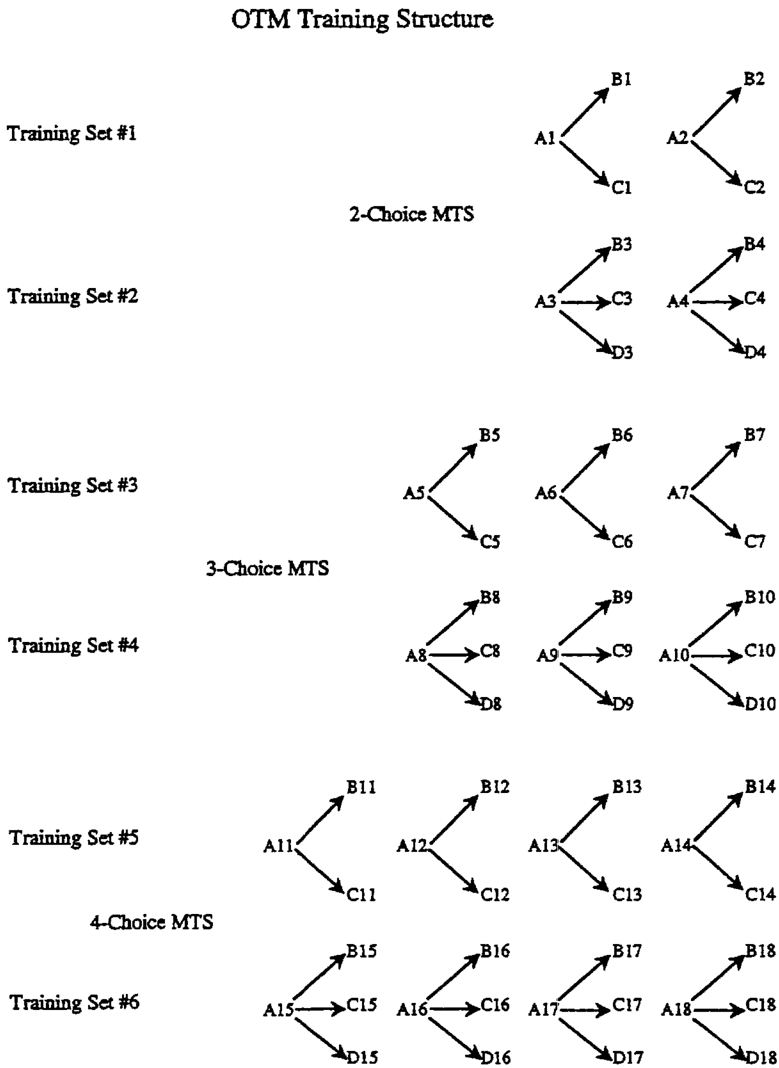

Figure 2. Schematics of the six training sets in the OTM training structure. Arrows point from samples to choice stimuli and represent trained relations. The numbering scheme shows that each training set contained unique stimuli.

in research on learning, while receiving a stipend of $10 per hour. Availability during normal work hours was a requirement and so we targeted people 55 and older. An initial telephone screening was used to exclude individuals who had medical conditions or took medications that might interfere with learning and memory. The ages of 2 male and 10 female participants ranged from 56 to 89 and the extent of formal education ranged from 8 to 18 years.

Apparatus

For the MTS training and equivalence testing, participants sat at a desk in an office cubicle facing a computer monitor. Training was conducted using a Macintosh computer with an ArtMedia 14 inch Trinitron CRT (Model TC1564). The monitor operated with Troll Touch Software (Version 1.8.8) whereby the participant could make a response by touching the screen. A software program was used to program stimulus display sequences and response consequences.

Participant Scheduling and Assignment

Training was scheduled at times and frequencies that were convenient for the participant and was conducted at the Parsons Research Center, located on the grounds of the Parsons State Hospital and Training Center. Participants attended two 90-min or three 1-hr training sessions per week. Six sets of conditional discriminations of increasing complexity were created for each training structure (LS, OTM, MTO), for a total of 18 training sets. Figure 2 provides schematics of the six training sets in the OTM structure as examples. Participants received training in all sets in one structure before proceeding to training in another structure. Two participants were unsystematically assigned to each of the six permutations of possible progression through the three structures, as shown in Table 1.

Training Structures, Stimuli, and Tests for the Properties of Equivalence

All stimuli in the training sets were abstract two-dimensional line

Table 1

Order of Training for Each Participant

Participant	1st structure trained	2nd structure trained	3rd structure trained
1	MTO	LS	OTM
2	MTO	LS	OTM
3	MTO	OTM	LS
4	MTO	OTM	LS
5	OTM	LS	MTO
6	OTM	LS	MTO
7	OTM	MTO	LS
8	OTM	MTO	LS
9	LS	OTM	MTO
10	LS	OTM	MTO
11	LS	MTO	OTM
12	LS	MTO	OTM

drawings similar to those illustrated in Figure 1, presented in black against a white background. Each training set employed unique stimuli. Throughout training, each choice stimulus served as the correct choice on some training trials and as one of the incorrect choices on other trials. Throughout training, the number of choice stimuli presented on a trial was equal to the number of experimenter-intended classes that could be derived from exposure to the training, as shown in Figure 2. Thus, Training Sets 1 and 2 were trained with two-choice MTS, Training Sets 3 and 4 with three-choice MTS, and Training Sets 5 and 6 with four-choice MTS. Table 2 provides an overview of the stimulus organization of the six training sets in each structure.

Table 2

Experimenter-Intended Organization of Stimuli in Training Sets 1-6 for All Training Structures

Training set	Classes intended	Stimuli/class intended	Total unique stimuli in set
1	2	3	6
2	2	4	8
3	3	3	9
4	3	4	12
5	4	3	12
6	4	4	16

Following the training, tests for the properties of equivalence were conducted. Following LS training, some test trials tested for the property of transitivity alone (e.g., AC) and some tested for the properties of symmetry and transitivity combined (e.g., CA). Following MTO and OTM training, only the combined tests could be employed (e.g., BC, CB); there is no test for transitivity alone in the MTO and OTM structures. Hereinafter, the set of test trials for each structure will be referred to collectively as the equivalence test for that structure. Tests for the property of symmetry alone were administered to some participants as explained below. Table 3 provides information on the composition of the equivalence and symmetry tests for each training set, including numbers of unique trial types, numbers of presentations of each unique trial type, and the resulting total number of test trials in each test. For equivalence tests and symmetry

Table 3

Organizations of Tests for Equivalence and Symmetry for Each Training Set

	Equivalence tests			Symmetry tests		
Training set	Unique test trial types	Presentations of each trial type	Total test trials	Unique test trial types	Presentations of each trial type	Total test trials
1	4	2	8	4	2	8
2	12	1	12	6	2	12
3	6	2	12	6	2	12
4	18	1	18	9	2	18
5	8	2	16	8	2	16
6	24	1	24	12	2	24

tests, test trials were presented intermixed among training trials. During tests, no trial-by-trial consequences followed responses to choice stimuli in either training or test trials.

Instructions to Participants

Prior to the start of training, participants were told, "Your task is to use the computer feedback to learn the relationships between the symbols." To begin training, a sample stimulus was presented in the middle of the screen. The participant was instructed to "Touch that symbol," and upon doing so, two choice stimuli were presented. The sample stimulus remained displayed during presentation of the choice stimuli (i.e., simultaneous MTS). The participant was then instructed to "Touch another symbol." These instructions were given only for the first trial in the first training set. Selection of the choice stimulus designated as correct resulted in removal of all stimuli from the screen, presentation of the printed word "correct" on the screen accompanied by a brief audible tone. Selection of a choice stimulus not designated as correct resulted in removal of all stimuli from the screen and presentation of the printed word "wrong" being displayed on the screen. Presentations of the words correct and wrong lasted approximately 1 s each and were followed by a 1-s intertrial interval with a blank screen. Thus, our trial-by-trial feedback was similar to that employed in much prior research on stimulus equivalence with normal adults (e.g., Fields, Adams, Verhave, & Newman, 1990; Fields et al., 1999; Fields, Landon-Jimenez, Buffington, & Adams, 1995; Steele & Hayes, 1991). For comparison, Arntzen and Holth (e.g., Arntzen & Holth, 2000) played taped music during the intertrial interval following correct responses.

Training

Training proceeded similarly in each structure. First, each conditional discrimination (e.g., AB) was taught in isolation to a proficiency criterion of 100% correct. Next, all the conditional discriminations in the training set were intermixed (e.g., AB trials with AC trials in the OTM structure) and trained to a similar proficiency criterion of 100% across a series of trials containing at least one example of each training trial type for that training set. As training sets varied in the number of unique trial types, as a function of the number of stimuli in the set, the proficiency criterion applied to more trials in the higher numbered training sets than the lower numbered sets. Next, training was continued, but without trial-by-trial feedback until the same proficiency criterion was met again.

Prior to training, the participants were told "Call me [the experimenter] if you feel that you are no longer making progress or are 'stuck' on a particular problem." If a participant provided this statement during training of the individual relations (e.g., AB), the session was briefly suspended while the experimenter examined the training record. If the training record revealed that no progress toward the proficiency criterion was evident, the experimenter ended training on that set and training on the next set was begun. The experimenter also could declare a participant to be stuck. This occurred when, at the end of training on a given day, the experimenter noted a pattern of no acquisition of a conditional discrimination. Similarly, a participant could become stuck during the mixed conditional discrimination training or the mixed training without feedback. In these latter cases, rather than terminate the training set, the experimenter reset the computer to provide retraining on selected individual conditional discriminations (e.g., AB) in isolation. The computer then presented the programmed sequence of mixed training trials and mixed training without feedback again, leading to the testing phase. If the proficiency criterion was not met with this remedial procedure, the experimenter discontinued that training set and initiated training with the next training set.

A test for equivalence followed training and is hereinafter referred to as Equivalence Test 1. As shown in Table 3, Sets 1, 3, and 5 had a small number of equivalence test trial types and each was tested twice in Equivalence Test 1. Training with four-member classes (Sets 2, 4, and 6) had more trial types and each trial type was presented once in Equivalence Test 1. Following Equivalence Test 1, the experimenter examined the test results. If the participant's test performance was 83% or more indicative of equivalence class establishment, testing was terminated and training with the next training set was initiated. With the larger training sets, 83% class consistent responding translates to as many as four responses not consistent with equivalence class establishment. If even a small number of inconsistent responses could be attributed to a single tested relation independent of direction (e.g., B1C1, C1B1), conclusions about class establishment would be equivocal, despite the overall high percentage of equivalence-indicative responding (cf., Sidman, 1987). Thus, in addition to a test performance of 83% or better correct, no more than one test error was permitted for any tested relation. Ultimately, only one test performance of 83% or better contained more than one error in any tested relation.

If the class-establishment criteria were not met in Equivalence Test 1, the experimenter initiated a symmetry test (e.g., BA, CA in the OTM structure), consisting of two presentations of each possible symmetry test trial type intermixed among training trials. If the participant's performance on the tests was 83% or more correct, the experimenter initiated a second equivalence test series, Equivalence Test 2. If symmetry was not demonstrated, testing in that training set was terminated. We reasoned that gradual emergence of correct responses in subsequent equivalence tests (e.g., Sidman, Kirk, & Willson-Morris, 1985) was unlikely if the trained relations did not have the property of symmetry. Completion of Equivalence Test 2 terminated testing regardless of the participant's test performance. Participants meeting the class-establishment criteria in Equivalence Test 1 were not tested on symmetry because the property of symmetry in the trained relations is demonstrated by correct performances on the combined tests for transitivity and symmetry. Participants were not informed about any aspect of their test results prior to the end of the study.

Results and Discussion

Equivalence Class Establishment

Table 4 summarizes the data from the training conditions. There were 10 instances in which training was discontinued because of the participant's lack of progress or becoming stuck, indicated in the table by the numbers in parentheses. Training trial data from these instances were

Table 4

Medians of Mean Training Trials to Criterion per Conditional Relation

Training set	Relations trained	OTM	MTO	LS	Configuration summary
1	4	23.5	32.5	15.5	Two-choise
2	6	17.5	15	24 (1)*	median = 19.5; range = 9 to 123
3	6	29.5	33.5	51	Three-choice
4	9	21.5	44	51 (2)	median = 31.5; range = 11 to 173
5	8	39 (1)	42 (1)	54	Four-choice
6	12	40 (2)	42 (1)	43 (2)	median = 42; range = 14 to 169
Training structure summary		OTM median = 33; range = 9 to 169	MTO median = 34; range = 11 to 123	LS median = 32; range = 9 to 173	

* Number of discontinued training sets in parentheses

excluded from the calculations shown in the table. Data for the table were derived as follows: Training trials presented prior to tests for equivalence were summed for each training set for each participant and divided by the number of conditional relations trained in the training set. The median of the participant means was calculated for each training set in each training structure. Further, medians were calculated for each training structure type overall (bottom row) and for all training sets with two-, three-, and four-choice configurations, respectively (far right column). Table 4 shows that the medians for the training structure types (OTM, MTO, and LS) were similar (32-34) although the ranges for the participant means were large for all structure types. Training sets with more choice stimuli per trial, and also more relations trained, tended to require more training trials per relation. In short, increased task complexity was associated with higher numbers of trials to criterion.

Analysis of Equivalence Test 1 results showed equivalence class establishment with 115/206 training sets. In the remaining 91 cases, symmetry tests were conducted and 71/91 tests confirmed that the trained relations had the property of symmetry. Equivalence Test 2 results showed equivalence class establishment with 19/71 training sets for which the property of symmetry had been confirmed. Thus, a total of 134 training sets (62%) led to test performances showing equivalence class establishment. Table 5 shows the results of the final equivalence

Table 5

Percentage Equivalence-Consistent Responses in Each Participant's Final Test in Each Structure*

Structure type	OTM						MTO						LS					
Intended classes**	2	2	3	3	4	4	2	2	3	3	4	4	2	2	3	3	4	4
Stimuli/class	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4
Participant																		
1	**100**	**92**	**92**	**100**	**94**	**100**	50	75	**83**	**94**	**100**	**88**	**100**	58	**100**	**83**	**88**	71
2	38	33	**100**	78	**100**	75	50	33	67	22	25	**88**	**88**	**100**	67	61	62	29
3	**88**	**83**	**92**	**89**	**100**	**88**	**88**	**92**	**83**	**100**	81	**96**	**100**	**83**	**100**	56	75	54
4	62	**100**	17	**83**	81	d/c	25	**100**	**92**	72	**88**	**92**	50	50	**83**	**83**	38	62
5	62	**100**	**100**	**94**	**88**	**92**	**100**	**100**	**100**	**100**	**88**	**96**	**100**	**100**	**83**	**89**	**100**	**88**
6	50	58	25	44	44	50	**100**	**92**	**100**	56	25	**96**	**100**	**92**	**100**	44	25	29
7	**100**	**83**	33	51	**88**	**92**	**100**	**100**	**92**	**100**	d/c	d/c	**100**	42	**83**	67	**94**	d/c
8	**100**	8	67	**89**	69	42	62	**100**	**100**	**89**	**94**	**96**	**100**	75	**100**	**83**	**88**	50
9	**100**	**100**	**100**	**94**	**100**	**83**	**100**	**100**	**100**	**83**	**100**	**100**	38	50	**100**	d/c	56	42
10	**100**	**100**	**100**	**100**	**94**	71	**100**	**100**	**100**	**94**	**100**	**96**	**88**	50	50	44	81	50
11	**100**	**92**	**92**	**94**	d/c	d/c	62	58	**92**	**89**	**94**	**86**	**100**	d/c	50	d/c	69	50
12	**88**	**100**	75	**94**	69	**88**	**100**	**83**	**100**	**100**	**100**	**100**	**100**	58	**83**	44	**100**	d/c

* Performances meeting test criteria are shown in bold and discontinued training sets are indicated with d/c.

** The number of choice stimuli presented in each trial equaled the number of intended classes.

test series for each participant, organized by training structure, with performances meeting the class-establishment criteria in bold. The table shows that when the criteria were met, the test trial percentage generally exceeded the 83% minimum requirement.

Analyses of errors on training trials during Equivalence Test 1 and 2 indicated that failures to show equivalence class establishment were not a function of lost baselines; that is, errors on training trials during testing were rare. Analyses were also conducted to estimate the likelihood that additional testing would have led to a higher percentage of training sets with equivalence class establishment (i.e., gradual emergence). For this analysis, Equivalence Test 1 and Equivalence Test 2 results were compared. Equivalence Test 2 results had the same or fewer correct responses than Equivalence Test 1 in 30/52 cases (58%). Of the 22 who showed some improvement from Test 1 to Test 2, 10 improved but nevertheless had 50% or fewer correct responses in Equivalence Test 2. Of the remaining 12, only 4 improved to within one correct response of meeting the criteria for class establishment in the Equivalence Test 2. This analysis suggests that further testing might have led to only a small number of additional criterion-level performances.

Figure 3 shows the number of training sets meeting the criteria for equivalence class establishment by training structure and number of choice stimuli per trial. Visual inspection of these data suggest that increasing the number of choice stimuli above two does not increase the likelihood of equivalence class establishment, except marginally in the MTO structure. These results were also subjected to statistical

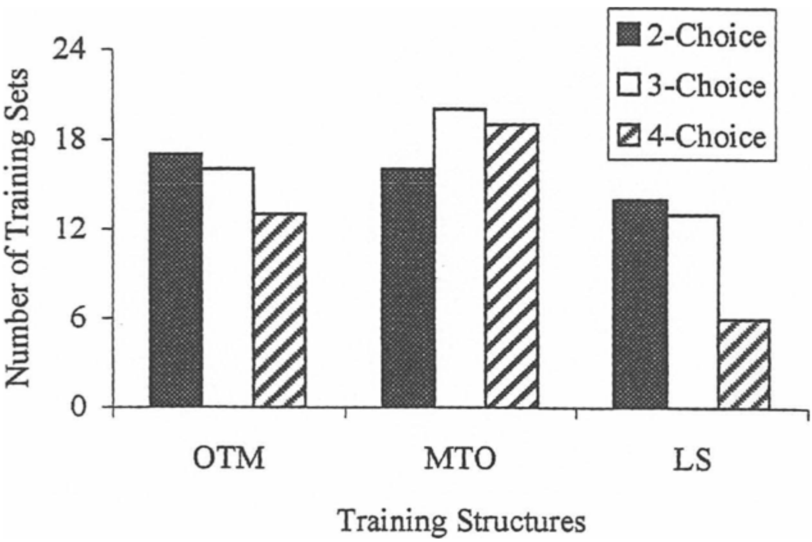

Figure 3. Graph showing number of training sets leading to equivalence class establishment by training structure type, broken down by number of choice stimuli per trial in training and testing in Experiment 1.

analysis. Table 6 shows the results organized as number of training sets leading to class establishment for the statistical analysis. The 10 training sets wherein training was not completed were counted as sets in which equivalence classes were not established. A Friedman analysis indicated that equivalence class establishment was not related to MTS configuration, $x^2 = 4.537$, $df = 2$, $p = .103$. A Friedman analysis is applicable to data from k related samples, such as number of tests passed, that are hypothesized to vary as a function of another variable, such as MTS configuration.

Table 6

Number of Training Sets Leading to Equivalence Class Establishment*

	MTS configurations		
Participant	2-choice	3-choice	4-choice
1	3	6	5
2	2	1	2
3	6	5	3
4	2	4	2
5	5	6	6
6	4	2	1
7	5	3	3
8	3	5	3
9	4	5	4
10	5	4	3
11	3	4	2
12	5	4	4

* Maximum of 6 per configuration

An Analysis to Control for the Effects of More Than One Nodal Stimulus per Class

The design of the present experiment permitted comparison of training structures counterbalanced for number of intended classes, number of stimuli per intended class, and number of choice stimuli per trial. Unfortunately, such counterbalancing precludes also counterbalancing for number of nodes. Nodes are the stimuli that provide the linkages between two other stimuli (Fields & Verhave, 1987). With the trained relations AB and BC, the B stimuli are the nodal stimuli, as shown in the LS schematic in Figure 1. Adding CD training in the LS structure creates a second set of nodal stimuli, the C stimuli. In the present design, Training Sets 2, 4, and 6 in the LS structure presented training such that there would be two nodal stimuli per class. An increase in the number of nodes in training structures has been linked to delayed or reduced establishment of equivalence classes (Fields et al., 1990, 1995). Our results were consistent with the literature. As shown in Table 5, the participants showed less equivalence class establishment in the LS structure in the columns headed by four stimuli per class (Training Sets 2, 4, and 6) than in the columns headed by three stimuli per class (Training Sets, 1, 3, and 5). Because of these differences, we repeated the Friedman analysis on only the results from structures with three stimuli and one node per intended class (Training Sets 1, 3, and 5). The data for this analysis are arrayed in Table 7. As in the analysis with the complete data set, the number of choices showed no significant interaction with test results, $x^2 = 2.600$, $df = 2$, $p = .273$.

Thus, overall, the results of the study demonstrate that training and testing with more than two choice stimuli did not improve the proportion of equivalence tests with positive results relative to two-choice procedures.

Analyses of Sample/S- Control of Responding

Carrigan and Sidman (1992) described how sample/S- control in a

Table 7

Number of One-NobeTraining Sets Leading to Equivalence Class Establishment*

	MTS configurations		
Participant	2-choice	3-choice	4-choice
1	2	3	3
2	1	1	1
3	3	3	1
4	0	2	1
5	2	3	3
6	2	2	0
7	3	2	2
8	2	2	2
9	2	3	2
10	3	2	2
11	2	2	1
12	3	2	2

* Maximum of 3 per configuration

two-choice two-node LS structure would likely affect test results. They outlined how consistent sample/S- control during training would lead to 0% experimenter-intended or correct responding on one-node tests and 100% correct responding on two-node tests. Thus, sample/S- control arising in a two-choice one-node LS structure (Training Set 1) is easy to evaluate. Sample/S- control should produce 0% correct responses on tests for equivalence. Only P4 and P9 failed to show equivalence-class establishment in the LS Training Set 1. Their correct test trial percentages were 50 and 38, respectively; thus, they did not show evidence of sample/S- control. The pattern expected from sample/S- control for a two-choice, two-node structure was observed in the data of only 1 participant. In LS Training Set 2, in Equivalence Test 1, P7 made 25% correct responses on the one-node trials and 100% correct responses on the two-node trials. In Equivalence Test 2, his two-node performance remained 100% and he made one correct response on the one-node trials (12.5%) for an overall 42% correct. As indicated by Sidman (1994), it is difficult to distinguish sample/S- control in the data from experiments with three or more intended classes in the LS structure, and in all OTM and MTO structures. Sample/S- control arising in these other structures can lead to indeterminate response patterns in terms of percentages correct.

Comparison of the Present Results with Previous Similar Research

The present study demonstrated the establishment of equivalence classes with 134/216 training sets. Table 8 shows the results of the

Table 8

Comparison of Percentages of Tests Showing Equivalence with Four Studies with Younger Adults*

Training structure	OTM		MTO		LS	
Intended outcome	Present study	Their studies	Present study	Their studies	Present study	Their studies
2 classes of 3 stimuli	67	100	58	50	83	
2 classes of 4 stimuli	75		75		33	
3 classes of 3 stimuli	58	94	92	73	75	24
3 classes of 4 stimuli	75	67	75	78	33	
4 classes of 3 stimuli	58		67		42	
4 classes of 4 stimuli	50		92		8	

* Arntzen and Holth (1997, 2000) and Holth and Arntzen (1998, 2000)

present study contrasted with the results reported by Arntzen and Holth (1997, 2000) and Holth and Arntzen (1998, 2000) with abstract stimuli, simultaneous MTS, and younger adult participants (note: studies employing familiar rather than abstract stimuli are not included). The comparison in the table is offered with a caveat: Arntzen and Holth applied a more stringent criterion for equivalence class establishment, but tested only a subset of the possible test trial types. Thus, the comparison is based on general rather than precise similarities in experimental protocols. The table shows that in Arntzen and Holth's research, training leading to classes of three stimuli each resulted in more equivalence class establishment in OTM than MTO, with the reverse being true for classes of four stimuli each. In the present study, we also had mixed results for MTO and OTM. Also, as noted above, across Arntzen and Holth's studies, the three-choice one-node LS structure led to equivalence class establishment infrequently, with 14/59 participants (24%) as shown in the table. In the present study, 9/12 or 75% showed equivalence class establishment with that structure (Training Set 3).

The present data from LS Training Set 3 support our hypothesis that the absence of AC test trials in Arntzen and Holth's test protocol might have contributed to their substantially lower percentage of equivalence-indicative test results. Using a test protocol referred to as "simultaneous," several investigators have observed low frequencies of equivalence class establishment (Buffington, Fields, & Adams, 1997; Fields et al., 1995; Fields et al., 1997) relative to a "simple-to-complex" protocol (e.g., Adams et al., 1993). In the simultaneous protocol, AC and CA test trial types are presented intermixed in the same test sessions, as we did in our LS test sessions; in the simple-to-complex protocol, participants are exposed to AC trial types in one or more sessions before encountering CA trial types. It may be that a protocol without any AC tests inhibits equivalence-indicative performances as much or more than the simultaneous protocol.

Arntzen and Holth studies tested younger participants (e.g., college students, currently employed adults) than did we, prompting speculation that age or an aspect of aging was a relevant variable with respect to performances. Indeed, Wilson and Milan (1995) trained and tested a group of young adults (19-22 yr) and a group of elderly adults (62-81 yr) with a three-choice procedure intended to produce three classes of three abstract stimuli in the OTM structure. The percentage showing equivalence class establishment for the two groups was 80% and 45%, respectively. Clearly additional research is needed for firm conclusions about the influences of aging on equivalence class establishment.

Experiment 2

Method

By the conclusion of Experiment 1, 6 additional senior citizens (all female between 58 and 75 years old) had volunteered for possible participation. We elected to replicate Experiment 1 with these individuals with one change in procedure. In Experiment 2, we employed a 0-s delayed MTS (DMTS) training and testing paradigm. We hypothesized that the absence of the sample stimulus during responding to the choice stimuli would make acquisition of the conditional discriminations more difficult and possibly have a negative impact on equivalence class establishment. In this paradigm, a touch to the sample stimulus simultaneously resulted in removal of the sample stimulus and presentation of the choice stimuli. With only 6 participants, we assigned 2 to each of these training orders:

MTO-LS-OTM, OTM-MTO-LS, and LS-OTM-MTO. All other procedures remained the same, as did the stimuli used.

Results and Discussion

Table 9 summarizes the training data from Experiment 2. The table shows that, contrary to our hypothesis, trials to criterion for testing were generally fewer in Experiment 2 than in Experiment 1, both for training structure types (median 15-25) and for training sets clustered by number of choice stimuli per trial. For example, the median training trials per relation for four-choice MTS training sets was 42 in Experiment 1, but only 24 in Experiment 2. Further research will be necessary to determine whether these differences were a function of the 0-s DMTS paradigm or a function of participant differences across the experiments.

Table 9

Medians of Mean Training Trials to Criterion per Conditional Relation

Training set	# relations trained	OTM	MTO	LS	Configuration summary
1	4	15	14 (1)*	55.5	Two-choice
2	6	11	15.5	16	median = 14; range = 8 to 266
3	6	25	16.5	24.5	Three-choice
4	9	13	12.5	23	median = 15; range = 9 to 123
5	8	22	25	22.5	Four-choice
6	12	19	25.5	33	median = 24; range = 10 to 74
Training structure summary		OTM median = 15; range = 8 to 123	MTO median = 16; range = 8 to 89	LS median = 25; range = 9 to 266	

* Discontinued training set

Overall, and also contrary to our hypothesis, 92/108 training sets (85%) led to criterion performances on tests for equivalence classes in Experiment 2. Table 10 shows the equivalence class establishment results for the individual participants, with performances meeting criteria in bold. The table shows that only one training set, with P13, was discontinued due to lack of progress in training. The table also shows that equivalence classes were established with all but one other training set in the OTM and MTO structures—considerably more frequently than seen in Experiment 1. The number of training sets meeting the criteria in the LS structure was slightly higher, also, in Experiment 2 than in Experiment 1. The data in Table 10 also provide two additional instances of apparent sample/S- control with two-choice LS procedures: P14 in the first LS training set with 0% equivalence-indicative responses and P17 in the second LS training set, with only 1/8 one-node responses consistent with

Table 10

Percentage Equivalence-Consistent Responses in Each Participant's Final Test in Each Structure*

Structure type	OTM						MTO						LS					
Intended classes**	2	2	3	3	4	4	2	2	3	3	4	4	2	2	3	3	4	4
Stimuli/class	3	4	3	4	3	4	3	4	3	4	3	**4**	3	4	3	4	3	4
Participant																		
13	**88**	**92**	**92**	**89**	**88**	**92**	d/c	**92**	**92**	**89**	50	**92**	**100**	**100**	75	67	56	46
14	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	0	50	**83**	**100**	**100**	**100**
15	**100**	**100**	**100**	**100**	**100**	**96**	**100**	**100**	**92**	**100**	**88**	**100**	**100**	**100**	**100**	**100**	**94**	**100**
16	**100**	**83**	**83**	**100**	**88**	**92**	**100**	**92**	**83**	**100**	**88**	**96**	75	**100**	**100**	**83**	**94**	75
17	**88**	**100**	**100**	**83**	**100**	**96**	**100**	**92**	**100**	**100**	**100**	**83**	50	25	58	**83**	**88**	71
18	**100**	**100**	**92**	**94**	**100**	**92**	**100**	**100**	**100**	**100**	**100**	**100**	**83**	**92**	**83**	78	**100**	54

* Performances meeting test criteria are shown in bold and discontinued training sets are indicated with d/c.
** The number of choice stimuli presented in each trial equaled the number of intended classes.

class establishment, but 3/4 two-node responses consistent with class establishment, for an overall 25%.

Figure 4 shows the number of training sets meeting the criteria for equivalence class establishment by training structure and number of choice stimuli per trial. It is clear from these data that training sets with three or four choice stimuli per trial did not improve test outcomes relative to two-choice procedures in Experiment 2. As with the training trial data, more research is needed to sort procedural effects from participant variables.

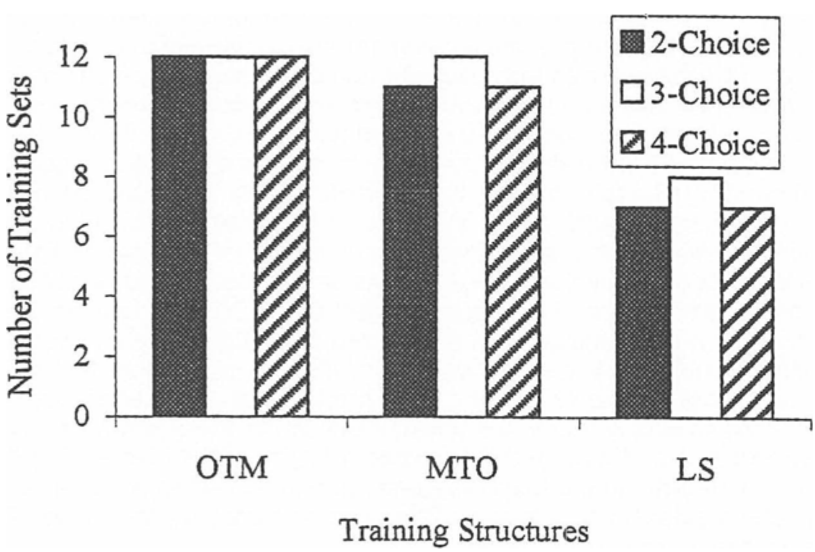

Figure 4. Graph showing number of training sets leading to equivalence class establishment by training structure type, broken down by number of choice stimuli per trial in training and testing in Experiment 2.

General Discussion

Our results generally replicate Arntzen and Holth's findings that three- and four-choice MTS configurations did not consistently increase equivalence class establishment overall relative to two-choice MTS configurations. We also demonstrated this outcome with training structures not evaluated by Arntzen and Holth, with the 0-s DMTS paradigm, and with older participants. Boelens (2002) opined that the evidence for abandoning two-choice procedures in favor or three- or four-choice procedures was weak. The present data provide convincing support for Boelens' opinion.

Across Experiments 1 and 2 in the present study, 5/18 participants failed to show class establishment in the LS structure leading to two classes of three stimuli each. Of these, only one had an error pattern indicative of sample/S- control. In the LS structure leading to two classes of four stimuli each, 10/18 failed to show class establishment. Of these, only two had a response pattern indicative of sample/S- control. These results are not conclusive proof that sample/S- control did not arise; certainly, mixed control can occur, resulting in something between 100% class-consistent responding and the pure patterns of sample/S- control described by Carrigan and Sidman (1992). Thus, the present data suggest either that sample/S- control rarely arose or arose in competition with sample/S+ control (i.e., mixed control), resulting in sub-criteria test performances.

A reasonable question is, "If sample/S- control rarely arose, why was equivalence class establishment not more common as the number of choice stimuli was increased"? A possible answer is that increasing the number of choice stimuli in the present experiment also increased the number of classes. This method also increased the number of stimuli in each training set. By this process, the number of possible simple discriminations in each training set increased also. As R. R. Saunders and Green (1999) pointed out, the LS and OTM training structures do not, during training, "require" discrimination of each stimulus in a potential class from each stimulus in every other potential class. Nevertheless, those discriminations not required during training are required for class-consistent responding in trials that test for equivalence. In short, the procedural changes that might reduce the emergence of sample/S- control during training also increase the number of discriminations not required during training but required for class-consistent test performances. Although training in the MTO structure requires acquisition of all the prerequisite discriminations, our method substantially increased the number of discriminations that had to be acquired and maintained throughout testing. Thus, increasing the number of choices in the MTS problems, at least as we accomplished it, may have resulted in no net gain with respect to increasing the probability of equivalence class establishment.

Acquisition of simple discriminations is only one possible source of variability. We employed brief, but fairly informative instructions to our participants. We told them that they were to learn relationships between symbols. We have no way of determining whether our reference to the stimuli as symbols or our suggestion that some symbols were related to others facilitated or impeded equivalence class establishment. Further research is required to determine how the instructions might have altered the equation. However, given the results of Experiment 2, in which the participants received the same instructions, one might conclude that other experimental variables may be far more influential than brief instructions. The participants in Experiment 2 had more frequent class establishment in the MTO and OTM structures with the DMTS procedure than did the participants of Experiment 1 with the simultaneous MTS procedure. The LS data from both experiments also point to the importance of training structure, particularly nodal aspects of structure, in the establishment of equivalence classes. Overall, the results of these and other experiments suggest many sources of stimulus control could be involved in any particular instance of success or failure of equivalence class establishment.

We failed to replicate Arntzen and Holth's results' pattern with the LS structure leading to three intended classes of three stimuli each. The contrast suggests that research explicitly aimed at the effects of test protocols is warranted. Our results with the OTM structure leading to three classes of three stimuli were similar, however, to those reported by Wilson and Milan (1995) with older participants. In Experiment 1, 58% of our participants showed equivalence class establishment in this structure, whereas 45% of theirs showed class establishment. It is interesting to note, however, that with the DMTS paradigm, 100% of our participants in Experiment 2 showed class establishment in this structure. This is closer to Arntzen and Holth's cumulative percentage (94%) with younger participants than Wilson and Milan's (80%) with younger participants.

The results with the 0-s DMTS are somewhat surprising. The DMTS paradigm led to fewer trials to criteria for testing than the simultaneous paradigm. DMTS participants also met criteria for equivalence class establishment in a higher percentage of training sets; indeed, meeting the criteria in every OTM training set. One possibility for these differences, other than participant characteristics, is that the DMTS paradigm prompts or promotes development of precurrent behavior early in the experiment (Holth & Arntzen, 1998, 2000). Skinner (1968) linked precurrent behavior to intellectual self-management. Precurrent behavior refers to behavior that changes the person or the environment to increase the likelihood that an ensuing response is reinforced. Skinner provided the example of learning to look at the sample stimulus simultaneously with pressing it as a precurrent to correct responding in a MTS task. It seems possible that the removal of the sample stimulus in the DMTS paradigm might engender sample naming, for example, during the interval between sample removal and responding to a choice stimulus. Sample naming and similar strategies or precurrents could occur during simultaneous MTS procedures as well. Perhaps, however, the DMTS paradigm prompted such strategies earlier in Experiment 2; thus, more training sets led to class establishment in Experiment 2. Naming should facilitate stimulus discrimination and that facilitation may explain the increase in class establishment in the MTO and OTM structures. If naming was employed in the LS structures, it clearly did not have the same facilitative effect on class establishment.

References

ADAMS, B. J., FIELDS, L., & VERHAVE, T. (1993). Effects of test order on intersubject variability during equivalence class formation. *The Psychological Record,* 43, 133-152.

ARNTZEN, E., & HOLTH, P. (1997). Probability of stimulus equivalence as a function of training design. *The Psychological Record,* 47, 309-320.

ARNTZEN, E., & HOLTH, P. (2000). Equivalence outcome in single subjects as a function of training structure. *The Psychological Record,* 50, 603-628.

BOELENS, H. (1994). A traditional account of stimulus equivalence. *The Psychological Record,* 44, 587-605.

BUFFINGTON, D. M., FIELDS, L., & ADAMS, B. J. (1997). Enhancing the formation of equivalence classes by pretraining of other equivalence classes. *The Psychological Record,* 47, 1-20.

CARRIGAN, P. F., & SIDMAN, M. (1992). Conditional discrimination and equivalence relations: A theoretical analysis of control by negative stimuli. *Journal of the Experimental Analysis of Behavior,* 58, 183-204.

FIELDS, L., ADAMS, B. J., VERHAVE, T., & NEWMAN, S. (1990). The effects of nodality on the formation of equivalence classes. *Journal of the Experimental Analysis of Behavior,* 53, 345-358.

FIELDS, L., HOBBIE-REEVE, S. A., ADAMS, B. J., & REEVE, K. F. (1999). Effects of training directionality and class size on equivalence class formation by adults. *The Psychological Record,* 49, 703-723.

FIELDS, L., LANDON-JIMENEZ, D. V., BUFFINGTON, D. M., & ADAMS, B. J. (1995). Maintained nodal-distance effects in equivalence classes. *Journal of the Experimental Analysis of Behavior,* 64, 129-154.

FIELDS, L., REEVE, K. F., ROSEN, D., VARELAS, A., ADAMS, B. J., BELANICH, J., & HOBBIE, S. A. (1997). Using the simultaneous protocol to study equivalence class formation: The facilitating effects of nodal number and size of previously established equivalence classes. *Journal of the Experimental Analysis of Behavior,* 67, 367-389.

FIELDS, L., & VERHAVE, T. (1987). The structure of equivalence classes. *Journal of the Experimental Analysis of Behavior,* 48, 317-332.

HOLTH, P., & ARNTZEN, E. (1998). Stimulus familiarity and the delayed emergence of stimulus equivalence or consistent nonequivalence. *The Psychological Record,* 48, 81-110.

HOLTH, P., & ARNTZEN, E. (2000). Reaction times and the emergence of class consistent responding: A case for precurrent responding? *The Psychological Record,* 50, 305-337.

JOHNSON, C., & SIDMAN, M. (1993). Conditional discrimination on equivalence relations: Control by negative stimuli. *Journal of the Experimental Analysis of Behavior,* 59, 333-347.

SAUNDERS, K. J., SAUNDERS, R. R., WILLIAMS, D. C., & SPRADLIN, J. E. (1993). An interaction of instructions and training design on stimulus class formation: Extending the analysis of equivalence. *The Psychological Record,* 43, 725-744.

SAUNDERS, R. R., DRAKE, K. M., & SPRADLIN, J. E. (1999). Equivalence class establishment, expansion, and modification in preschool children. *Journal of the Experimental Analysis of Behavior,* 71, 195-214.

SAUNDERS, R. R., & GREEN, G. (1999). A discrimination analysis of training-structure effects on stimulus equivalence outcomes. *Journal of the Experimental Analysis of Behavior,* 72, 117-137.

SAUNDERS, R. R., WACHTER, J., & 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. (1971). Reading and auditory-visual equivalences. *Journal of Speech and Hearing Research,* 14, 5-13.

SIDMAN, M. (1987). Two choices are not enough. *Behavior Analysis,* 22, 11-18.

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

SIDMAN, M., KIRK, B., & WILLSON-MORRIS, M. (1985). Six-member stimulus classes generated by conditional-discrimination procedures. *Journal of the Experimental Analysis of Behavior,* 43, 21-42.

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

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.

SPRADLIN, J. E., & SAUNDERS, R. R. (1986). The development of stimulus classes using match-to-sample procedures: Sample classification versus comparison classification. *Analysis and Intervention in Developmental Disabilities,* 6, 41-58.

STEELE, D., & HAYES, S. C. (1991). Stimulus equivalence and arbitrarily applicable relational responding. *Journal of the Experimental Analysis of Behavior,* 56, 519-555.

URCUIOLI, P. J., & ZENTALL, T. R. (1993). A test of comparison-stimulus substitutability following one-to-many matching by pigeons. *The Psychological Record,* 43, 745-759.

URCUIOLI, P. J., ZENTALL, T. R., JACKSON-SMITH, P., & STEIRN, J. N. (1989). Evidence for common coding to many-to-one matching: Retention, intertrial interference, and transfer. *Journal of Experimental Psychology: Animal Behavior Processes,* 15, 264-273.

WILSON, K. M., & MILAN, M. A. (1995). Age differences in the formation of equivalence classes. *Journal of Gerontology,* 50B, 212-218.