Psychol Rec (2015) 65:615–625
DOI 10.1007/s40732-015-0132-5

ORIGINAL ARTICLE

Sorting: An Alternative Measure of Class Formation?

Erik Arntzen[1] · **Aina Norbom**[1] · **Lanny Fields**[2]

Published online: 18 June 2015

Abstract In initial sorting tests, 16 participants did not assign stimuli to experimenter-defined classes. Then, the baseline relations for 5-member equivalence classes were trained using matching-to-sample (MTS) trials. Follow-up MTS tests assessed class formation. Regardless of outcome, another sorting test assessed delayed class formation if classes had not formed or class-maintenance if classes emerged during the MTS test. Classes were not formed by 11 participants, emerged on a long-delayed basis in the sorting test for 2, emerged on a delayed basis in the first MTS test for 2 others, and emerged immediately in the MTS test for 3 others. The latter three participants then attempted to form new equivalence classes. After baseline training, emergence was assessed with a sorting test administered immediately thereafter, and was followed serially with an MTS and final sorting test. Responding in the first sorting tests demonstrated the immediate emergence of the stimulus classes for these participants. The MTS test results implied that the classes that emerged in the sorting test were actually equivalence classes. For the two participants who showed delayed class formation, class integrity was maintained during the follow-up sorting and MTS tests. Two other participants showed class formation in the last and final sorting test. The remaining nine did not show class formation. Because the sorting tests were completed 90 % faster than MTS tests, they provided a quick and reliable alternative to MTS tests for the tracking of class formation.

Keywords Sorting · Categorization · Equivalence class formation · College students

Introduction

An equivalence class contains a finite number of perceptually disparate stimuli, all of which have become related to each other and can be used in an interchangeable manner (Sidman 1994). The vast majority of all experimental studies of equivalence class formation involve the training of baseline conditional relations among the stimuli in a set that is followed by testing for the emergence of untrained relations that can be derived from the stimuli in the baseline relations. Collectively, the untrained relations have also been called emergent or derived relations. In addition, the production of class-consistent responding assessed by derived-relations probes documents the properties of reflexivity, symmetry, and transitivity (Fields and Verhave 1987; Sidman 1994; Sidman and Tailby 1982).

Most articles that have explored equivalence classes have focused on the identification of variables that have influenced the likelihood of class formation. Other articles have used equivalence classes in combination with other behavioral processes to build laboratory models that simulate complex forms of human behavior such as syntax (Mackay and Fields 2009), meaning (e.g., Arntzen et al. 2015; Fields et al. 2012; Tyndall et al. 2004), and contextually determined symbol classification (Bush et al. 1989; DeRosse and Fields 2010). Yet other articles have used equivalence-based instruction to establish

Author Note Parts of the results of this study were presented at the 7th International Conference of the Association for Behavior Analysis, Merida, Mexico.

✉ Erik Arntzen
erik.arntzen@equivalence.net

[1] Department of Behavioral Science, Oslo and Akershus University College, PO Box 4, St. Olavs Plass, 0130 Oslo, Norway

[2] Queens College and The Graduate School of the City University of New York, Flushing, NY, USA

college level course content (e.g., Fields et al. 2009b; Fienup and Critchfield 2011; Fienup et al. 2015; Pytte and Fienup 2012; Walker and Rehfeldt 2012). Most of these studies used training and testing trials conducted in a matching-to-sample (MTS) format (Arntzen 2004), with many fewer conducted with trials administered in a stimulus pairing yes-no format (Fields et al. 2009a).

In many experiments, the test trials presented to some of the participants produced class-indicative responding upon their first presentation, which documented the immediate emergence of the equivalence classes. In contrast, the same test trials presented to other participants produced class-indicative responding only after extended testing, thereby documenting the delayed emergence of the equivalence classes (e.g., Fields and Watanabe-Rose 2008; Holth and Arntzen 1998; Kennedy 1991).

As noted above, in most studies of equivalence class formation, training and testing trials were administered in the same format. These results might suggest that the formation of equivalence classes requires the use of training and testing trials that are of the same format. The results of a few recent experiments, however, prove that that is not the case. In these experiments, trials in different formats were used to establish the baseline relations and test for the emergence of the derived relations that document the formation of an equivalence class. For example, respondent-type training was used to establish baseline relations among the stimuli in a set, after which trials in an MTS format showed the emergence of the derived relations that document equivalence class formation (e.g., Leader et al. 1996). While supportive, the results of one experiment do not definitively prove a point. One goal of the present experiment is to determine whether similar outcomes can be obtained with another mix of trial formats: MTS-based trials to establish the baseline relations followed by a sorting test to assess the emergence of equivalence classes. A positive outcome would provide additional support for the view that the formation of equivalence classes is relatively independent of the trial formats used to establish baseline relations and measure the emergence of the classes (Dymond and Rehfeldt 2001; Fields et al. 2014).

A sorting test can take a number of forms (Arntzen et al. 2011; Eikeseth et al. 1997; Grimm 2011; Smeets et al. 2000). One form involves the use of a deck of cards, each of which contains one of the stimuli that is a member of each potential equivalence class. The deck of cards is shuffled to "randomize" the ordering of the cards with respect to class membership and structural position in a class, and then given to a participant with the instruction to "put these stimuli into groups as you feel like." The presence of all of the experimenter-defined classes would be documented by the placement of the cards into the number of stacks that corresponds to the number of experimenter-defined classes, with each stack containing all of the cards for one of the experimenter-defined classes only.

To date, a number of experiments have incorporated this particular form of sorting test to study of equivalence class formation (Arntzen et al. 2011; Dickins 2011; Eilifsen and Arntzen 2009; Fields et al. 2012, 2014). Most of these experiments began with the administration of a sorting test to determine whether the experimenter-defined classes were present prior to training. In all cases, class-consistent performances were never obtained. Thereafter, baseline relations were established using MTS training trials, and were followed by MTS test trials to assess the emergence of the equivalence classes. Some participants did show class formation, but others did not. When the sorting test was re-administered at the completion of the MTS test, the performances were concordant with the MTS tests. Specifically, the participants who showed class formation in the previously administered MTS test stacked the cards in accordance with equivalence, whereas most of the participants who did not show class formation in the MTS test did not stack the stimuli in accordance with equivalence. For the remaining few who did not show class formation in the MTS test, class-indicative responding occurred during the sorting test, which showed the delayed emergence of the classes (Arntzen et al. 2011). Finally, the sorting tests were completed in 10 % of the time needed to complete the MTS test.

Although the use of sorting to study equivalence classes is promising, a number of issues require clarification. In all of the experiments that have used sorting, the results of the sorting tests could not be used to document the formation of the classes for the following reason. The establishment of the baseline relations was followed immediately by MTS-based emergent relations tests and then the sorting test was presented. If class-consistent responding occurred during the MTS test, it would document the immediate emergence of the equivalence classes. Thus, class formation could not be assessed with the sorting test because it would have already emerged in the previously administered MTS test. Because of its placement after both baseline-relation training and administration of the MTS test, the sorting test could only document the maintenance of the equivalence classes in a different testing format. These limitations, however, are not an inherent property of a sorting test. The present experiment determined whether a sorting test could be used to measure the immediate emergence of equivalence classes by establishing the baseline relations for three 3-node, 5-member equivalence classes followed by the administration of a sorting test, the results of which could document the immediate emergence of the classes.

By its nature, a sorting test involves assessing the presence of only a small proportion of the relations in an equivalence class. Thus, even if the sorting performances are indicative of equivalence classes, they do not assess the emergence of all class-based relations. To obviate this potential interpretive limitation, the sorting test in the present experiment was followed by a traditional MTS test for equivalence class

formation. The presence of class-indicative responding therein would support the view that the results of the sorting test reflected the immediate emergence of the equivalence classes. Result such as these, then, would constitute a first demonstration of the use of a sorting test to document the immediate emergence of equivalence classes. Finally, the sorting test was also used to track the delayed emergence of stimulus classes.

Method

Participants

The participants in the present experiment were 16 individuals, some of whom were university students and others who had different occupations, who ranged in age from 20–43 years (average=28.4 years). Seven of the participants were females and nine were males. The participants were recruited through personal contacts at the university college. The participants had not previously participated in similar experiments and all had no explicit knowledge of stimulus equivalence.

Upon arrival at the laboratory, all participants received a written consent form, which they were told to read, and if they agreed, to sign the document. Participants were informed briefly about the experiment in which they were about to participate through the consent form. They were notified when recruited that the experiment could take up to four hours. After signing the consent form, participants were asked if they had any questions. Additionally, they were told their results would be anonymous. Finally, they were informed that they could withdraw from the experiment at any time without penalty. All participants were debriefed after the experimental session and they were presented with their results from the MTS and sorting programs.

Setting and Apparatus

The experiments were conducted in a 1×1.5-m cubicle located in a quiet office that was approximately 3×4.5-m in size. The cubicle contained a chair for the participant that faced a desk with a computer. During the experimental sessions, the experimenter sat outside the office in order to avoid being asked questions about the tasks and did not interrupt the participants.

All aspects of the experiment were administered on an HP EliteBook laptop computer (Model 8760w) that contained an Intel Core i5-2540 M processor running at 2.6 GHz. The computer screen was 36.8 cm wide by 23 cm high. Each stimulus was presented in an area that was 5 cm high and 5 cm wide and was solid black on a white background. The sample stimulus was presented in the approximate center of the screen with four "invisible" squares in the corners of the screen. On each trial, three comparisons were presented in three of the invisible squares in randomly determined corner locations. Responding involved placing the cursor on any part of a stimulus and clicking the mouse button. Figure 1 displays the stimuli used as the members of classes 1, 2, and 3. Figure 2 has the same format as Fig. 1 and displays the stimuli used in classes 4, 5, and 6.

Experimental Design

The phases of the experiment are indicated in Table 1. The experiment began with a sorting test (PreSRT-1) to determine whether the stimuli in experimenter-defined classes were related to each other prior to training. Thereafter, the simultaneous protocol was used to establish equivalence classes (Fields et al. 1997). The protocol was used because it typically produces relatively poor yields, and, thus, is a preparation that is sensitive to variables that influence the likelihood of equivalence class formation, such as mode of testing. First, all of the baseline relations were established on a concurrent basis for three 3-node, 5-member classes that had a linear series training structure represented as A/B/C/D/E (training of baseline relations, or TBR). Thereafter, the emergence of equivalence classes was assessed by the presentation of the two test blocks MTS-1a and MTS-1b, each of which contained baseline relations and symmetry, transitivity, and equivalence probes. All of the trials in these test blocks were presented in an MTS format. The MTS-1b test was then followed by a sorting test (SRT-1).

The outcomes on these tests determined the remaining procedures administered to each participant. One outcome was that the MTS-1a test produced mastery levels of class-consistent responding, which indicated the immediate emergence of equivalence classes 1, 2, and 3. Further, the continued occurrence of mastery-level responding in the MTS-1b and SRT-1 tests would document the maintenance of the classes. These three performances are indicated as "y/y/y" in the outcomes column in Table 1. Any participant who showed this outcome then attempted to form three new equivalence classes that consisted of stimulus sets 4, 5, and 6. This procedure began with a sorting test with the stimuli in classes 4, 5, and 6 (PreSRT-2), which was followed by the concurrent training of the baseline relations that were the prerequisites of equivalence classes 4, 5, and 6 (TBR). Acquisition of the baseline relations was followed immediately with the administration of a sorting test (SRT-2) to assess the immediate emergence of the new equivalence classes. That sorting test was then followed by two MTS tests (MTS-2a and 2b) to determine whether the classes that emerged during the sorting test (SRT-2) remained intact and also showed the defining properties of an equivalence class. The procedure ended with the re-administration of another post-class-formation sorting test (SRT-3). If the classes did not emerge on an immediate basis

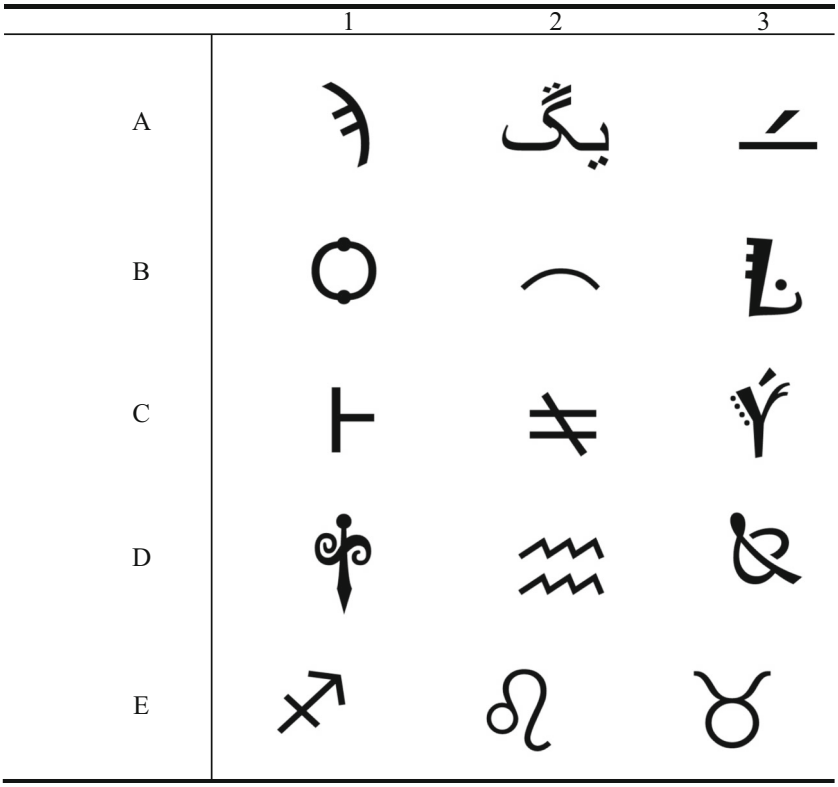

Fig. 1 Stimulus Set 1 used as members of potential classes 1, 2, and 3. The columns (1–3) represent the classes and the rows (A–E) represent the members of the class; that is, A1/B1/C1/D1/E1 were in the same experimenter-defined class

during SRT-2, the sequence of tests conducted after SRT-2 permitted the tracking of equivalence class formation on a long-delayed basis.

It was also possible that classes 1, 2, and 3 did not emerge on an immediate basis during the MTS-1a test. Those performances that did not show class-indicative responding on

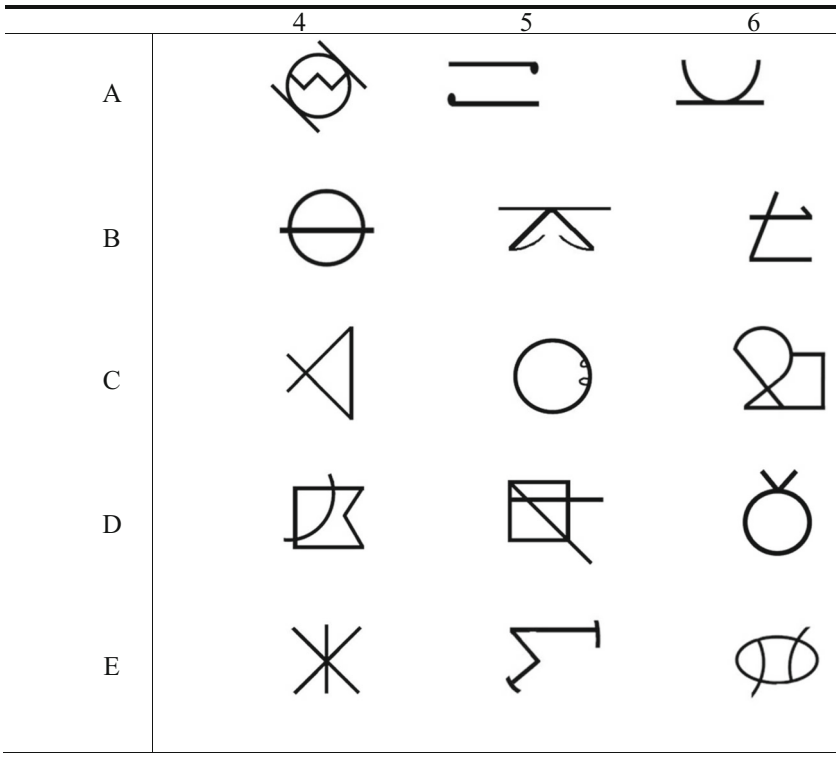

Fig. 2 Stimuli used as members of potential classes 4, 5, and 6. The columns represent the stimuli in classes 4, 5, and 6, and the rows (A–E) represent the members of each particular experimenter-defined class (e.g., A4/B4/C4/D4/E4)

Table 1 An Overview of Training and Testing in Phases 1 and 2

Stimulus Sets	Phase 1	Possible Outcome	Stimulus Sets	Phase 2
1, 2, and 3	PreSRT/TBR/MTS-1a/MTS-1b/SRT-1	n/n/n or n/y/y	1, 2, and 3	MTS-2a/MTS-2b/SRT-2
		y/y/y	4, 5, and 6	PreSRT/TBR/SRT-2/MTS-2a/MTS-2b/SRT-3

Notes. All participants started with Stimulus Sets 1, 2, and 3. Depending on the outcome, they were assigned either of two sequences: one sequence with repeated training and testing with the same stimulus sets, or another sequence with training and testing on new stimulus sets. "TBR"=training baseline relations; "MTS"=matching-to-sample. "PreSRT"=sorting test conducted prior to the training of baseline relations for classes 1, 2, and 3 or 4, 5, and 6. "MTS-1a" and "MTS-1b" each represent half of the MTS tests for the emergence of classes 1, 2, and 3. "SRT-1" is a sorting test used to assess the presence or emergence of classes 1, 2, and 3. "MTS-2a" and "MTS-2b" each represent half of the matching-to-sample tests for the emergence of classes 4, 5, and 6. "SRT-2" and "SRT-3" are sorting tests used to assess the presence or emergence of classes 4, 5, and 6. Each "n" and "y" in the outcome column indicates the absence or presence, respectively, of mastery level class-indicative performances in the MTS-1a, MTS-1b, and SRT-1 tests. The first, second, and third y or n in each string indicates the possible outcome of the MTS-1a, MTS-1b, and SRT-1 tests, respectively

MTS-1a, MTS-1b, or SRT-1 tests are represented as "n/n/n" in the outcomes column in Table 1. Those performances that indicated the delayed emergence of the classes in the MTS-1b test and maintenance in the subsequent SRT-1 test are represented as "n/y/y." In these conditions, the SRT-1 test was followed with a second administration of two MTS tests, MTS-2a and MTS-2b, and a final sorting test (SRT-2), all of which enabled the tracking of class formation on a long-delayed basis or the maintenance of already emergent classes.

Phases in the Experiment The experiment included some combination of the following phases: a sorting test, a phase used to establish the baseline relations for the equivalence classes with trials conducted in an MTS format, and an MTS test for the emergence of the equivalence classes. The details of each of these phases are described next.

Sorting Test Once a participant signed the consent form, the first of three sorting tests was administered with a deck of cards in which each card contained one of the stimuli in each of the classes used in the experiment. The sorting test was presented on the computer. It began with the presentation of the instruction, "put these into groups as you feel like," in the presence of the stimuli presented on the screen. The stimuli were presented as a deck of cards. Participants were informed that to see the next stimulus in the deck, the top card had to be moved to a different location on the screen that was close to other related stimuli so that they formed a cluster that was separated from other clusters that contained stimuli from different sets. The stimuli could be moved around using a computer mouse. When the participant was finished, he or she called the experimenter, who took a screenshot of all of the stimuli on the screen. That information was then printed and used to determine which stimuli were clustered together as participant-defined classes.

Baseline Acquisition After the first sorting test, the participants were presented with the MTS-based program to train the baseline relations. They were asked to push the start button when ready, and after pushing "start," an instruction for the MTS training was shown in Norwegian, which is presented as an English translation:

> A stimulus will appear in the middle of the screen. Click on this by using the computer mouse. Three other stimuli will then appear. Choose one of these by using the computer mouse. If you choose one of the stimuli we have defined as correct, words like "very good," "excellent" and so on will appear on the screen. If you press an incorrect stimulus, the word "wrong" will appear on the screen. At the bottom of the screen, the number of correct responses you have made will be counted. During some stages of the experiment, the computer will not tell you if your choices are correct or incorrect. However, based on what you have learned, you can get all the tasks correct. Please do your best to get everything right. Good Luck!

All of the baseline relations were established on a concurrent basis, which involved the presentation of training trials in a training block that was repeated until all relations were acquired. The experiment involved the formation of AB, BC, CD, and DE relations for each of the three classes, which yielded 12 relations to be learned. Thus, the training block contained 60 trials that included five iterations of each relation. The relations in a block were presented in a randomized sequence. The block was repeated until the baseline relations were acquired, which was defined as responding correctly on at least 90 % of the trials in a block. During these blocks, programmed consequences were provided for all trials during the acquisition of these relations. The stimuli used as programmed consequences included text such as "great," "good," or "excellent," etc. when responses were correct, and "wrong" whenever responses were incorrect, as defined by the experimenter.

Maintenance of Baseline Relations Once the baseline relations were acquired, they were maintained in the presence of

reduced programmed consequences and eventually no programmed consequences. This was accomplished with re-administration of the same training block but with the "thinning" of programmed consequences from 100 to 75 % of the trials in the block. That block was re-administered until accuracy in a block was at least 90 % correct. Thereafter, the cycle was repeated with only 25 % programmed consequences, and eventually with 0 % programmed consequences. When a participant responded accurately on at least 90 % of the trials in a block with 0 % programmed consequences, the MTS test for derived relations was presented.

MTS Test Each MTS test evaluated the emergence of the symmetrical relations BA, CB, DC, and ED, transitive relations AC, AD, AE, BD, BE, and CE, and equivalence relations CA, DA, EA, DB, EB, and EC. The test consisted of 300 trials, divided into 60 baseline trials, 60 symmetry trials, 90 transitivity trials and 90 equivalence trials. Class formation was documented if at least 95 % of all trials in the test block produced class-consistent selections. For purposes of analysis, the 300 test trials were divided into 2 test blocks, with the first containing trials 1–150 and the second containing trials 151–300.

Results

Sorting Tests Figure 3 is a screen shot of the outcome of one sorting test by one participant. The stimuli were distributed into four clusters that contained 3, 3, 5, and 4 stimuli, respectively. The stimuli in each cluster corresponded to one participant-defined class. For example, the cluster in the bottom right contains the stimuli, D1, C3, A2, and B3. The 3-digit string beneath each cluster is a summary representation of the stimuli in the cluster, where the first, second and third digits indicate the number of cards in the cluster that were members of experimenter-defined classes 1, 2, and 3, respectively. For example, the string designated as 321 represents a participant-defined 6-member class that contains three stimuli from class 1, two stimuli from class 2, and one stimulus from class 3. While not produced by this participant, 3-digit strings designated as 500, 050, and 005 would represent three participant-defined classes that corresponded to the stimuli in the 5-member experimenter-defined classes 1, 2, and 3, respectively.

Sorting and Classes Prior to Training The results of the sorting test conducted prior to any training are presented in the PreSRT segment of Table 2. Each row in an SRT segment contains 3-digit strings that correspond to the clusters produced by one participant. The pre-class formation-sorting test (PreSRT) produced 67 clusters across participants. Taken as a group, the participants produced three to seven clusters that contained one to six stimuli per stack. Most of the stacks contained stimuli from one to three of the experimenter-defined classes. Assuming that each cluster of at least three stimuli corresponds to a participant-defined class, none of participants showed the presence of an experimenter-defined class prior to training.

Formation of Equivalence Classes 1, 2, and 3 The upper portion of Table 2 shows the results for participants who did and did not show the immediate emergence of equivalence classes 1, 2, and 3. Each row represents the data for one participant. With respect to columns, the left, middle, and rightmost segments labelled PreSRT, SRT-1, and SRT-2, respectively, summarize the results of the sorting tests conducted prior to training (Pre-SRT), in the first sorting test conducted after baseline acquisition (SRT-1), and the second sorting test (SRT-2). The columns labelled MTS-1a and -1b list the accuracy of responding during the first and second half of the first MTS test and assessed the emergence of derived relations. The columns headed MTS-2a and 2b list the accuracy of

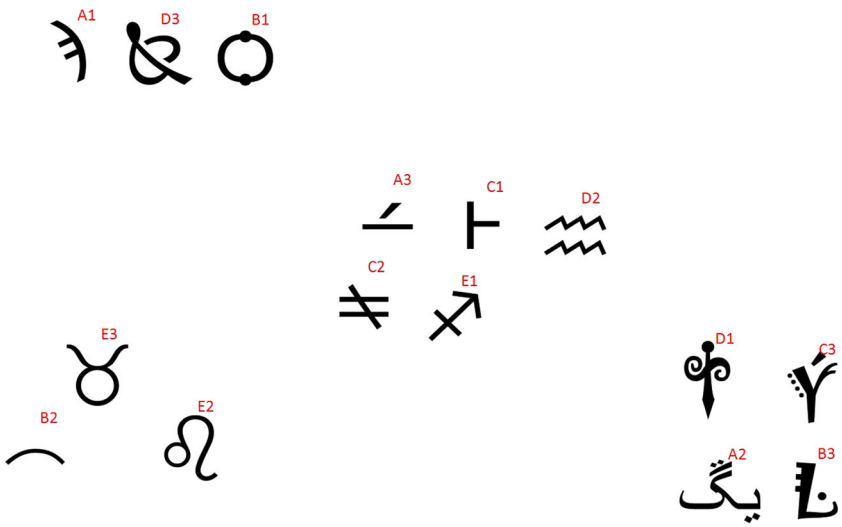

Fig. 3 Screen shot of how one of the participants organized the cards in the presorting test

Table 2 The Results for All Participants

PN	Pre-SRT-1							TBR	MTS-1a	MTS-1b	SRT-1							MTS-2a	MTS-2b	SRT-2					
5203	321	031	203					360	**98**	**100**	**500**	**050**	**005**												
5207	321	221	013					360	**97**	**100**	**500**	**050**	**005**												
5210	210	201	021	112	011			260	**99**	**100**	**500**	**050**	**005**												
5212	221	221	113					480	84	**99**	**500**	**050**	**005**					**99**	**100**	**500**	**050**	**005**			
5202	310	011	021	012	201			1080	77	**98**	**500**	**050**	**005**					**100**	**100**	**500**	**050**	**005**			
5208	210	201	021	012	111			540	65	76	**500**	040	015					78	79	**500**	**050**	**005**			
5204	311	122	122					420	51	62	**500**	051	004					75	87	**500**	**050**	**005**			
5211	201	021	012.	321				660	66	65	300	200	030	020	002	003		65	65	300	230	025			
5201	201	221	021	112				840	57	57	220	120	203	012				57	57	320	220	015			
5213	401	021	021	113				1020	54	60	300	210	021	012	012			60	60	201	210	121	012	011	
5214	201	010	010	002	011	110	211	540	49	46	200	200	020	110	003	002	020	46	46	323	232				
5216	211	101	120	111	012			480	49	38	211	101	120	012	111			38	38	221	223	111			
5209	021	221	302	011				720	47	24	313	221	012					24	24	313	221	021			
5206	121	221	012	201				480	46	42	300	221	022	012				42	42	300	221	022	012		
5205	021	012	321	201				960	37	36	311	200	020	010	012	002		36	36	311	020	200	010	002	012
5215	400	010	122	113				1020	35	40	410	022	121	002				40	40	410	122	023			
PN	**Pre-SRT-2**							**TBR**			**SRT-2**							**MTS-2a**	**MTS-2b**	**SRT-3**					
5203	221	031	203					300			**500**	**050**	**005**					**93**	**97**	**500**	**050**	**005**			
5207	230	121	112					420			**500**	**050**	**005**					**95**	**99**	**500**	**050**	**005**			
5210	211	212	021	123				480			**500**	**050**	**005**					**100**	**100**	**500**	**050**	**005**			

Notes. Each row is for a different participant. The top section shows performances when participants attempted to form equivalence classes 1, 2, and 3. The bottom section contains data for three participants who attempted to form equivalence classes 1, 2, and 3 in the first phase. The row heads indicate participant number (PN). Names of each phase with the left-to-right listing indicate the temporal order of administration. In the SRT test, each 3-digit triad represents one cluster of stimuli produced by a participant. The triads in **bold** font indicate subject-defined classes that are the same as the experimenter-defined classes. "PreSRT-1" and "-2"=sorting tests conducted prior to training of the baseline relations for classes 1, 2, and 3 (PreSRT-1) and classes 4, 5, and 6 (PreSRT-2), respectively. "TBR"=training of baseline relations; "MTS"=trials presented in a matching-to-sample test; "SRT"=a sorting test conducted after training

responding during the first and second half of the second MTS test and assessed the emergence of derived relations. Finally, the TBR column contains the trials needed to acquire the baseline relations.

Three of the participants (5203, 5207, and 5210) responded in accordance with stimulus equivalence during the first two MTS tests (MTS-1a and MTS-1b), thereby demonstrating the immediate emergence of the equivalence classes 1, 2, and 3. In the subsequently administered sorting test (SRT-1), each of the three clusters contained cards from only one experimenter-defined class, thereby documenting the maintenance of the classes in different testing contexts (SRT vs. MTS).

Two of the participants (5212 and 5202) showed subcriterion performances in the first MTS test block (MTS-1a) that increased to mastery in the second MTS test block (MTS-1b). These data demonstrated the rapid albeit delayed emergence of equivalence classes 1, 2, and 3. The performances in the subsequently administered sorting test (SRT-1) and the second set of MTS tests (MTS-2a and MTS-2b) showed the maintenance of these classes across testing formats.

Two other participants (5208 and 5204) responded with low levels of class-indicative responding in the first two MTS tests (MTS-1a and -1b). The performances in the subsequently administered sorting test (SRT-1) approached mastery, with three five-member subject-defined clusters, each of which contained at least four of the five stimuli in an experimenter defined class. This was followed by a return to lower levels of class-indicative responding in the next two MTS tests (MTS-2a and MTS-2b), but they were somewhat higher than the values observed in the first two MTS tests. Finally, the last SRT test (SRT-2) produced class-indicative responding that documented the long-delayed emergence of equivalence classes 1, 2, and 3. Thus, the sorting test tracked the delayed emergence of stimulus classes. Because no MTS test was presented thereafter, it was not possible to determine whether these stimulus classes had the properties of equivalence classes. The results of prior experiments, however, support the expectation that such a test, if it had been conducted, would have shown criterion-level responding and the delayed emergence of the equivalence classes.

Finally, nine of the participants did not show class-indicative responding in any of the MTS or sorting tests after the acquisition of the baseline relations. None of these participants formed equivalence classes 1, 2, or 3.

To summarize, no experimenter-defined classes were documented before training. Three participants showed the immediate emergence of equivalence classes, four showed the delayed emergence of the classes, and nine did not form classes. When immediate emergence occurred, it took place in the MTS test that immediately followed training. The delayed emergence of these classes occurred during the MTS tests for two participants and in the sorting tests for two other participants.

Sorting and the Immediate Emergence of Equivalence Classes 4, 5, and 6 After participants 5203, 5207, and 5210 formed equivalence classes 1, 2, and 3 on an immediate basis, they were placed into a new condition designed to determine whether a sorting test could document the immediate emergence of new equivalence classes. The results of this condition are presented in the lower portion of Table 2. It began with a sorting test (SRT-1) conducted with the stimuli from the three new sets (4, 5, and 6). In this test, the participants produced three or four participant-defined classes, each of which contained stimuli from at least two of the experimenter-defined classes. None of the participants, however, produced any experimenter-defined classes.

The subsequent acquisition and maintenance of the baseline relations was followed immediately with the administration of a sorting test (SRT-2) to assess the emergence of stimulus classes 4, 5, and 6. In this sorting test, all three of the participants distributed the cards into three clusters, each of which contained the five stimuli that belonged to one of the experimenter-defined classes. This performance then documented the immediate emergence of three stimulus classes.

While this sorting test documented the formation of stimulus classes, the classes may not have been equivalence classes because only an indeterminate subset of all of the derived relations was presented in the sorting test (but see Bortoloti et al. 2014; McIlvane and Dube 1990). This interpretive option was evaluated by the subsequent administration of the MTS-2a and MTS-2b tests of derived relations, which contained all baseline, symmetry, transitivity, and equivalence relations from the classes. Seen in the MTS-1a and -1b columns of the lower half of Table 2, mastery in the first MTS tests documented the presence of all relations in the equivalence classes, along with the maintenance of the previously emergent classes. Most likely, then, the classes that emerged in the sorting test were actually equivalence classes. These results support the view that a sorting test can document equivalence class formation.

The final sorting test (SRT-3) also resulted in the placement of the stimuli into clusters that corresponded to those that were members of the experimenter-defined classes. Thus, the classes were maintained regardless of test type. These results support the view that equivalence classes are relatively independent of the testing procedures used to document their establishment and maintenance.

Acquisition of Baseline Relations The baseline relations for classes 1, 2, and 3 were acquired in 260–1080 trials across participants (see TBR column in the upper half of Table 2). After training, the three participants who showed the immediate emergence of the classes (5203, 5207, and 5210) took an average of 327 trials to acquire the baseline relations, whereas the remaining 13 participants acquired the baseline relations in an average of 711 trials. Thus, the speed of acquiring the

baseline relations was significantly faster for the participants who formed equivalence classes on an immediate basis than for those who did not ($t(14)=2.65$, $p=0.019$). In addition, a one-way Pearson r test found that immediate emergence of the equivalence class was inversely correlated with baseline acquisition speed ($r(16)=.492$, $p=0.025$). Indeed, baseline acquisition speed accounted for 24 % of the variance (r^2) in the immediate emergence of equivalence classes 1, 2, and 3. The baseline relations for classes 4, 5, and 6 were acquired after an average of 400 trials. While more than 327 trials were needed to learn the baseline relations for classes 1, 2, and 3, the difference was not significant.

Discussion

Replication of Prior Experiments Prior experiments have shown a very high concordance between the formation of equivalence classes measured using MTS-formatted derived-relations tests and the maintenance of those classes measured with post-class-formation sorting tests (Arntzen et al. 2014; Fields et al. 2014; Nartey et al. 2014; Nedelcu et al. 2015; Travis et al. 2014). When classes were formed, post-class formation sorting tests showed class maintenance; when classes did not form, the post-class-formation sorting tests also showed the absence of experimenter defined categorization of the stimuli. The same concordance was found in the present experiment for those who did and did not form equivalence classes 1, 2, and 3.

Furthermore, the results replicated the finding that a rather low number of participants respond in accordance with stimulus equivalence when trained baseline relations are arranged in the linear-series training structure using a simultaneous training and testing protocol (see Arntzen, 2012 for an overview). These low yields occurred during the administration of test blocks that involved the random presentation of all baseline relations along with probes to assess the emergence of symmetrical, transitive, and equivalence relations. In contrast, when the simple-to-complex protocol has been used, and all of the probes have been presented serially and in a programmed order, yields are very high (Adams et al. 1993). Thus, it appears that the low yields obtained when using the simultaneous protocol can be attributed to the concurrent presentation of all derived-relations probes in a given test block.

Sorting and the Immediate Emergence of Equivalence Classes As just mentioned (Arntzen et al. 2014; Fields et al. 2014), sorting tests demonstrated the maintenance of equivalence classes. The sorting tests in previous experiments, however, could not document the emergence of equivalence classes because they were not administered right after the training of baseline relations. This matter was addressed in the present experiment by the administration of a sorting test right after the establishment of the baseline relations for classes 4, 5, and 6; all three of the participants responded in a manner that showed the immediate emergence of stimulus classes 4, 5, and 6. This outcome is the first demonstration of the immediate emergence of stimulus classes by use of a sorting test instead of a derived-relations test conducted in an MTS format. Were these classes, however, actually equivalence classes?

That matter was addressed in the present experiment by considering the results of the MTS tests that were administered after the sorting test. The MTS tests assessed the properties of symmetry, transitivity, and the combined properties of symmetry and transitivity among the stimuli in the nominally defined 3-node, 5-member classes, and produced 95–100 % class-consistent responding. By implication, then, the classes documented by the outcomes of the sorting tests appear to have the definitional properties of equivalence classes.

The generality of this finding can also be extended by replication with more participants, classes that have different sizes and nodal structures, training conducted using procedures like those described by the simple-to-complex and complex-to-simple protocols (Adams et al. 1993), and sorting tests conducted in a number of formats, modes of administration, and in the context of a wide range of instructions (Arntzen et al. 2011; Eikeseth et al. 1997; Grimm 2011; Smeets et al. 2000).

As noted above, the tracking of equivalence class formation with a sorting test was successful for participants who had previously formed other equivalence classes. Whether sorting can be used to confirm equivalence class formation for participants who have not shown prior equivalence class formation can be determined with additional research.

Finally, an additional experiment can be conducted to evaluate whether a stimulus class documented by a sorting test has other properties of an equivalence class, such as function transfer. For example, many experiments have shown that a response trained to one member of an equivalence class generalizes completely to all of the remaining members of that class and not to the members of other equivalence classes (Augustson and Dougher 1997; Augustson et al. 2000; Belanich and Fields 2003; Fields and Garruto 2009). Thus, after documenting the emergence of a class using a sorting test, would a response trained to one class member generalize to the other members of the same class but not to the members of other classes? If so, the class defined by the outcome of a sorting test would be manifesting another property of an equivalence class. Thus, such an outcome would further strengthen the view that a sorting test can document the formation of equivalence classes.

Sorting and Delayed Emergence of Equivalence Classes In addition to tracking the immediate emergence of equivalence classes, the results obtained with two participants

(5204 and 5208) showed, for the first time, that sorting tests also tracked the delayed emergence of classes. The sorting tests, however, were intermingled with MTS-based tests. Thus, it is possible that the outcomes of the sorting tests could have been influenced by the intervening MTS tests and/or the sheer number of tests used to track class formation.

When the outcomes of the MTS and sorting tests were compared, for some participants, the results of the sorting tests showed the delayed emergence of the classes whereas those of the MTS tests did not. Thus, it is possible that the sorting test was more sensitive to the delayed emergence of the classes than the traditional MTS tests. Additional research will be needed to evaluate this possibility.

Economy of Sorting Tests A sorting test is very easy to administer and to negotiate by a participant. Of equal or perhaps more import, the sorting test was completed in 1.5 to 2 min on average, whereas the MTS test was completed in an average of 30 min. Thus, the time needed to establish equivalence classes in applied and basic settings could be substantially reduced by tracking the emergence of the classes with sorting tests instead of MTS tests. When a sorting test is administered, a participant is presented with a randomly shuffled deck of cards that contains one card for each stimulus in a class of N members. Assuming that the participant then assigns the top card in the deck to a stack that corresponds to a potential class, the participant is presented with N-1 relations for a given class: the stimulus in a stack plus the stimulus at the top of the deck. In contrast, for a class of N stimuli, the typical derived relations test involves the monitoring of the emergence of (N^2 – N) relations in a potential class: N-1 baseline relations, N-1 symmetry probes, and (N^2 -2 N -1) transitive and/or equivalence relations (Fields and Verhave 1987). Thus, a sorting test involves the assessment of class formation using many fewer probes than the number used during a typical derived relations test.

Further, because the sorting test involves the presentation of a randomly shuffled deck, the probes that are being evaluated have to be randomly selected for a given participant and across participants. In contrast, a typical derived relations test assesses the emergence of all symmetry relations, transitive relations, and equivalence relations. Thus, one could question the reliability of a sorting test as a measure of equivalence class formation. That issue, however, does not appear to be critical because the outcomes of the MTS tests conducted after the sorting tests showed the presence of all derived relations. Indeed, these results raise a question regarding the necessary number and types of probes that are needed to document the emergence of an equivalence class. Additional research will be needed to obtain an answer to that question.

Trial Format and Class Formation As noted in the introduction, most studies of equivalence class formation have used the same type of trial format to establish the baseline relations to assess the emergence of the derived relations that confirmed class formation. Leader et al. (1996), however, showed equivalence class formation by establishing the baseline relations with respondent-type stimulus pairings and testing for class formation using MTS trials. Thus, the formation of equivalence classes does not depend on the use of trials presented in the same format for training and testing.

The present experiment also showed the formation of equivalence classes by using training and testing conducted in different trial formats; the baseline relations were established using trials administered in an MTS format, whereas class formation was assessed using trials presented in a sorting format. The positive results obtained in the present experiment, then, provide further support for the view that similarities or differences in the conditions of training and testing do not constrain the definition of equivalence classes (Fields et al. 2012, 2014; Smeets et al. 2000).

Summary This experiment represents the first demonstration of the use of a sorting test to document equivalence class formation. This finding reduces any procedural constraints that have to be considered when defining equivalence classes. Sorting tests were completed 10 times faster than a typical MTS-based test for class formation. Thus, the use of sorting tests should significantly reduce the time needed to establish equivalence classes in applied and basic research settings. The sorting test tracked class formation by sampling a subset of emergent relations. Thus, documenting all emergent relations may not be needed to show equivalence class formation.

Conflicts of Interest There are no conflicts of interest to declare concerning the three authors.

References

Adams, B. J., Fields, L., & Verhave, T. (1993). Effects of test order on intersubject variability during equivalence formation. *The Psychological Record, 43*, 133–152.

Arntzen, E. (2004). Probability of equivalence formation: Familiar stimuli and training sequence. *The Psychological Record, 54*, 275–291. Retrieved from http://thepsychologicalrecord.siuc.edu/index.html.

Arntzen, E. (2012). Training and testing parameters in formation of stimulus equivalence: methodological issues. *European Journal of Behavior Analysis, 13*, 123–135. Hentet fra http://www.ejoba.org/.

Arntzen, E., Braaten, L. F., Lian, T., & Eilifsen, C. (2011). Response-to-sample requirements in conditional discrimination procedures. *European Journal of Behavior Analysis, 12*, 505–522. Retrieved from Retrieved from http://www.ejoba.org/.

Arntzen, E., Nartey, R. K., & Fields, L. (2014). Identity and delay functions of meaningful stimuli: Enhanced equivalence class formation. *The Psychological Record*, 349–360. doi: 10.1007/s40732-014-0066-3.

Arntzen, E., Nartey, R. K., & Fields, L. (2015). Enhancing responding in accordance with stimulus equivalence by the delayed and relational

properties of meaningful stimuli. *Journal of the Experimental Analysis of Behavior, 103*, 524–541. doi:10.1002/jeab.152.

Augustson, E. M., & Dougher, M. J. (1997). The transfer of avoidance evoking functions through stimulus equivalence classes. *Journal of Behavior Therapy and Experimental Psychiatry, 28*, 181–191. Retrieved from ISI:A1997XZ11200002.

Augustson, E. M., Dougher, M. J., & Markham, M. R. (2000). Emergence of conditional stimulus relations and transfer of respondent eliciting functions among compound stimuli. *The Psychological Record, 50*, 745–770. Retrieved from ISI: 000165227100008.

Belanich, J., & Fields, L. (2003). Generalized equivalence classes as response transfer networks. *The Psychological Record, 53*, 373–413. Retrieved from http://thepsychologicalrecord.siuc.edu/index.html.

Bortoloti, R., Pimentel, N., & de Rose, J. (2014). Electrophysiological investigation of the functional overlap between semantic and equivalence relations. *Psychology and Neuroscience, 7*, 183–191. doi:10.3922/j.psns.2014.015.

Bush, K. M., Sidman, M., & de Rose, T. (1989). Contextual control of emergent equivalence relations. *Journal of the Experimental Analysis of Behavior, 51*, 29–45. doi:10.1901/jeab.1989.51-29.

DeRosse, P., & Fields, L. (2010). The contextually controlled, feature-mediated classification of symbols. *Journal of the Experimental Analysis of Behavior, 93*, 225–246. doi:10.1901/jeab.2010.93-225.

Dickins, D. W. (2011). Transitive inference in stimulus equivalence and serial learning. *European Journal of Behavior Analysis, 12*, 523–555. Retrieved from Retrieved from http://www.ejoba.org/.

Dymond, S., & Rehfeldt, R. A. (2001). Supplemental measures and derived stimulus relations. *Experimental Analysis of Human Behavior Bulletin, 19*, 8–12. Retrieved from http://www.eahb.org/NewSitePages/BulletinHomepage.htm.

Eikeseth, S., Rosales-Ruiz, J., Duarte, A., & Baer, D. M. (1997). The quick development of equivalence classes in a paper-and-pencil format through written instructions. *The Psychological Record, 47*, 275–284.

Eilifsen, C., & Arntzen, E. (2009). On the role of trial types in tests for stimulus equivalence. *European Journal of Behavior Analysis, 10*, 187–202. Retrieved from http://www.ejoba.org/.

Fields, L., & Garruto, M. (2009). Optimizing linked perceptual class formation and transfer of function. *Journal of the Experimental Analysis of Behavior, 91*, 225–251. doi:10.1901/jeab.2009.91-225.

Fields, L., & Verhave, T. (1987). The structure of equivalence classes. *Journal of the Experimental Analysis of Behavior, 48*, 317–332. doi:10.1901/jeab.1987.48-317.

Fields, L., & Watanabe-Rose, M. (2008). Nodal structure and the partitioning of equivalence classes. *Journal of the Experimental Analysis of Behavior, 89*, 359–381. doi:10.1901/jeab.2008-89-359.

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. doi:10.1901/jeab.1997.67-367.

Fields, L., Doran, E., & Marroquin, M. (2009a). Equivalence class formation in a trace stimulus pairing two-response format: Effects of response labels and prior programmed transitivity induction. *Journal of the Experimental Analysis of Behavior, 92*, 57–84. doi:10.1901/jeab.2009.92-57.

Fields, L., Travis, R., Roy, D., Yadlovker, E., de Auguiar-Rocha, L., & Sturmey, P. (2009b). Equivalence formation: A method for teaching statistical interaction. *Journal of Applied Behavior Analysis, 42*, 575–593. doi:10.1901/jaba.2009.42-575.

Fields, L., Arntzen, E., Nartey, R. K., & Eilifsen, C. (2012). Effects of a meaningful, a discriminative, and a meaningless stimulus on equivalence class formation. *Journal of the Experimental Analysis of Behavior, 97*, 163–181. doi:10.1901/jeab.2012.97-163.

Fields, L., Arntzen, E., & Moksness, M. (2014). Stimulus sorting: A quick and sensitive index of equivalence class formation. *The Psychological Record, 64*, 487–498. doi:10.1007/s40732-014-0034-y.

Fienup, D. M., & Critchfield, T. S. (2011). Transportability of equivalence-based programmed instruction: Efficacy and efficiency in a college classroom. *Journal of Applied Behavior Analysis, 44*, 435–450. doi:10.1901/jaba.2011.44-435.

Fienup, D. M., Wright, N. A., & Fields, L. (2015). Optimizing equivalence based instruction: Effects of training protocols on equivalence class formation. *Journal of Applied Behavior Analysis*. (in press)

Grimm, L. J. (2011). *Emergence of equivalence relations: Comparing sorting and match-to-sample procedures.* Nortwestern University. Applied Behavioral Analysis Master's Theses. Paper 53. http://hdl.handle.net/2047/d20001017

Holth, P., & Arntzen, E. (1998). Stimulus familiarity and the delayed emergence of stimulus equivalence or consistent nonequivalence. *The Psychological Record, 48*, 81–110. Retrieved from http://thepsychologicalrecord.siuc.edu/index.html.

Kennedy, C. H. (1991). Equivalence class formation influenced by the number of nodes separating stimuli. *Behavioural Processes, 24*, 219–245.

Leader, G., Barnes, D., & Smeets, P. M. (1996). Establishing equivalence relations using a respondent-type training procedure. *The Psychological Record, 46*, 685–706.

Mackay, H. A., & Fields, L. (2009). Syntax, grammatical transformation, and productivity: A synthesis of stimulus sequences, equivalence classes, and contextual control. In R. A. Rehfeldt & Y. Barnes-Holmes (Eds.), *Derived relational responding. Applications for learners with autism and other developemental disabilities* (pp. 209–236). Oakland, CA: New Harbinger Publications, Inc.

McIlvane, W. J., & Dube, W. V. (1990). Do stimulus classes form before they are tested? *The Analysis of Verbal Behavior, 8*, 13–18.

Nartey, R. K., Arntzen, E., & Fields, L. (2014). Two discriminative functions of meaningful stimuli that enhance equivalence class formation. *The Psychological Record, 64*, 777–789. doi:10.1007/s40732-014-0072-5.

Nedelcu, R. I., Fields, L., & Arntzen, E. (2015). Conditional discriminative functions of meaningful stimuli and enhanced equivalence class formation. *Journal of the Experimental Analysis of Behavior, 103*, 349–360. doi:10.1002/jeab.141.

Pytte, C. L., & Fienup, D. M. (2012). Using equivalence-based instruction to increase efficiency in teaching neuroanatomy. *The Journal of Undergraduate Neuroscience Education, 10*, A125–A131.

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. doi:10.1901/jeab.1982.37-5.

Smeets, P. M., Dymond, S., & Barnes-Holmes, D. (2000). Instructions, stimulus equivalence, and stimulus sorting: Effects of sequential testing arrangements and a default option. *The Psychological Record, 50*, 339–354. Retrieved from http://thepsychologicalrecord.siuc.edu/index.html.

Travis, R. W., Fields, L., & Arntzen, E. (2014). Discriminative functions and over-training as class-enhancing determinants of meaningful stimuli. *Journal of Experimental Analysis of Behavior*. doi:10.1002/jeab.91.

Tyndall, I. T., Roche, B., & James, J. E. (2004). The relation between stimulus function and equivalence class formation. *Journal of the Experimental Analysis of Behavior, 81*, 257–266. doi:10.1901/jeab.2004.81-257.

Walker, B. D., & Rehfeldt, R. A. (2012). An evaluation of the stimulus equivalence paradigm to teach single-subject design to distance education students via blackboard. *Journal of Applied Behavior Analysis, 45*, 329–344. doi:10.1901/jaba.2012.45-329.