Journal of Applied Behavior Analysis

JOURNAL OF APPLIED BEHAVIOR ANALYSIS 2017, **9999,** n/a–n/a NUMBER 9999 ()

TRAINING CLASS INCLUSION RESPONDING IN TYPICALLY-DEVELOPING CHILDREN AND INDIVIDUALS WITH AUTISM

Siri Ming, Teresa Mulhern and Ian Stewart

NATIONAL UNIVERSITY OF IRELAND, GALWAY

Laura Moran

BROTHERS OF CHARITY, ROSCOMMON

AND

Kellie Bynum

CHRYSALIS ACADEMY, PHOENIX, AZ

In a *class inclusion* task, a child must respond to stimuli as being involved in two different though hierarchically related categories. This study used a Relational Frame Theory (RFT) paradigm to assess and train this ability in three typically developing preschoolers and three individuals with autism spectrum disorder, all of whom had failed class inclusion tests. For all subjects, relational training successfully established the target repertoire and subsequent testing demonstrated both maintenance and generalization. Limitations and future research directions are discussed.

Key words: class inclusion, hierarchy, nonarbitrary relational training, relational frame theory

---

Categorization, defined as “differential responding to classes of nonidentical, though potentially discriminable, stimuli” (Zentall, Galizio, & Critchfield, 2002, p. 238), has been described as fundamental to thinking and concept learning (e.g., Lakoff, 1987). For example, classification and sorting of sets and subsets is considered an important competency for older preschoolers and kindergarten-age students as a foundation for mathematical skills (California Department of Education, 2008). One feature of such advanced categorization repertoires is hierarchical categorization—responding to categories contained within other categories. One test of appropriate responding in this respect is the so-called *class inclusion* task commonly used in mainstream psychological tests. In this task, a child is shown an array of stimuli from a particular class that includes two different subclasses, with a greater quantity of stimuli from one of the two subclasses. They are then asked whether there are more members of the more populous of the two subclasses or more members of the entire class. For example, they might be shown pansies and violets with more pansies than violets present and asked, “Are there more pansies or are there more flowers?” This type of question probes responding to stimuli based on both class and subclass.

Previous researchers have tried to teach class inclusion responding to young (4- to 7-year-old) typically developing (TD) children (e.g., Judd &

Authorship order for the first and second authors was determined by coin toss. This study was conducted as part of both authors’ doctoral theses at the National University of Ireland, Galway. The first author conducted a portion of this study with support from experiment.com backers, with particular thanks to Julia Fiebig, Eric Fox/Foxylearning, Catherine Green, Michelle Kelly, Richard Laitinen, Denny Luan, Christina Lovaas, Mandy Mason, Josh Pritchard, Greg Stikeleather, Vicci Tucci and Regina Wielenska. The second author conducted a portion of this study with support from the Galway Doctoral Research Scholarship.

Address correspondence to: Siri Ming, School of Psychology, NUI Galway, Ireland. Email: siri@siriming.com

doi: 10.1002/jaba.429

© 2017 Society for the Experimental Analysis of Behavior

Mervis, 1979) using various combinations of quantification (i.e., counting the items in the sets), feedback, and reinforcement. Despite some success, there is a lack of data on generalization and maintenance and those data reported are less convincing than desired (e.g., McCabe & Siegel, 1987). Furthermore, no work has been conducted with children with developmental delay.

A behavior-analytic approach to conceptualizing and teaching class inclusion responding as a core aspect of hierarchical categorization might yield greater success in establishing a generalized repertoire in young children both with and without developmental delay. Such an approach was adopted in this study, informed by relational frame theory (RFT; Hayes, Barnes-Holmes, & Roche, 2001). RFT proposes that generalized relational responding is a key process in complex human behavior, including hierarchical categorization. For RFT, advanced categorizing requires an advanced relational framing repertoire including hierarchical relational framing. Hierarchical or categorical relations are rooted in simpler relations including containment (e.g., X contains Y) and comparison (X is larger than Y). Correct performance on categorization tasks (e.g., the class inclusion task) requires sufficient experience with these simpler relations and their combination in the context of categories (e.g., seeing that larger classes can contain smaller classes).

Strengthening the foundations of categorization ability in young children with ASD is likely to improve their everyday functioning as well as accelerating their acquisition of intellectual skills more generally. As suggested, learning class inclusion is a relatively important feature of this repertoire. Heretofore, however, no investigation or remediation of class inclusion had been undertaken in this population. RFT provides a relatively clear conceptualization of class inclusion in terms of relational responding and would suggest that multiple-exemplar training of the combination of containment and comparison relations in a nonarbitrary relational context can provide a useful means by which to establish and strengthen this repertoire. The aim of the current research was to demonstrate this RFT-based approach to assessing and training class inclusion in typically developing children and individuals (both children and adults) with autism spectrum disorders.

METHOD

Participants and Setting

Three typically-developing (TD) children (T1, T2, and T3) and three individuals diagnosed with autism spectrum disorder (ASD; A1, A2, and A3) participated. T1 (age 3 years, 6 months), T2 (age 4 years, 1 month), and T3 (age 3 years, 5 months) were enrolled in an Irish preschool. A1 (age 8 years, 1 month; Peabody Picture Vocabulary Test, Fourth Edition [PPVT-4; Dunn & Dunn, 2007] age-equivalency 7 years, 2 months) was enrolled in a specialized school for children with ASD in India. A2 (age 19 years; PPVT-4 age-equivalency 7 years, 11 months) and A3 (age 9 years, 7 months, PPVT-4 age-equivalency 6 years, 5 months) were enrolled in a specialized school for children with ASD in the United States. All participants had tact, listener, and intraverbal repertoires consistent with age (TD) or PPVT-4 scores (ASD); A1, A2, and A3 could read and write short sentences. Teachers, school behavior analysts, and the first and second authors conducted sessions two to three times weekly in a separate room in the participants' schools.

Experimental Design

A nonconcurrent multiple baseline design was used with participants preassigned to one of three baseline lengths (three, five, or seven sessions).[1]

[1]Watson and Workman (1981) and Christ (2007) recommend the use of randomly preassigned baseline durations to improve the internal validity of nonconcurrent multiple-baseline designs; a potential drawback of this method is that responding may not be stable on a given baseline prior to the predetermined introduction of an intervention.

Materials

Materials included colored flashcards and plastic containers. The flashcards were 5.5 cm x 5.5 cm pictures of items from four different categories (animals, fruit, clothing, and vehicles), with six exemplars per category (e.g., for animals: dog, cat, horse, cow, pig, and sheep). The set of clear plastic containers were labeled using whiteboard markers and included two smaller containers (for the exemplars) which fit into one larger container (for the category).

Two sets of flashcards were used to randomize presentation of trials in terms of (a) stimuli used and (b) class inclusion trial types. Stimulus cue cards contained all combinations of quantities 1-6 and exemplars for each of the categories used, for a total of 30 unique combinations per category (e.g., "four lemons, five oranges"); thus, each trial included different stimuli from previous trials. Eight variations of class inclusion questions were used to counterbalance questions between asking about the larger or smaller number of exemplars, asking "more" versus "less" questions, and varying word order to prevent undesirable stimulus control. For example, given four dogs and six cats, potential questions might include, "Are there more dogs or more animals?", "Are there less animals or less cats?", "Are there more animals or more cats?", etc. (see next section for all eight trial types). We interspersed a variety of mastered questions, including quantitative comparisons between subclass stimuli (e.g., "Are there more dogs or more cats?"). Trials were presented randomly, in a ratio of one mastered trial to one class inclusion trial.

Measurement

Accuracy of responding to class inclusion questions was measured for the first presentation of each class inclusion trial-type for a total out of eight, and then converted to a percentage. Trial types included: More [category] or more [smaller subclass], more [category] or more [larger subclass], less [category] or less [smaller subclass], less [category] or less [larger subclass], more [smaller subclass] or more [category], more [larger subclass] or more [category], less [smaller subclass] or less [category], and less [larger subclass] or less [category]. Each trial type was presented once in baseline, whereas trial types were presented multiple times in intervention based on need for corrective feedback.

Procedures

Screening. Participants were screened using tabletop discrete trial procedures to ensure they could: (a) tact all stimuli; (b) answer yes/no stimulus identification questions (e.g., "Is this a cat?"); (c) tact the category of all stimuli (e.g., "What category does this [picture of a cat] belong to?" "Animals"); (d) tact quantities of stimuli from 1-10; and (e) answer questions of quantitative comparison between stimulus sets (e.g., "Are there more cats or more dogs?"). Screening questions were used as interspersal questions during the intervention.

Baseline. At the start of each baseline session, each set of cue cards (trial type and stimulus selection) was first shuffled; the stimulus selection card deck included all four categories of stimuli (animals, fruit, clothing and vehicles) and thus categories were randomly paired with question variations. The participant was then asked to select a stimulus card, and the administrator laid out the two stimulus sets on the table as described by the stimulus card, such as one pile of picture cards consisting of three cats, and another consisting of five horses. The experimenter then selected a trial type card as described above and presented the relevant trial (class inclusion question or interspersal question). Once the participant responded, stimulus sets were removed, a new stimulus card and new trial type card were selected, and the next trial began.

During baseline, nonspecific praise was provided for all trials (e.g., "You're working really hard!"; "I like how you're paying attention!"), and reinforcement for participation was provided on the schedule identified by the participant's teacher as appropriate to a teaching session. No feedback or reinforcement contingent on correct responding was provided. Only responses to class inclusion questions were recorded. A session was terminated after all eight types of class inclusion questions had been asked. As class inclusion and interspersal questions were presented in a ratio of 1:1, there were 16 trials per session. Baseline sessions typically lasted approximately 10 min, and were conducted two to three times per week.

Intervention. Multiple exemplar training (full protocol available from first author) was provided using nested boxes to promote saliency of the relation of "containment" of the smaller category within the larger category. For each trial, a new stimulus set was used (e.g., three cows, five pigs) based on random selection from the stimulus flashcards. Only the animal stimulus set was used for intervention. Otherwise, trials were arranged as in baseline. Intervention included two phases. Phase 1 included a number of pretrial requirements to enhance the saliency of the boxes. These pretrial requirements were faded in Phase 2.

Phase 1 (pretrial prompting): The experimenter began by describing the larger box as being for the category and asking the participant to tact the category of the flashcard stimuli (animals). For each trial, the participant was instructed that the specific stimuli used for that trial (e.g., three cats, six horses) were all animals (e.g., "Cats and horses are both animals"), belonged to the animal category, and went inside the animal category box. The participant was then asked to place the flashcards in the two smaller boxes, place the smaller boxes inside the larger box, and select the box containing the stimulus type for the trial (e.g., "Show me the horse box") and the category box (i.e., "Show me the animal category box"). Errors were corrected by gesturing to the correct selection and allowing another opportunity to answer independently. Once both boxes had been selected correctly, the experimenter presented the trial question (e.g., "Are there less horses or less animals?") while lifting up each of the boxes.

Correct responses were followed by specific praise (e.g., "You got it, there are less horses than animals!"), while lifting up relevant boxes. Incorrect responses were followed by repeating the requirement to select the stimulus type and category boxes and corrective feedback detailing the relation between the items and the category while picking up relevant boxes. For example, "Horses and cats are types of animals, so they all go inside the big animal category box. They all belong to the animal category, but only these are horses, so there are less horses in the horse box than there are animals in the animal category box." The trial was then re-presented, and the same trial type was repeated on the following trial but with a new combination of animal in different quantities. This process continued until the participant responded correctly on the first trial with new stimuli. A new trial type was then selected.

Phase 2 (reduced prompting). During the next intervention phase, the pretrial requirement to select the relevant boxes was eliminated, and verbal feedback was reduced to eliminate explicit reference to the size of the boxes (i.e. "big" category box, "small" subclass box) or that the subclass boxes "go inside" the category box. For example, corrective feedback statements were reduced to stating, while picking up the boxes, that "they all belong to the animal category, but only these are [subclass stimulus type, e.g., horses], so there are more [or less] [subclass] in the [subclass] box than there are animals in the animal category box."

Intervention sessions (lasting 30-45 min) in each phase continued until the participant responded correctly on each of the eight class inclusion trials. Each intervention phase continued until participants responded correctly to the first trial presentation of each trial-type.

Postintervention probes. Once participants reached the criterion for the final intervention phase, generalization was assessed using the same procedures as in baseline, first for animals and then for all four category types interspersed. Maintenance was tested 4 to 8 weeks later.

Procedural Fidelity and Interobserver Agreement

Procedural fidelity checks and interobserver agreement (IOA) were determined for all session types, including baseline, intervention, generalization and maintenance sessions, by a trained research assistant. Procedural fidelity was assessed through the use of a fidelity checklist in which each trial presentation was scored as either correct or incorrect; correct presentation required adherence to all relevant procedural criteria based on trial type/phase of intervention, including instructional presentation and use of the appropriate feedback script. Interobserver agreement was calculated on a trial-by-trial basis for each class inclusion trial within the session. Procedural fidelity and interobserver agreement (IOA) were assessed during 100% of baseline, intervention, generalization, and maintenance sessions with TD participants and during 20% of sessions with participants with ASD. Procedural fidelity ranged from 87.5% to 100% (M = 98%). IOA ranged from 87.5% to 100% (M = 99%).

RESULTS AND DISCUSSION

Baseline performance for all participants was near chance levels (Figure 1). For T1, T2, and T3, baseline performance was similar on animal questions (targeted for intervention) and questions related to other categories (Table 1). Baseline data by category are not available for A1, A2, and A3, but post-hoc review of available videos (approximately 10% of all baseline sessions) indicated that correct and incorrect responses were distributed across all category types. All participants were then successfully trained in class inclusion responding and demonstrated generalization and maintenance across several categories.

All participants showed immediate improvements in performance once intervention began, and met criterion in three to seven sessions. T1, T2, T3, and A1 subsequently showed 100% correct responding during the postintervention probe for the trained and untrained categories and 100% maintenance at 1 month to 6 weeks. A2 responded with 100% accuracy during the initial postintervention probe, but performance on the trained category returned to baseline levels in the maintenance probe. After one additional Phase 2 intervention session, A2 was successful in postintervention probes immediately following training and in a second session two days later. In 2-week maintenance probes A2 made one error with the trained category and none with the untrained categories; a 6-week maintenance probe showed 100% accuracy for both trained and untrained categories. A3 made one error in the first postintervention probe with animals. Following an additional Phase 2 intervention session he showed 100% success on postintervention and maintenance probes.

Two limitations must be noted. First, ensuring baseline stability before intervention would have provided a stronger demonstration of experimental control. Second, the role of the individual components of the intervention remains to be investigated (e.g., use of the nested boxes). Nevertheless, this study is the first to implement training procedures for class inclusion responding with individuals with ASD, and to show generalization and maintenance with typically developing children.

This is also the first study to use an approach informed by RFT to teach class inclusion responding by drawing on relevant foundational nonarbitrary relational respondings. Categorization skills have primarily been addressed within ABA programs strictly from the

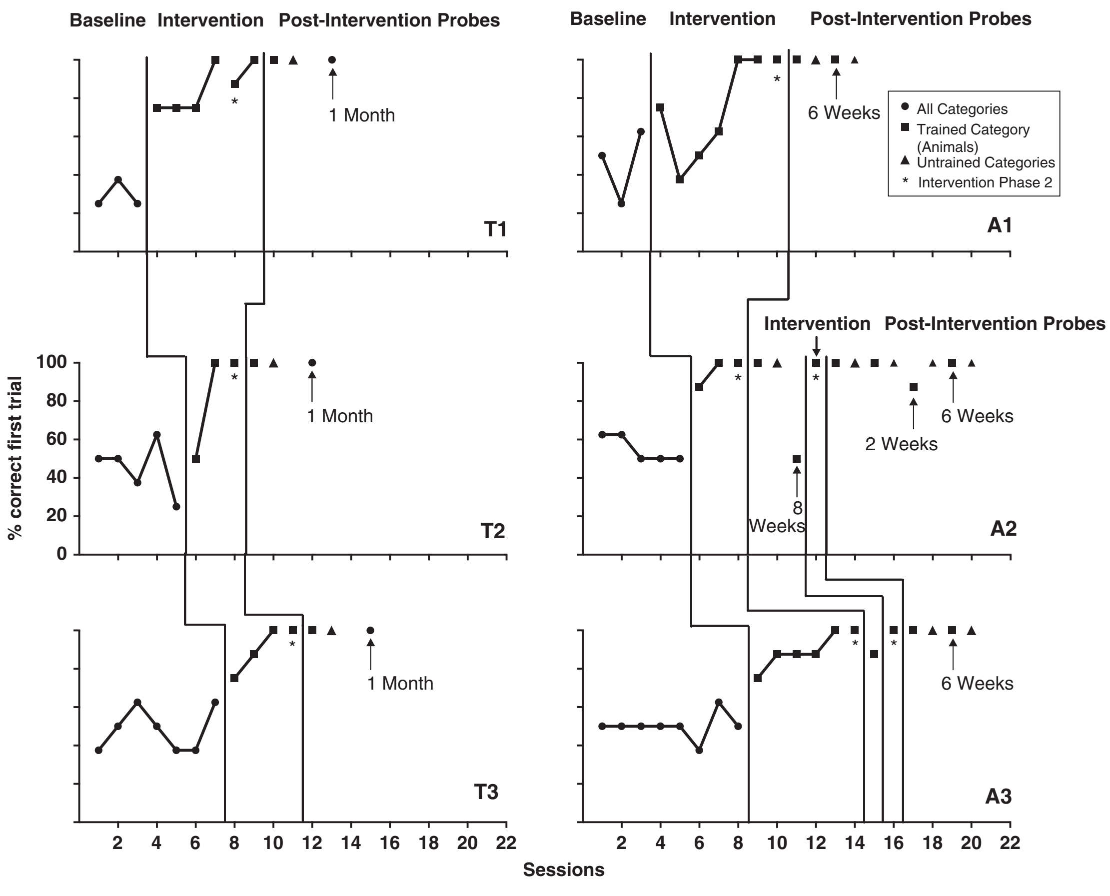

Figure 1. Results for class inclusion trials; chance level responding is 50%.

Table 1
Baseline Scores by Category for Typically Developing Participants

Participant	Trained (animals)	Untrained	All
T1	2/6 (33%)	5/18 (28%)	7/24 (29%)
T2	4/10 (40%)	14/30 (47%)	18/40 (45%)
T3	7/14 (50%)	20/42 (48%)	27/56 (48%)

perspective of associating names of stimuli with names of the categories to which the stimuli belong (e.g., Miguel, Petursdottir & Carr, 2005; Petursdottir, Carr, Lechago & Almason, 2008). However, it can be argued that if children cannot respond in accordance with the relations *between* categories and members, rooted in relevant nonarbitrary relations, then their repertoire of categorization is inadequate and teaching a category name might simply be like teaching a different name for the same object. In fact, the pattern of responding seen with several of the participants in these studies would seem to bear this out—responding during baseline seemed to reflected a pattern of stimulus control in which participants were responding simply on the basis of the quantity of each stimulus type, and were comparing the stimulus type asked about to the other stimulus type rather than the larger category, as if the category was irrelevant to the relation. Once intervention was begun, the participants' performances immediately began to reflect responding in accordance with the relation between the category and the particular stimulus type/subcategory referred to in the question, rather than in accordance with the relation between the two stimulus types/subcategories. Incorporating relevant nonarbitrary containment relations may provide an important level of support.

One way in which the current work might be extended is, as already suggested, by examining the role of the individual components of the intervention. This would be helpful in probing further the exact role played by the nonarbitrary relational dimension and how extensive this need be. Beyond this, it would be useful to gauge how class inclusion responding might interact with other repertoires. For example, might training in class inclusion skills have a noticeable influence on children's classification repertoire more generally (as assessed by such mainstream instruments as the Children's Category Test [CCT; Boll, 1993] or similar) or might broader, more extensive training in hierarchical categorization relations be needed? Similarly, it would be beneficial to determine if training in class inclusion might facilitate other repertoires of hierarchical responding. For example, Newsome, Berens, Ghezzi, Aninao, and Newsome (2014) provided children aged 9-12 with an intervention to strengthen their abilities to discriminate hierarchical relations on the basis of same/different relations (e.g., "How is rice different from/same as watermelon?"). This intervention targeted skills that Newsome et al. note as critical for reading comprehension, and indeed the participants' performance on reading comprehension measures improved, along with fluency and novelty of their responding to the tasks. It would be informative to examine whether and how training in nonarbitrary class inclusion tasks might facilitate the types of arbitrary hierarchical responding studied by Newsome et al., as well as what impact such training might have on other academic skills.

In summary, while this is preliminary work, it represents the first explicitly RFT-based study into classification in young, typically developing children and individuals with developmental delays. While there is much yet to be examined with respect to the full range of hierarchical relational responding, the results contribute to our understanding of early emergent relational responding repertoires, and are promising for future curriculum development for language intervention for children and adults with autism and other developmental delay.

REFERENCES

Boll, T. (1993). *Children's category test: Assessment tools for the neuropsychological assessment of children*. San Antonio, CA: The Psychological Corporation.

California Department of Education. (2008). *California preschool learning foundations: Vol. 1*. Sacramento, CA: CDE Press.

Christ, T. J. (2007). Experimental control and threats to internal validity of concurrent and nonconcurrent multiple baseline designs. *Psychology in the Schools*, *44*, 451-459. https://doi.org/10.1002/pits.20237

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test, Fourth Edition*. Minneapolis, MN: Pearson Assessments.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition*. New York, NY: Plenum Press.

Judd, S., & Mervis, C. (1979). Learning to solve class-inclusion problems: The roles of quantification and recognition of contradiction. *Child Development, 50*, 163-169. https://doi.org/10.2307/1129052

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago, IL: University of Chicago Press.

McCabe, A., & Siegel, L. (1987). The stability of training effects in young children's class inclusion reasoning. *Merrill-Palmer Quarterly, 33*, 187-194. Retrieved from http://www.jstor.org/dis-cover/10.2307/23086328?uid=3738232&uid=2&uid=4&sid=21102457621977

Miguel, C. F., Petursdottir, A. I., & Carr, J. E. (2005). The effects of multiple-tact and receptive-discrimination training on the acquisition of intraverbal behavior. *The Analysis of Verbal Behavior*, 27-41. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2774100/.

Newsome, K. B., Berens, K., Ghezzi, P. M., Aninao, T., & Newsome, W. D. (2014). Training relational language to improve reading comprehension. *European Journal of Behavior Analysis, 15,* 165-197.

Petursdottir, A. I., Carr, J. E., Lechago, S. A., & Almason, S. M. (2008). An evaluation of intraverbal training and listener training for teaching categorization skills. *Journal of Applied Behavior Analysis*, *41*(1), 53-68. https://doi.org/10.1901/jaba.2008.41-53

Watson, J., & Workman, E. A. (1981). The non-concurrent multiple baseline across-individuals design: An extension of the traditional multiple baseline design. *Journal of Behavior Therapy and Experimental Psychiatry, 12*, 257-259. https://doi.org/10.1016/0005-7916(81)90055-0

Zentall, T. R., Galizio, M., & Critchfied, T. S. (2002). Categorization, concept learning, and behavior analysis: An introduction. *Journal of the Experimental Analysis of Behavior, 78*, 237-248. https://doi.org/10.1901/jeab.2002.78-237

Received November 29, 2015
Final acceptance December 23, 2016
Action Editor, Anna Petursdottir