

The Psychological Record

Teaching Real World Categories using Touchscreen Equivalence Based Instruction.

--Manuscript Draft--

Manuscript Number:	PSRE-D-17-00041R2	
Full Title:	Teaching Real World Categories using Touchscreen Equivalence Based Instruction.	
Article Type:	Empirical article	
Funding Information:	Dublin City University	Not applicable
Abstract:	<p>Although researchers have evaluated the application of equivalence based instruction (EBI) to naturalistic teaching settings, few have examined individualized educational programs of application. The current study therefore employed an EBI procedure to teach individual categorization lessons to pre-school age children (n=5). A category sort test was used to identify stimuli for each child which were incorrectly categorized pre-intervention. These stimuli consisted of real world stimuli from non-overlapping categories (e.g., toys, clothing, and fruit). Participants were trained in six conditional discriminations tested for the emergence of three three-member classes (A1, B1, C1: A2, B2, C2; A3, B3, C3) using a computerized touch-screen matching to sample (MTS) procedure. Participants subsequently were trained to identify receptively the category label for the C stimulus in each class. Following training and testing, the category sort test was re-administered. All participants demonstrated categorization of the directly trained class members and further generalization to addition unknown stimuli. The results show that little training was required in the use of touch-screen responding indicating that the use of such devices may provide a simple means of computerized teaching in young populations.</p>	
Corresponding Author:	Ronda Barron Dublin City University IRELAND	
Corresponding Author Secondary Information:		
Corresponding Author's Institution:	Dublin City University	
Corresponding Author's Secondary Institution:		
First Author:	Ronda Barron	
First Author Secondary Information:		
Order of Authors:	Ronda Barron	
	Sinéad Smyth	
	Julian Leslie	
Order of Authors Secondary Information:		
Author Comments:		
Response to Reviewers:	<p>We would like to take this opportunity to thank the reviewers for their extremely helpful feedback on this manuscript.</p> <p>Response to reviewer feedback is attached in word document for ease of use.</p>	

Running head: Equivalence based instruction

Teaching Real World Categories using Touchscreen Equivalence Based Instruction.

Ronda Barron,
Dublin City University

Julian C. Leslie
University of Ulster

Sinéad Smyth,
Dublin City University
Ireland

Corresponding author. Tel.: +353 (0) 1 700 6030. E-mail addresses: ronda.barron2@mail.dcu.ie, rondabarron@gmail.com (R. Barron), jc.leslie@ulster.ac.uk (J. Leslie), sinead.smyth@dcu.ie (S. Smyth).

Conflict of Interest. All of the authors declare that they have no conflict of interest.

Informed consent. Informed consent was obtained from parents or guardians of the all individual participants, and the participants themselves assented to taking part in the study.

Acknowledgments. This research was conducted as part of the first author's doctoral studies. The School of Nursing and Human Sciences at Dublin City University provided funding of this project.

Abstract

Although researchers have evaluated the application of equivalence based instruction (EBI) to naturalistic teaching settings, few have examined individualized educational programs of application. The current study therefore employed an EBI procedure to teach individual categorization lessons to pre-school age children (n=5). A category sort test was used to identify stimuli for each child which were incorrectly categorized pre-intervention. These stimuli consisted of real world stimuli from non-overlapping categories (e.g., toys, clothing, and fruit). Participants were trained in six conditional discriminations tested for the emergence of three three-member classes (A1, B1, C1; A2, B2, C2; A3, B3, C3) using a computerized touch-screen matching to sample (MTS) procedure. Participants subsequently were trained to identify receptively the category label for the C stimulus in each class. Following training and testing, the category sort test was re-administered. All participants demonstrated categorization of the directly trained class members and further generalization to addition unknown stimuli. The results show that little training was required in the use of touch-screen responding indicating that the use of such devices may provide a simple means of computerized teaching in young populations.

Key words: touch-screen, categorization, children, equivalence based instruction, match-to-sample

[Click here to view linked References](#)

1
2
3 Stimulus equivalence has been extensively studied in the field of behavior analysis in recent decades (e.g., Devany,
4 Hayes, & Nelson, 1986; Dixon et al., 2016; Galizio, Stewart & Pilgrim, 2004), using a procedure whereby the training
5 of at least two conditional discriminations results in the derivation of others. Specifically, having been taught the
6 relation A1-B1 (i.e., given a stimulus arbitrarily designated as A1, select another stimulus, B1) and the relation B1-
7 C1, an individual may, without further instruction, derive reflexive (A1-A1, B1-B1, C1-C1), symmetrical (B1-A1,
8 C1-B1), transitive (A1-C1), and combined transitivity and symmetry/equivalence (C1-A1) relations between those
9 stimuli (Sidman & Tailby, 1982). The prevailing method of training and testing equivalence relations has been through
10 use of match-to-sample (MTS) tasks (see Groskreutz, Karsina, Miguel, & Groskreutz, 2010 for a description).
11
12
13
14
15
16
17
18
19

20 As many of these studies were laboratory simulations of real world scenarios, typically using arbitrary stimuli
21 to model how real world learning might take place, they may not be truly reflective of real-world learning (Pytte &
22 Fienup, 2012), and the application of stimulus equivalence to real-world teaching (referred to as equivalence based
23 instruction; EBI), has received less attention. EBI has, however, been used to teach adults in university settings include
24 algebra and trigonometry (Ninness, Dixon, Barnes-Holmes, Rehfeldt, Rumph, McCuller, & McGinty, 2009; Ninnes,
25 Barnes-Holmes, Rumph, McCuller, Ford, Payne, & Elliott, 2006), statistical interactions (Fields, Travis, Roy,
26 Yadlovker, Aguiar-Rocha, & Sturmey, 2009), statistical inferences (Fienup & Critchfield, 2010) and brain-behavior
27 relations (Fienup, Covey, & Critchfield, 2010).
28
29
30
31
32
33
34
35

36 Although these studies used real-world (rather than arbitrary) stimuli, they were still mainly laboratory-based
37 and so lacked ecological validity (Pytte & Fienup, 2012; Rehfeldt, 2011). Recognizing the shortage of naturalistic
38 research, Pytte and Fienup (2012) successfully taught classes of neuroanatomical associations using EBI with
39 university students using a standard lecture format. The results provided evidence of the effectiveness of EBI in a
40 natural educational environment and this suggested that EBI may be an effective educational tool. The application of
41 EBI in teaching educational goals in younger populations has produced an even smaller body of work. Research with
42 young children has included other approaches to categorization, such as naming theory (Horne & Lowe, 1996),
43 whereby typically developing children have demonstrated stimulus categorization after being taught the relevant
44 listener and speaker behaviors separately. In this study a table top based category sort test was used to assess emergent
45 conditional relations of geographical relations (LeBlanc, Miguel, Cummings, Goldsmith, & Carr, 2003; Miguel,
46 Petursdottir, Carr, & Michael, 2008). Both typically developing children (Miguel et al., 2008) or children with a
47 diagnosis of Autism Spectrum Disorder (ASD) (LeBlanc et al., 2003) have been studied, and the majority of studies
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 have used so-called table-top MTS procedures where the researcher sits across a table from the child and places cards
4
5 on the table for them to respond to. For example, Groskreutz et al. (2010) demonstrated equivalence class formation
6
7 via table top procedures with a group of children diagnosed with ASD. The skills were chosen from individualized
8
9 educational goals in the participants' individual educational plans.

10
11 EBI is readily implemented on computers in educational settings, but published evaluations of their efficacy
12
13 are limited. In one study, Haegele, McComas, Dixon, Mark, & Burns (2011) adapted computer-based MTS training
14
15 often used in interventions with children with ASD with typically developing pre-school aged children. One group of
16
17 six participants was taught A-B relations where the A stimuli were the spoken words "eleven", "twelve" and "thirteen",
18
19 and the B stimuli were the written numerals, 11, 12 and 13, and was also taught A-C relations where C stimuli were
20
21 the written words for 11, 12 and 13 in one endangered Native American Language (Ojibwa or Dakota). For another
22
23 group of six participants the training was the same except that the C stimuli were the written words for 11, 12 and 13
24
25 in another endangered Native American Language. Training on A-B and A-C relations continued until a mastery
26
27 criterion was reached. Both prior to the training and afterwards, the children were tested on all relations (A-B, B-A,
28
29 A-C, C-A, B-C, C-B) for the three classes that had been trained. While a comparison group (which had been receiving
30
31 relevant classroom instruction) remained around chance level, the children who had received MTS training improved
32
33 from 41% to 94% accuracy averaged across relations after the training. This suggests that EBI can outperform the
34
35 effects of conventional classroom instruction in achievement of mainstream educational goals. Furthermore, EBI leads
36
37 to the acquisition of additional untrained relations by design, while conventional classroom instruction through
38
39 exemplars and rules does not necessarily have this generative outcome.

40
41 While laboratory-based simulation studies use arbitrary stimuli, often nonsense words or shapes, to avoid
42
43 effects of pre-experimental learning history (Fields & Moss, 2008), this is not appropriate for EBI where participants
44
45 are being taught actual relations between objects, concepts, or events that they may encounter in the real world. In
46
47 complex natural categories and indeed many lexical classes, class members often do not share physical characteristics,
48
49 however, this is not always the case and class members frequently do share perceptual characteristics across multiple
50
51 exemplars (e.g., all snakes have a forked tongue, limbless body, cylindrical shape and scaly skin) (Fields, & Moss,
52
53 2008; Fields, Reeve, Adams & Verhave, 1991; Fields, Verhave, & Fath, 1984). Stimuli from within and between
54
55 classes often have types of variations such as color, size and brightness therefore categorization can be conceived in
56
57 the same terms as stimulus discrimination and generalisation (Zentall et al., 2002). Rosch and Mervis (1975) described
58
59
60

1
2
3 categories as having family memberships, the category ‘games’ may include a board game, or chasing in the park.
4
5 Therefore, the contextual and culturally specificity of categories must be accounted for, when someone is asked ‘what
6
7 do you sit on?’ responses may differ from a chair, to a sofa or to a cushion. Rosch, Mervis, Gray, Johnson, and Boyes-
8
9 Braem. (1976) argued that family memberships had implications for the understanding of hierarchies in concepts.
10
11 Specifically, this argument centred on the correlation structure of features, that features tended to be shared across
12
13 several instances of a category member. This structure created what they referred to as ‘chunks’ or ‘clusters’ that
14
15 correspond to basic categories (Rosch et al., 1976). To illustrate, within the animal kingdom having feathers is
16
17 associated with nesting in a tree but having gills is associated with living under water. Typicality of a class
18
19 member is based upon its closeness to the prototype, which describes the category best, but this typicality decreases
20
21 with distance from the category prototype.
22

23
24 Categories at basic level (e.g., dog) can be combined to form super-ordinate level categories (e.g., animals),
25
26 or can be divided to form subordinate level categories (e.g., Alsatian). Perceptual features or commonality of shape
27
28 are commonly expected at the basic level and therefore the errors seen in natural categorization with children often
29
30 can be attributed to reliance on perceptual commonality (Quinn & Eimas, 1997). When natural categories are used,
31
32 children may indeed sort stimuli on the basis of other levels such as at super-ordinate level (animals). Furthermore,
33
34 non-similarity-based classes can often be described as arbitrary as they may be related by other functions than physical
35
36 similarities. Zentall et al. (1996) discussed how even across a broad range of stimuli minimal change is found in
37
38 response in similarity based classes. However, at the boundary of the category where similarity becomes more difficult
39
40 to determine, an abrupt change in response strength is demonstrated. Previous studies have attempted to control for
41
42 this through the use of pre-experimental testing. For example, Fienup et al. (2010) used a criterion of below 70%
43
44 responding at pre-test for inclusion of stimuli in their study, while Haegel et al. (2011) additionally used pre- and post-
45
46 testing as a measure of pre-experimental knowledge and to measure acquisition following the experiment.
47

48
49 The method of responding is also important in an EBI protocol. Difficulties have been reported regarding the
50
51 use of click-and-point devices, such as a mouse, in young children and the time required to train use of the device
52
53 (Hourcade, Perry & Sharma, 2008; Shimizu, Yoon & McDonough, 2010). One type of response which has received
54
55 little mention within the literature is interactive touch-screen technology which enables the child to reach and touch a
56
57 stimulus directly, rather than having to respond elsewhere (cf. Arntzen & Holth, 1997; Saunders, Drake, & Spradlin,
58
59 1999, who used an early type of fixed touch-screen; see also Nason & Zabrocky, 1988, for an even earlier attempt to
60

1
2
3 use this type of technology). The widespread availability of this type of technology now means that devices are easily
4 transportable and no longer require a fixed location. This expands research and learning opportunities via computer-
5 based instruction to a much younger population and additionally to individuals with special educational needs, such
6 as those with physical or intellectual disability (Still, Rehfeldt, Whelan, May & Dymond, 2014). It is important that
7 behavior analytic teaching methods assess the utility of technologies which are now in widespread use in educational
8 settings. An investigation of touchscreen technology with EBI is currently absent from the literature.
9
10
11
12
13
14

15
16 Rehfeldt (2011) called for ecologically valid research to investigate the application of EBI to naturalistic
17 educational settings. The purpose of the current study was to take some steps towards this by employing a touch-
18 screen computerized MTS program to teach functional real world categorization skills to young neurotypical children.
19 Each child was pre-screened and then taught individualized categories which were previously unknown to them but
20 appropriate to their developmental level, thus focusing on existing deficits and making training relevant for the learner.
21 To ensure generalization from the touchscreen to familiar classroom settings and activities, category sorting was
22 assessed in a table top task prior to any computer-based training. Furthermore, generalization to other, untrained,
23 category members was measured after training and testing with a table top task.
24
25
26
27
28
29
30
31

32 33 **Method**

34 35 **Participants**

36
37 Five typically developing preschool-age children (3 boys aged 4 years 3 months, 4 years 6 months &
38 4 years 2 months; 2 girls aged 4 years 2 months & 3 years 9 months) who attended a day care service took part.
39 Inclusion criteria were good receptive language, no major visual or motor problems, and no pre-existing knowledge
40 of the categories to be trained. All participants spoke English as their first language, and Participant 3 was bilingual,
41 speaking both English and Polish in the home. The standardized and norm referenced assessment of language, the
42 Preschool Language Scale – Fourth Edition (PLS – 4; Zimmerman, Steiner, & Pond, 2005), was used to assess baseline
43 levels of language ability. These assessments showed that the children had language skills closely matching norms
44 for their chronological ages. The ages and PLS-4 age equivalents (each in years: months) were as follows for
45 Participants 1 to 5 respectively: 4:3 and 4:2, 4:6 and 4:3; 4:2 and 3:10, 4:8 and 4:0, 4:2 and 3:10. Ethical approval for
46 this study was granted by Dublin City University Research Ethics Committee, and consent was obtained from parents
47 or guardians of the children who took part. Each child also gave assent to take part in the study prior to commencement.
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3 **Setting and materials**
4

5 All participants were trained and tested individually. All sessions took place in a small quiet area located in the
6 corner of the preschool classroom. Experimental sessions took place twice weekly and session duration was set at a
7 maximum of 30 min inclusive of breaks for each participant. A picture card (5 x 5cm, showing a computer with the
8 word “work” written underneath) representing the experiment was added to the participants’ daily visual schedules
9 located in the classroom, as was a yellow card (8 x 5 cm) with the word “break” on it in black. Tokens of yellow stars
10 (3 cm dimension), a token board with eight grids, and a standard kitchen countdown timer were also used. A visual
11 instruction boards (described further in the MTS pre-training section) contained written instructions for the
12 computerized experiment accompanied by corresponding pictures.
13
14
15
16
17
18
19
20

21 Prior to any other testing or training, a two-choice preference assessment (Fisher et al., 1992) was conducted
22 using a variety of developmentally appropriate toys. Pictures of the toys were affixed to a board later presented as
23 choices for participants following an experimental training block. Access to highly preferred toys was contingent upon
24 performance during training, with each token earning 30 s of play following completion of a block, and, in order to
25 maintain motivation, less preferred toys were presented for 60 s if no tokens had been earned in that block.
26
27
28
29
30

31 All picture stimuli used for either category sorting, training or computerized phases of the experiment were
32 obtained from Picture This 3.0 Professional Edition (CD©; Silver Lining Multimedia, Inc, 1999). All stimuli were
33 colored images presented on a white background in printed form as 7 x 10 cm laminated pictures. The same images
34 were used for the computer-based training phases in the form of bitmaps. Different stimuli were employed for each
35 participant depending on the outcome of the pre-experimental category sort tests. A 22.6 cm Asus Eee PC T91
36 Touchscreen Netbook ® with Windows XP operating system was used in the same location for the training phases.
37 The matching to sample programme was written using Microsoft Visual Basic 6.0 by S Smyth. This programme
38 recognizes a touch to the screen in the same way as a mouse click and so the input was simply coded as a mouse click.
39
40
41
42
43
44
45
46
47
48
49

50 **Procedure**
51

52 The experiment consisted of one pre-experimental category sort test, eight training stages and two post-
53 experimental tests.
54
55

56 *Pre experimental Category Sort Test.* The purpose of this test was to identify, for each participant, three
57 developmentally appropriate categories of which they had little or no knowledge. Participants were required to sort
58
59
60

1
2
3 27 picture cards once into three corresponding categories (9 cards per category). A fourth container was presented
4 and given a category label although no stimuli from this category were included. During the category sort task, the
5 participant sat at a table upon which four containers were placed. The experimenter sat out of view behind to the right
6 or left side of the participant. Participants were given the following instructions, “I want you to sort these pictures into
7 these containers.” The experimenter then pointed to each container and named what category was to be placed in each
8 container. For example, “Animals” (pointing at Container 1), “Fruit” (pointing at Container 2), “Transport” (Pointing
9 at Container 3), and “Toys” (Pointing at Container 4). The labels were repeated a second time again pointing to each
10 container. After instruction, the experimenter shuffled the picture cards. Each picture card was handed to the
11 participant individually while he/she was simultaneously asked to ‘sort’. Correct responding was defined as placing
12 the picture card in the corresponding category container, and self –corrected errors were accepted as correct. Incorrect
13 responses were placing the picture card in a non-corresponding container, at any other location on the table/floor, or
14 making no attempt to place the card within 10 s. No feedback was given at any stage during testing and no corrective
15 actions were undertaken by the experimenter. If the participant did not respond within 10 s the experimenter removed
16 the picture card and immediately placed the next picture in the participant’s hand, issuing the instruction ‘sort’.

17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32 Categories for the computerized stages of the study were chosen based on the results of the pre-experimental
33 category sort test. Category exclusion occurred when a participant placed more than three of the nine pictures from
34 the same category set into the same container, regardless of the container’s category label (ie even if they
35 miscategorised 3 or more pictures together). This conservative strategy was employed to minimize the possibility of
36 prior knowledge of a to-be-taught category. Participants were tested across several categories until three sets had been
37 identified as meeting the criterion outlined. Those categories identified were finally tested together to ensure that they
38 still met criteria and ensure minimal overlap in topography of images. The inclusion criterion for the exact stimuli to
39 be trained and tested using the computer-based programme was that none of the three chosen should have been sorted
40 in the same container (See Table 1 for the categories and stimuli chosen for each participant).

41
42
43
44
45
46
47
48
49
50 ***MTS Pre-training.*** The purpose of this stage was to familiarize the participants with the touch-screen.
51 Participants were directly trained to match three identical shapes, presented in a quasi-random order on the tablet PC.
52 The participant was informed that he/she would be asked to match pictures. The experimenter directed the participant
53 to the tablet PC and explained verbally and with the use of the visual instruction board. The spoken instructions were
54 as follows:
55
56
57
58
59
60

- 1
- 2
- 3
- 4 1. *It's time for our computer work.*
- 5
- 6 2. *It's time to match pictures on the computer.*
- 7
- 8
- 9 3. *When you match pictures, you can earn tokens.*
- 10
- 11
- 12 4. *We can swap our tokens for some toys or [named preferred item/activity].*
- 13
- 14
- 15 5. *If you need to you can ask for a break.*
- 16
- 17

18 The experimenter pointed to each associated item on the visual instruction board to show the participant while
19 reading the instructions. After reading the instructions was completed, the experimenter demonstrated for one trial
20 how to select the correct comparison while the tablet PC was placed on the table. During each trial, a black shape
21 (triangle, square or circle) appeared in the top center of the screen. The participant was required to touch this shape.
22 It immediately disappeared from the screen and an array of three comparison shapes (triangle, square and circle)
23 appeared in the lower area of the screen, one to the right, one to the center and one to the left. The location of the
24 correct comparison stimuli was randomized across trials. The experimenter issued the instruction 'Match' at the
25 beginning of each trial (appearance of the sample) and if no response was made after 5 s, the instruction was reissued.
26 No participant failed to respond following the second instruction. A correct response was recorded by the computer
27 if the participant touched the screen and selected the identical picture from the lower portion of screen (e.g., if the
28 sample was circle, select circle from the array of three comparison shapes). Correct responding was followed by the
29 appearance of a green symbol on the screen. An incorrect response was counted if the participant selected a non-
30 corresponding picture from the lower portion of screen. If an incorrect response was made, a red symbol appeared on
31 the screen. The correct and incorrect symbols disappeared after 3 s and the next sample appeared on the screen
32 immediately. As well as the visual feedback following a correct response, the experimenter delivered a token and
33 verbal feedback (e.g., 'good matching', 'nice matching', or 'super matching'). No other feedback was given for
34 incorrect responses. Each MTS pre-training block consisted of 12 trials during which each of the three shapes was
35 presented as a sample four times. At the end of the block the word "finished" appeared on the screen. The
36 experimenter pointed to the screen and stated, 'matching is finished'. A minimum break of 5 min occurred between
37 each block of trials, this was inclusive of time with access to toys earned. Participants were required to make 11 correct
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3 responses in a 12-trial block in order to proceed to the first experimental phase. Any participant who did not reach
4 criterion was re-exposed to the MTS pre-training. Participants 1, 2, 3 and 5 required only one exposure to MTS pre-
5 training, and Participant 4 met criterion after two exposures.
6
7

8
9 **Phase 1. A-B baseline training.** For their individual stimuli see Table 2. Participants were directly trained to
10 match the A and B stimuli (A1-B1, A2-B2, and A3-B3) using an MTS procedure as previously described. For
11 example, choosing B1 from an array (B1, B2, and B3) was reinforced following the presentation of A1. Correct
12 responding on trials was reinforced by the experimenter delivering a token onto the token board. Incorrect responding
13 was followed by the presentation of a red symbol on the screen, as described in the MTS pre-training. Twelve stimulus
14 trials were presented per block with each of the three trial types present four times in a quasi-random order.
15 Participants were required to reach a criterion of 11 correct responses over a 12-trial block to move forward to Phase
16 2. Participants always met criterion after four or fewer training blocks in each phase.
17
18

19 **Phase 2. B-A testing.** Testing for derived symmetry (B1-A1, B2-A2, B3-A3) followed. Criterion performance
20 was as in Phase 1. If the participant did not meet criterion, Phase 1 training was repeated. Instructions for testing phases
21 differed from training because no reinforcement was provided. The participant was again informed that he/she was to
22 match pictures. The experimenter directed the student to the tablet PC and read out the following instructions:
23
24

- 25 1. *It's time for our computer work.*
- 26 2. *It's time to match pictures on the computer.*
- 27 3. *You will not earn tokens this time for matching.*
- 28 4. *When we are finished matching we will go back to the classroom.*

29
30 The experimenter pointed to each associated item on the visual instruction board to show the participant while
31 reading the instructions aloud. The experimenter did not give feedback to the participant at any stage of testing. If the
32 participant did not respond within 5 sec, the experimenter stopped the testing and returned to the prior training phase.
33
34 When the participant met criterion, he/she proceeded to Phase 3.
35
36

37 **Phase 3. B-C training.** The procedure was the same as in Phase 1 with the exception that participants were
38 trained to match B1, B2 and B3 to C1, C2 and C3, respectively. When the participant met criterion, he/she proceeded
39 to Phase 4.
40
41
42
43
44

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Phase 4. C-B testing. Tests for derived symmetry were the same as in Phase 2 with the exception that testing was for C1-B1, C2-B2 and C3-B3 relations. When the participant met criterion, he/she proceeded to Phase 5.

Phase 5. A-B, B-C Mixed training. This combined training Phases 1 and 3. Participants were required to match all six A-B and B-C relations (A1-B1, A2-B2, A3-B3, B1-C1, B2-C2 and B3-C3). A 12-trial training block was employed, in which each of the six trial types were present twice in a quasi-random order. If the participant did not meet criterion, B-C training was repeated prior to a return to this phase. When the participant met criterion, he/she proceeded to Phase 6.

Phase 6. Mixed B-A, C-B testing. The testing procedure combined Phases 2 and 4. Participants were tested for all derived symmetry relations (B1-A1, B2-A2, B3-A3, C1-B1, C2-B2, and C3-B3). There were 12 trials in each test block and each of the six trial types was presented twice in a quasi-random order. The same criterion was used as in Phase 2. If Participants did not meet this requirement they were re-exposed to Phase 5 mixed training. When the participant did meet criterion, he/she proceeded to Phase 7.

Phase 7. A-C and C-A transitivity and combined symmetry and transitivity (equivalence) tests. The procedure for Phase 7 was the same as in Phase 6 with the exception that participants were tested on the previously untested A-C and C-A relations (A1-C1, A2-C2, A3-C3, C1-A1, C2-A2, C3-A3). The test block exposure consisted of 12 trials in which each trial type was presented twice in a quasi-random order. When the participant met criterion, he/she proceeded to Phase 8.

Phase 8. Category training. Table-top procedures were used to train appropriate category labels to each of the C stimuli. For example, if C1 was *apple* then the category label trained was *fruit*. Using verbal instructions, the participant was informed that he/she would be requested to choose pictures of objects when the experimenter named the category. The experimenter directed the participant to the three pictures representing the C stimuli, e.g., C1 (apple), C2 (ball), and C3 (car), and explained, 'I will give you a name, such as fruit, and you will give me the one that is the same'. The experimenter then placed the three C stimuli on the table in front of the participant and stated the category label (e.g., "fruit"). The participant was required to respond by choosing the picture that was related to this category, C1 (apple) in the example given. No feedback was given to the participants for incorrect responses. Correct responses were followed by the experimenter saying, 'that was giving me [category label]', and delivering a token as used in Phases 1, 3 and 5.

1
2
3
4 Initially, participants were not expected to have category knowledge, and a most-to-least prompt regime was
5 put in place to facilitate learning (Libby, Weiss, Bancroft & Ahearn, 2008). The most-to-least prompting hierarchy
6 (MTL) included four prompting levels: full physical - hand-overhand, light physical - light touch or shadow by the
7 elbow, gestural – instructor pointing at correct card and independent responding - no prompt. For each trial, the
8 experimenter noted the level of prompt required. Each block consisted of 12 trials during which each of the three C
9 stimuli were presented four times in a quasi-random order. Prompting was faded across blocks and the criterion for
10 reducing the level of prompt was 11/12 correct trials (92%) over one block. A correct response to a category training
11 trial was recorded if the participant selected the corresponding picture at the designated prompt level, or unprompted
12 within 5 s of the instruction.
13
14
15
16
17
18
19
20

21
22 Corrective action was taken within trial when an incorrect response was made. Incorrect selections were
23 blocked and the experimenter stepped back a level of intrusiveness in the prompt hierarchy while reissuing the
24 instruction. An incorrect response was recorded if a) a participant selected an incorrect picture card and the
25 experimenter had prompted at a level higher than designated, or b) if the participant made no attempt to select a card
26 within 5 s of the instruction. If a participant achieved less than 6/12 correct trials over three blocks, the experimenter
27 moved back a step in the prompt hierarchy intrusiveness. This phase of the study ended when the participant met the
28 criteria of 11 out of 12 trials correct independently – no prompt, and participants moved to final phase of the study,
29 post-experimental category sort test.
30
31
32
33
34
35
36
37

38 ***Post-experimental category sort test.*** A category sort test was used to test for the generalized derived transfer
39 of the categories trained in Phase 8. This phase involved the participants sorting the stimuli that had been tested pre-
40 experimentally. It included the three stimuli chosen for training during the experiment and an additional six from each
41 of the three categories 27 stimuli in total, to one of the three categories established for the C stimuli or to a fourth
42 irrelevant category. This exercise therefore tested for the generalization to untrained and untested stimuli. The
43 procedure and the instructions used were the same as in the pre-experimental category sort test, each of the stimuli
44 were presented once.
45
46
47
48
49
50
51

52 **Results**

53 **Pre-Experimental Category Sort Testing**

54
55
56
57
58
59
60
61
62
63
64
65

1
2
3 The results of the pre-experimental category sort test can be seen in Figure 1. A total of 27 stimuli were tested across
4 three categories, nine stimuli per category for each participant. The greatest number of correctly sorted stimuli for any
5 one category across participants was two stimuli, with Participant 2 correctly sorting two stimuli for each of the three
6 tested categories.
7
8
9

10 **MTS Pre-training**

11
12 Participant performances in all MTS phases of the experiment can be seen in Figure 2. Participants 1, 2 and
13
14 3 only required one exposure to the MTS pre-training before meeting the criterion to move on to the experimental
15 sessions. Participants 4 and 5 required two exposures.
16
17
18

19 **MTS Training and testing**

20
21 Participant 1 (Top left panel of Figure 2) required ten blocks of A-B training to meet the criterion of correct
22 responding on 11 out of 12 trials in a block. This participant failed the subsequent B-A test for the emergence of
23 symmetry relations. Following a second exposure to training (3 blocks), Participant 1 passed the B-A test. The
24 participant later met the criterion for B-C training after exposure to six blocks and subsequently passed the C-B test
25 for the emergence of symmetry relations at the first exposure. Only one exposure to A-B, B-C mixed training and
26 testing was required; following eight blocks the participant met criterion on the Mixed Test. The participant went on
27 to pass the A-C and C-A equivalence test. The other participants showed a similar pattern (see Figure 2) except that
28 none of them failed the first B-A test. For these four participants, A-B training took 3 to 8 blocks, B-C training took
29 3 to 6 blocks, and mixed training took 1 to 5 blocks, and, as with Participant 1, all met criterion on the first blocks of
30 transitivity and equivalence testing.
31
32
33
34
35
36
37
38
39
40
41
42

43 **Category Training**

44
45 Table 2 shows that each participant proceeded from full physical prompt to light physical prompt after one
46 block of trials, and then to gestural prompts after one further block. Participants required the most training blocks at
47 the gestural prompt stage. Four of the participants (Participants 1-4) only required one trial block in the final stage
48 where independent responses had to be made.
49
50
51
52

53 **Post-Experimental Category Sort Testing**

54
55 The results of the post-experimental sort test for all participants can be seen in Figure 3. Nine stimuli were
56 used for each of the three categories (27 stimuli in total). These consisted of the three stimuli in each category that
57
58
59
60

1
2
3 had been trained and tested in the MTS procedure, and an additional six category members that not been used in the
4 intervention but that had been tested in the pre-experimental category sort test. Overall, categorization scores were
5 very high with Participants 1 to 5 scoring 24, 26, 26, 26 and 27, respectively out of a possible 27, with all participants
6 correctly sorting all nine of the stimuli used in the MTS training.
7
8
9

10
11 The three categories chosen for Participant 1 were Insects, Fruit and Electrical (see Table 1 for actual stimuli),
12 and the stimuli used in MTS training were all correct sorted indicating a transfer of the category function from the
13 three C stimuli to their respective B and A stimuli. Three stimuli in total were not sorted correctly, two in the insect
14 category and one in the electrical category. The three categories chosen for Participant 2 were Accessories, Utensils,
15 and Fruit (See Table 1). AS well as correctly sorting the nine stimuli used in the MTS training phase, Participant 2
16 also correctly sorted 17 out of 18 additional stimuli. The only stimulus miscategorized at post testing was “‘scrunchie’
17 (fabric hair tie) which was miscategorized into the same category set fruit. The three categories chosen for Participant
18 3 were Dairy, Vegetables, and Furniture. As for the other participants, all stimuli used during training were correctly
19 sorted however one of the untrained stimuli, ‘margarine’ was mis-categorized into category set, furniture. For
20 Participant 4, the three categories chosen were sweets (candy), vegetables, and furniture. Only one stimulus from the
21 vegetable category ‘sweet corn’ was miscategorized at post-testing (into category set furniture). The three categories
22 tested for Participant 5 were Body parts, Musical instruments, and Furniture. Participant 5 sorted all stimuli correctly
23 at post testing.
24
25
26
27
28
29
30
31
32
33
34
35
36
37

38
39 As can be seen from Figure 3 all participants demonstrated categorization for the individualized real world
40 categories that had been pre-experimentally identified as unknown to each individual. During match-to-sample pre-
41 training (Figure 2) the participants acquired the skills to operate the touch-screen device with relative ease and only
42 two participants required a second block of training. All participants passed the computerized training and testing
43 phases, the largest number of training trials to criterion for all training and test phases required were for Participant 1
44 (396 trials) the shortest was for Participant 2 (156 trials). Participant 3 and 4 required 204 trials, and Participant 5
45 required 300 trials.
46
47
48
49
50
51
52

53 The category training which followed trained the corresponding category labels (e.g., animals) to each of the
54 three C stimuli (e.g., C1 dog). All participant passed this training phase, the highest trials to criterion being required
55 by Participant 5 (102 trials) and the lowest by Participant 4 (60 trials). Upon returning to the category sort test which
56
57
58
59
60
61
62
63
64
65

1
2
3 again tested the nine pre-experimentally tested stimuli a transfer of category function from the C stimuli to the A and
4
5 B stimulus was demonstrated. This was the case for all three equivalence classes for each participant. At post testing
6
7 additional untrained stimuli were also correctly sorted.
8
9

10 **Discussion**

11 The main purpose of the current study was to employ an EBI protocol in an ecologically valid manner.
12
13 Although previous research has indicated that EBI can be highly effective, this has typically been in laboratory
14
15 settings. It is important to increase our knowledge of EBI in real life settings and with functionally important stimuli.
16
17 The current study achieved this, thereby adding to the existing literature by tailoring category training to each
18
19 individual participant rather than teaching generic skills across participants, and, as well as using computer screen
20
21 stimulus presentation with touch screen responding, tested generalization to other media by using table-top
22
23 presentation of cards. The current methodology was successful in teaching previously unknown, real world categories
24
25 that were age appropriate and relevant to participants. In this experiment, all participants successfully passed all
26
27 training and testing phases without any need for remedial action. Furthermore, at post-testing all participants
28
29 accurately categorized the stimuli that were pre-experimentally known, the stimuli that were trained during match-to-
30
31 sample phases and all participants successfully matched at least 15 of 18 additional untrained stimuli (Participant 1
32
33 matched 15 stimuli, while Participant 5 matched all 18 and the other three participants matching 17 of the 18). Overall,
34
35 the results were encouraging because the procedures were not only effective but also efficient in terms of the total
36
37 time or number of trials taken to complete the training phases. Future research should further explore the efficiency
38
39 of EBI by comparing it with traditional teaching strategies such as verbal instruction plus trial-and-error learning. This
40
41 would help determine if that type of training could generate the extent of generalization to other category members
42
43 seen here.
44
45

46 The children who participated in the current study may have had a pre-experimental history of encountering
47
48 many of the experimental stimuli in their day-to day-life. Thus, the use of the pre-experimental category testing to
49
50 identify previously unknown categories for each participant was an important experimental control. Strict exclusion
51
52 criteria were employed in the pre-test. Due to the type of stimuli and the age of the participants this was of particular
53
54 importance as the children may have sorted the stimuli correctly together as a category but under an incorrect category
55
56 name. Furthermore, the use of an additional null container given a dummy name helped to ensure that that when
57
58
59
60

1
2
3 stimuli were sorted correctly this was based on accurate categorization. The null container allowed a stringent
4 exclusionary criterion to be put in place and reduced the probability of category exclusion due to chance. In the
5 experiment, knowledge acquisition demonstrated during match-to-sample and category training phases showed that
6 the stimuli chosen for each participant were indeed not associated pre-experimentally. Specifically, no participant
7 passed Phase 1 the first time and all demonstrated clear progression following each training phase.
8
9

10
11
12
13 The data from the A-B, B-C mixed training and testing phases show that only Participants 2 and 4 passed
14 this phase on the first attempt. Even though the participants were being trained and tested on previously trained
15 relations A-B and B-C, all other participants required additional training blocks. This failure to pass these phases on
16 the first attempt does provide some evidence that performance must be established using more stringent mastery
17 criteria before tests for equivalence take place (Arnzten & Holth., 1997; Devany et al., 1986).
18
19
20
21
22

23
24 Although the pre-experimental category sort test appeared to be successful in identifying categories unknown
25 to the participants, more than three sorted correctly excluded a category set. However, it also meant that two stimuli
26 from a given category could be paired within the same container – either the correct category container or an incorrect
27 one. If this was the case, then only one of those stimuli was chosen for inclusion in the trained conditional
28 discriminations. In the post-experimental sort test (administered after MTS training and testing and category function
29 training for the C stimuli) all participants correctly sorted all trained and tested stimuli. This indicated a transfer of
30 the category function from the three C stimuli to the B and A stimuli for each of the three categories. In addition, the
31 current study also examined generalization for up to six additional (untrained) members for each category, dependent
32 on how many stimuli each participant correctly sorted pre-experimentally. While transfer of function was seen for all
33 five participants, generalization to additional stimuli, that were not previously sorted correctly at pretesting or trained
34 during MTS, did not consistently occur. Participant 2, 3 and 4 miscategorized one stimulus each. Participant 1
35 miscategorized two stimuli, and Participant 5 was the only child to correctly sort all 27 stimuli. However, the
36 methodology employed did result in the both successful transfer of function and generalization of the stimuli
37 previously known and directly trained during MTS phases.
38
39
40
41
42
43
44
45
46
47
48
49
50
51

52
53 Interestingly, these generalization effects and errors demonstrated within the current study are similar to those
54 seen in young populations across disciplines (Quinn et al., 1997; Quinn, Eimas & Tarr, 2001). Furthermore, the
55 findings relate to the typicality effects seen in stimulus generalization in both equivalence (Galizio et al. 2004) and
56
57
58
59
60
61
62
63
64
65

1
2
3 prototype research (Rosch & Mervis, 1975; Rosch et al., 1976). Natural categories such as those targeted in the present
4 study have fuzzy boundaries, containing members that are not solely perceptual, therefore members may not be
5 entirely at basic category level. Therefore, it would not be expected that membership of any additional stimuli, that
6 participants have firstly, no experience with or secondly, do not do provide clear relatedness or similarity to known
7 members, become category members.
8
9

10
11
12
13
14 One other aspect to consider is that the current sort test was closed, in that the category names and containers
15 were given to the participants. A closed sort test is typically employed when taxonomy is already established for a
16 class or category (e.g., animals) and examines what knowledge participants demonstrate about the given taxonomy
17 Participants in the present study thus did not have any means to sort stimuli into a different type of sub category (e.g.,
18 furry things) as is generally examined using open sort testing (Fields et al., 1991). However, it should be noted that
19 the selection of both categories and stimuli for training and testing test did employ strict exclusionary criteria which
20 accounted for the possibility of the children sorting the cards as they may in an open test. Other means such as a mixed
21 test for conditional discriminations could be examined in future research to identify unknown stimuli.
22
23
24
25
26
27
28
29

30
31 Typically following training and testing under linear MTS protocols, tests for generalization are immediately
32 conducted. However, the current methodology uniquely included an additional aspect whereby the category name was
33 receptively trained to the C stimuli. This was purposefully trained to the C stimuli; the last to be trained following
34 MTS computerized stages and, given the participants' ages, was most likely to have been retained. The pre-
35 experimental sort test had identified two aspects of pre-experimental knowledge. Firstly, if the participants were
36 sorting based on the category name, or secondly, if sorting was as a result of reliance on perceptual commonalities
37 (Quinn et al., 1997; 2001; Zentall et al., 1996). In the present study the post category sort test was not conducted until
38 after the receptive training, and this may be a limitation of the procedure. It could be argued that the children may
39 have demonstrated categorization for the directly trained or derived relations found in the MTS procedure. Many of
40 the additional stimuli had been pre-experimentally paired during the sort test, but these were sorted incorrectly and
41 therefore not under the control of the category name. If the receptive category name was not trained to the C stimuli,
42 generalization effects to these additional class members would not be expected to occur as no relation between the
43 additional members (other than perhaps some perceptual similarity) exists. Future research might introduce a sort test
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3 prior to training the category names to establish the role that the directly trained members have on class expansion in
4 terms of perceptual commonality.
5
6

7
8 Another theoretical explanation for the generalization effects found in the present study could be Horne and
9 Lowe's (1996) common naming mechanism. During the studies, the category name was receptively trained to the C
10 stimuli for each of the three category sets before the post-category sort test was conducted. Further investigation could
11 be conducted to examine if category membership was found before the name was receptively trained to the C stimuli
12 through the introduction of additional sort testing probes. Furthermore, additional testing could be conducted to
13 establish if training the category name as speaker behavior would produce similar results.
14
15
16
17
18
19

20 The use of portable touch-screen devices, such as the one employed in the current study, is increasing among
21 preschool populations and there are many touch-screen educational applications aimed at these ages groups. All
22 participants in the current study had previous exposure to touch-screen technology apart from Participant 4. This meant
23 that training participants in the response mode was very quick. Comparable studies using fixed computers and a
24 computer mouse as a response device have reported considerable difficulties with point and click responses and the
25 long duration of training required by young children with and without developmental delays (Hourcade et al., 2008;
26 Shimizu et al., 2010). The current study provides an important step in disseminating research on evidence-based
27 teaching protocols using portable touch-screen methodologies. Future programs need not be confined to experimental
28 or educational settings as such technology provides a platform across settings such as within the home or in group
29 community.
30
31
32
33
34
35
36
37
38
39

40 The current study has demonstrated that EBI is an effective way to establish real world categories. The ability
41 to tailor the programme to target individualized skills in this populations means that the skills taught are age
42 appropriate and relevant for the learner. The ease with which the children learned to use the touch screen response
43 system is of particular interest, and offers further opportunity to examine even younger children and children with a
44 variety of disabilities.
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3 Ethical approval: “All procedures performed in studies involving human participants were in accordance with the
4 ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and
5 its later amendments or comparable ethical standards.”
6
7
8
9

10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65

- Arntzen, E., & Holth, P. (1997). Probability of stimulus equivalence as a function of training design. *The Psychological Record, 47*, 309-320. doi.org/10.1007/BF03395227
- Barnes, D., McCullagh, P. D., & Keenan, M. (1990). Equivalence class formation in non-hearing impaired children and hearing impaired children. *The Analysis of Verbal Behavior, 8*, 19. doi.org/10.1901/jeab.1986.46-243
- Devany, J. M., Hayes, S. C., & Nelson, R. O. (1986). Equivalence class formation in language- able and language- disabled children. *Journal of the Experimental Analysis of Behavior, 46*(3), 243-257. doi.org/10.1901/jeab.1986.46-243
- Dixon, M. R., Belisle, J., Stanley, C. R., Speelman, R. C., Rowsey, K. E., Kime, D., & Daar, J. H. (2016). Establishing derived categorical responding in children with disabilities using the PEAK- E curriculum. *Journal of Applied Behavior Analysis, 43*(1), 19-33. doi.org/10.1002/jaba.355
- Fienup, D. M., Covey, D. P., & Critchfield, T. S. (2010). Teaching brain—behavior relations economically with stimulus equivalence technology. *Journal of Applied Behavior Analysis, 43*(1), 19-33. doi.org/10.1901/jaba.2010.43-19
- Fienup, D. M., & Critchfield, T. S. (2010). Efficiently establishing concepts of inferential statistics and hypothesis decision making through contextually controlled equivalence classes. *Journal of Applied Behavior Analysis, 43*(3), 437-462. doi.org/10.1901/jaba.2010.43-437
- Fields, L., & Moss, P. (2008). Formation of partially and fully elaborated generalized equivalence classes. *Journal of the Experimental Analysis of Behavior, 90*(2), 135-168. doi.org/10.1901/jeab.2008.90-135
- Fields, L., Reeve, K. F., Adams, B. J., & Verhave, T. (1991). Stimulus generalization and equivalence classes: A model for natural categories. *Journal of the Experimental Analysis of Behavior, 55*(3), 305-312. doi.org/10.1901/jeab.1991.55-305

- 1
2
3 Fields, L., Travis, R., Roy, D., Yadlovker, E., Aguiar- Rocha, L. D., & Sturmey, P. (2009). Equivalence class
4 formation: A method for teaching statistical interactions. *Journal of Applied Behavior Analysis*, 42(3), 575-
5 593. doi.org/10.1901/jaba.2009.42-575
6
7
8
9
10 Fields, L., Verhave, T., & Fath, S. (1984). Stimulus equivalence and transitive associations: A methodological
11 analysis. *Journal of the Experimental Analysis of Behavior*, 42(1), 143-157. doi.org/10.1901/jeab.1984.42-143
12
13
14 Fisher, W., Piazza, C. C., Bowman, L. G., Hagopian, L. P., Owens, J. C., & Slevin, I. (1992). A comparison of two
15 approaches for identifying reinforcers for persons with severe and profound disabilities. *Journal of Applied*
16 *Behavior Analysis*, 25(2), 491–498. doi:10.1901/jaba.1992.25-491
17
18
19 Galizio, M., Stewart, K. L., & Pilgrim, C. (2004). Typicality effects in contingency-shaped generalized equivalence
20 classes. *Journal of the Experimental Analysis of Behavior*, 82(3), 253-273. doi.org/10.1901/jeab.2004.82-253
21
22
23 Groskreutz, N. C., Karsina, A., Miguel, C. F., & Groskreutz, M. P. (2010). Using complex auditory—visual samples
24 to produce emergent relations in children with autism. *Journal of Applied Behavior Analysis*, 43(1), 131-136.
25
26
27 doi.org/10.1901/jaba.2010.43-131
28
29
30 Haegele, K. M., McComas, J. J., Dixon, M., & Burns, M. K. (2011). Using a stimulus equivalence paradigm to teach
31 numerals, English words, and Native American words to preschool-age children. *Journal of Behavioral*
32 *Education*, 20(4), 283-296. doi.org/10.1007/s10864-011-9134-9
33
34
35
36 Horne, P. J., & Lowe, C. F. (1996). On the origins of naming and other symbolic behavior. *Journal of the Experimental*
37 *Analysis of Behavior*, 65(1), 185–241. doi:10.1901/jeab.1996.65-185
38
39
40 Hourcade, J. P., Perry, K. B., & Sharma, A. (2008, June). PointAssist: helping four year olds point with ease.
41 In *Proceedings of the 7th international conference on Interaction design and children* (pp. 202-209). ACM.
42
43
44 doi.org/10.1145/1463689.1463757
45
46
47 LeBlanc, L. A., Miguel, C. F., Cummings, A. R., Goldsmith, T. R., & Carr, J. E. (2003). The effects of three stimulus-
48 equivalence testing conditions on emergent US geography relations of children diagnosed with
49 autism. *Behavioral Interventions*, 18(4), 279-289. doi.org/10.1002/bin.144
50
51
52 Libby, M.E., Weiss, J.S., Bancroft, S., Ahearn, W.H., 2008. A Comparison of Most-to-Least and Least-to-Most
53 Prompting on the Acquisition of Solitary Play Skills. *Behav. Anal. Pract.* 1, 37–43.
54
55
56 doi:10.1007/BF03391719 Miguel, C. F., Petursdottir, A. I., Carr, J. E., & Michael, J. (2008). The role of naming
57
58
59
60
61
62
63
64
65

- 1
2
3 in stimulus categorization by preschool children. *Journal of the Experimental Analysis of Behavior*, 89(3), 383-
4 405. doi.org/10.1901/jeab.2008-89-383
5
6
7 Nason, S., & Zabrucky, K. (1988). A program for comprehension monitoring of text using HyperCard for the
8 Macintosh. *Behavior Research Methods, Instruments, & Computers*, 20, 499-502.
9 doi.org/10.3758/bf03202704
10
11
12
13
14 Ninness, C., Barnes- Holmes, D., Rumph, R., McCuller, G., Ford, A. M., Payne, R., ... & Elliott, M. P. (2006).
15 Transformations of mathematical and stimulus functions. *Journal of Applied Behavior Analysis*, 39(3), 299-
16 321. doi.org/10.1901/jaba.2006.139-05
17
18
19
20 Ninness, C., Dixon, M., Barnes- Holmes, D., Rehfeldt, R. A., Rumph, R., McCuller, G., ... & McGinty, J. (2009).
21 Constructing and deriving reciprocal trigonometric relations: A functional analytic approach. *Journal of*
22 *Applied Behavior Analysis*, 42(2), 191-208. doi.org/10.1901/jaba.2009.42-191
23
24
25
26 Pytte, C. L., & Fienup, D. M. (2012). Using Equivalence-Based Instruction to Increase Efficiency in Teaching
27 Neuroanatomy. *Journal of Undergraduate Neuroscience Education*, 10(2), A125.
28
29
30
31 Quinn, P., & Eimas, P. (1997). A reexamination of the perceptual-to-conceptual shift in mental representations. *Review*
32 *of General Psychology*, 1(3), 271. doi.org/10.1037/1089-2680.1.3.271.
33
34
35 Quinn, P., Eimas, P., & Tarr, M. (2001). Perceptual categorization of cat and dog silhouettes by 3-to 4-month-old
36 infants. *Journal of Experimental Child Psychology*, 79(1), 78-94. doi.org/10.1006/jecp.2000.2609.
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

- 1
2
3 Shimizu, H., Yoon, S., & McDonough, C. S. (2010). Teaching skills to use a computer mouse in preschoolers with
4
5 developmental disabilities: Shaping moving a mouse and eye–hand coordination. *Research in Developmental*
6
7 *Disabilities, 31*(6), 1448-1461. doi.org/10.1016/j.ridd.2010.06.013
8
9
10 Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: An expansion of the testing
11
12 paradigm. *Journal of the Experimental Analysis of Behavior, 37*(1), 5-22. doi.org/10.1901/jeab.1982.37-5
13
14 Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech & Hearing Research, 14*(1), 5-13.
15
16 doi.org/10.1044/jshr.1401.05
17
18 Still, K., Rehfeldt, R. A., Whelan, R., May, R., & Dymond, S. (2014). Facilitating requesting skills using high-tech
19
20 augmentative and alternative communication devices with individuals with autism spectrum disorders: A
21
22 systematic review. *Research in Autism Spectrum Disorders, 8*(9), 1184-1199.
23
24 doi.org/10.1016/j.rasd.2014.06.003
25
26 Zentall, T. R., & Smeets, P. M. (Eds.). (1996). *Stimulus class formation in humans and animals*. Elsevier.Zentall, G.
27
28 A. (2002). Categorization, Concept Learning, and Behavior Analysis: An Introduction. *Journal of the*
29
30 *Experimental Analysis of Behavior, 78* (3), 237-248. doi.org/10.1901/jeab.2002.78-237
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Table 1. Category sets and stimuli for each participant.

Participant	Stimuli	Set 1	Set 2	Set 3
1	A	Ant	Cherry	Iron
	B	Fly	Strawberry	TV
	C	Grasshopper	Watermelon	Vacuum
		Accessories	Utensils	Fruit
2	A	Belt	Masher	Kiwi
	B	Ring	Grater	Peach
	C	Watch	Ladle	Pineapple
		Dairy	Vegetables	Furniture
3	A	Butter	Broccoli	Wardrobe
	B	Swiss Cheese	Carrots	Sofa
	C	Milk	Potato	Table
		Sweets	Vegetables	Furniture
4	A	Candy floss	Carrots	Wardrobe
	B	Lollipop	Peas	Sofa
	C	Popcorn	Potato	Desk
		Body parts	Musical Instruments	Furniture
5	A	Hand	Drums	Bed
	B	Leg	Piano	Wardrobe
	C	Eye	Guitar	Desk

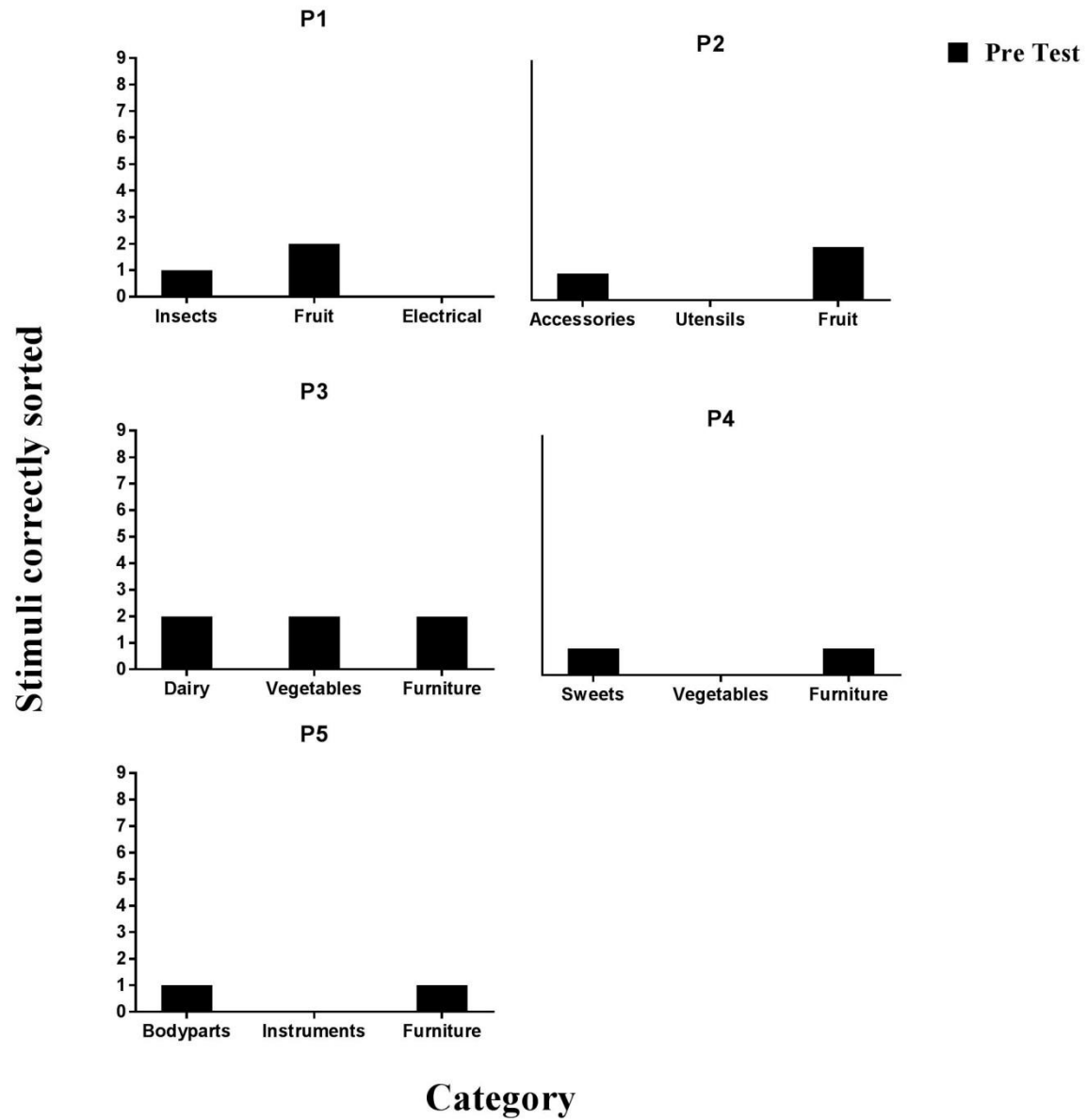
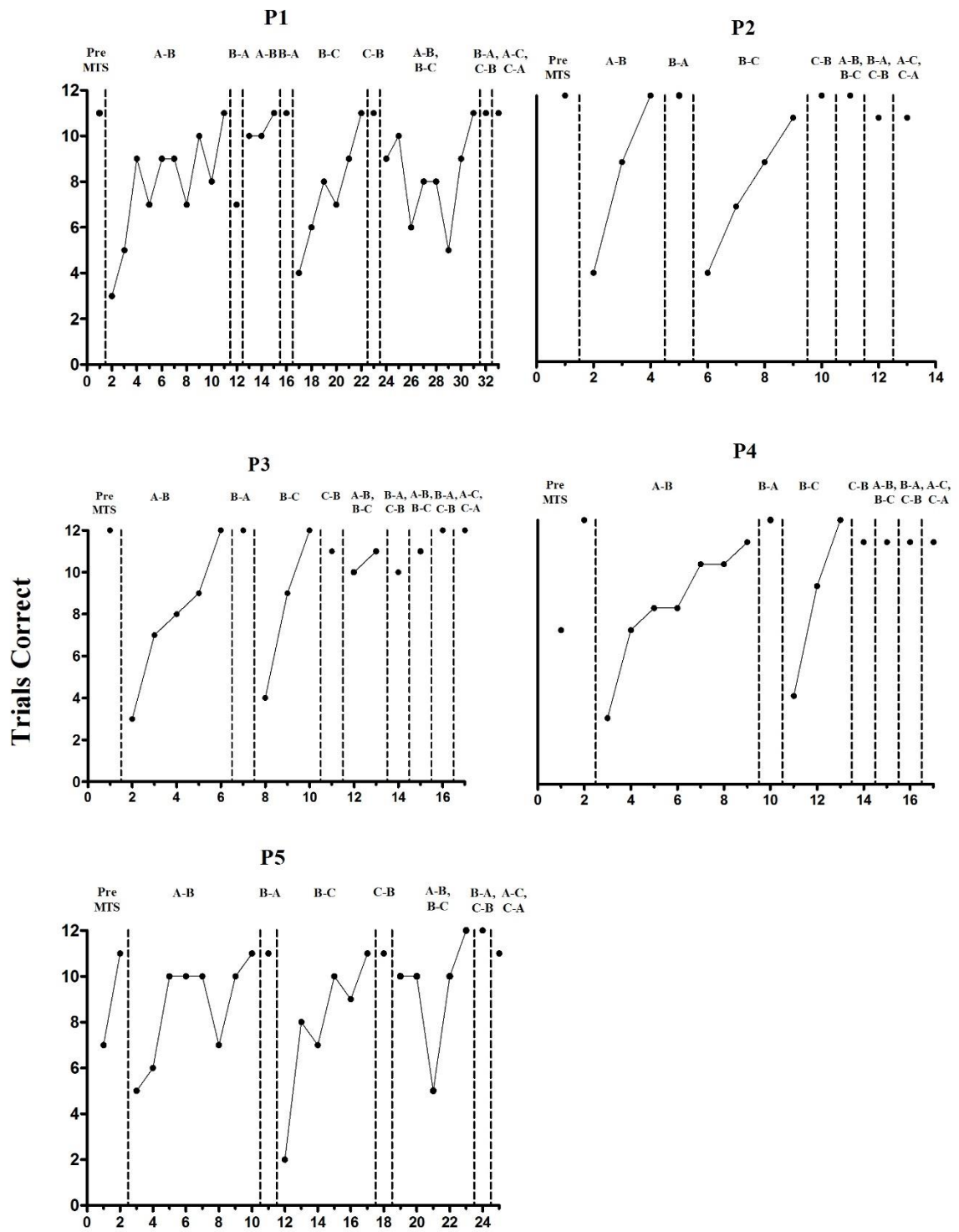


Figure 1. Results of pre-experimental category sort test, stimuli correctly sorted for each of the three categories (9 stimuli per category) tested for Participant 1-5.



Blocks

Figure 2. Participant 1-5 results for MTS Pre-training, A-B Training, B-A Testing, B-C Training, C-B Testing, Mixed A-B, B-C Training, Mixed B-A, C-B Testing and A-C, C-A Symmetry Transitivity (Equivalence) Testing.

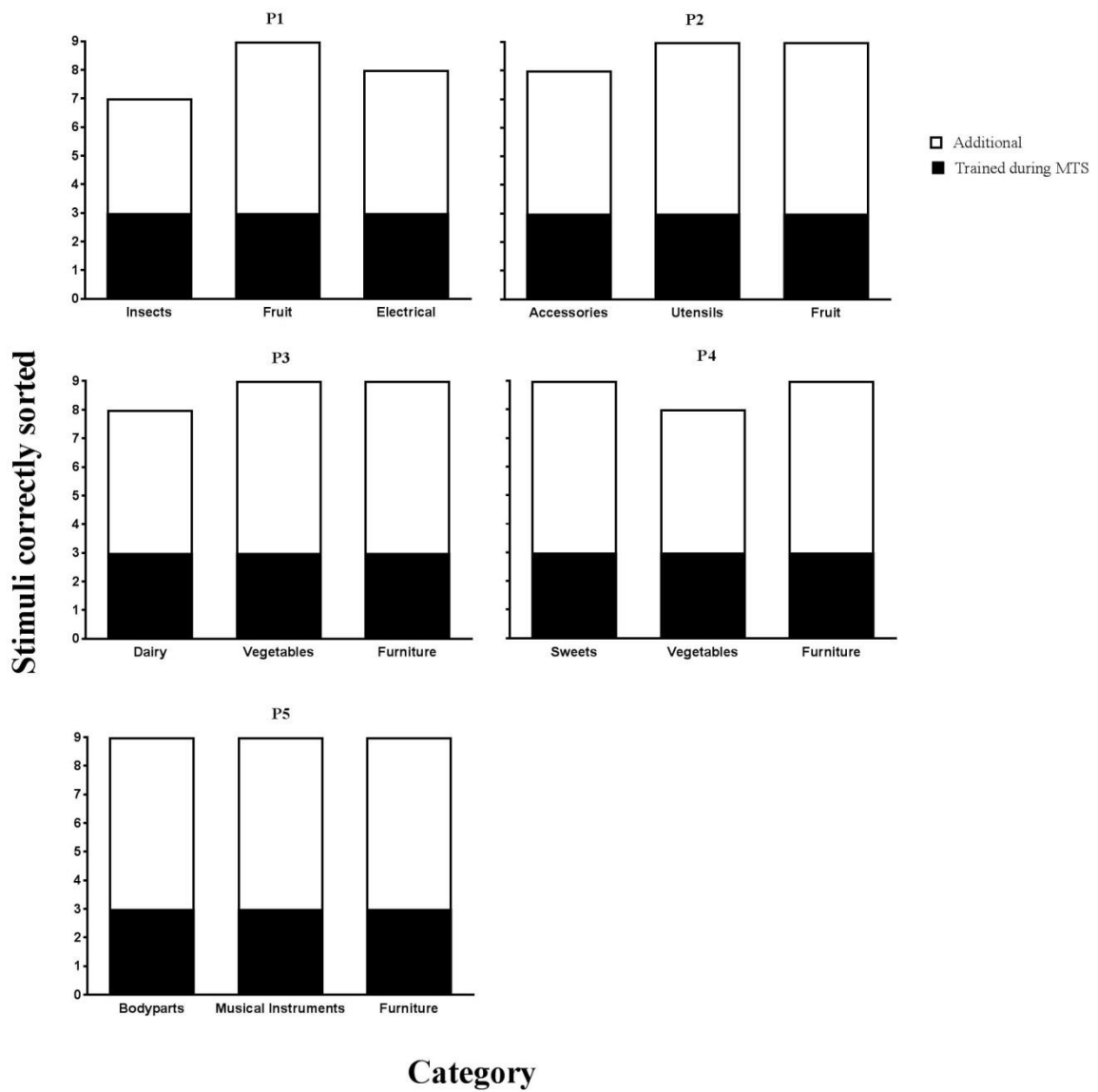


Figure 3. Category sort test results for all participants showing the total stimuli correctly sorted out of nine for each of the three category sets. Results show the three stimuli training during MTS training in black and additional untrained relations which were sorted correctly at post-testing in white.

Sample Revisions Matrix

Introduction:

Revision Suggestion	Modification Made/Comment	Page/ Reference
<p>Editor: On p. 3 of your submission, the following details regarding touch-screen technology are described: One type of response which has received little mention within the literature is that of interactive touch-screen technology (cf. Arnzten & Holtz, 1997; Saunders, Drake, & Spradlin, 1999, who used an early type of fixed touch-screen). In might be mentioned that as early as 1988, Nason and Zabucky conducted an investigation with an attempt to employ an approximation of touch-screen instructional technology (HyperCard operating on Mac Plus, Mac SE, or Mac II). Admittedly, these units were rather clunky by current standards. More to your point, this now antiquated system required mouse clicks rather than having a student simply touch the screen. Nevertheless, it seems reasonable to provide the reader with some historical context regarding a few of the early attempts at developing computer-interactive instructional technology.</p>	<p>Authors: Additional information has been added on p.3 of the Introduction. This outlines the historical context and importance of using new technology. Full changes can be seen on p. 3</p> <p>One type of response which has received little mention within the literature is interactive touch-screen technology which enables the child to reach and touch a stimulus directly, rather than having to respond elsewhere (cf. Arntzen & Holth, 1997; Saunders, Drake, & Spradlin, 1999, who used an early type of fixed touch-screen; see also Nason & Zabucky, 1988, for an even earlier attempt to use this type of technology).</p>	P. 3
<p>Editor: One of the primary contributions of your submission revolves around the advantage of using touch-screen technology to facilitate EBI protocols; however, in various parts of your methodology, you describe a series of table-top assessments. I am confused as to why pre and post testing was not conducted by way of your touch-screen system. Along these same lines, Reviewer 1 notes: Given that a stated goal of the research was to explore the utility of touch-screen technology in this research area, it seems somewhat odd that the category (pre and post) testing did not occur on the touchscreen. Since your study focused on an evaluation of EBI “real-world categories,” perhaps table top assessment was considered more consistent with generalized outcomes? In</p>	<p>Authors: The use of table top was designed to assess generalisation to other settings. This has been clarified on p.3</p> <p>Furthermore, EBI leads to the acquisition of additional untrained relations by design, while conventional classroom instruction through exemplars and rules does not necessarily have this generative outcome.</p> <p>And p. 5 Line 11</p> <p>To ensure generalization from the touchscreen to familiar classroom settings and activities, category sorting was assessed in a table top task prior to any computer-based training. Furthermore, generalization to other, untrained, category members was measured after training and testing with a table top task</p> <p>and p. 14 Line.1</p> <p>or to a fourth irrelevant category. This exercise</p>	<p>p. 3</p> <p>And p. 5 Line 11</p> <p>p. 14 Line 1</p>

<p>any event, a revised version of your manuscript should address this particular issue within the introduction section of your paper.</p>	<p>therefore tested for the generalization to untrained and untested stimuli.</p>	
<p>Editor: Your submission becomes somewhat difficult to follow during your descriptions of how assessments were conducted. For example, Reviewer 1 states: ...only when reading the method was it clear to me that the authors assessed generalization in their post-testing (i.e., their "additional" items). I think the manuscript could be strengthened by discussing this issue in their intro, helping readers understand the value of this experiment.</p>	<p>Authors: This has been addressed in the Introduction (final paragraph p. 3) and further clarification is offered in the Method as outlined in the above response.</p> <p>To ensure generalization from the touchscreen to familiar classroom settings and activities, category sorting was assessed in a table top task prior to any computer-based training. Furthermore, generalization to other, untrained, category members was measured after training and testing with a table top task.</p>	<p style="text-align: center;">final paragraph p. 3</p> <p style="text-align: center;">P 10</p>
<p>Reviewer 1: First, the introduction needs to be modified such that it (1) condenses what is currently presented and (2) adds some new content. Regarding the first point, the current text can be streamlined such that the authors more directly (and concisely) state what is unique about their study. This reduced presentation will be helpful because it will allow readers to more effectively compare the current paper to the published literature. I was not as convinced as the authors that there is a considerable amount of novelty here.</p>	<p>The final paragraph of the introduction has been amended to focus on the unique aspects of the manuscript.</p> <p>Rehfeldt (2011) called for ecologically valid research to investigate the application of EBI to naturalistic educational settings. The purpose of the current study was to take some steps towards this by employing a touch-screen computerized MTS program to teach functional real world categorization skills to young neurotypical children. Each child was pre-screened and then taught individualized categories which were previously unknown to them but appropriate to their developmental level, thus focusing on existing deficits and making training relevant for the learner. To ensure</p>	<p style="text-align: center;">p. 3</p>

	<p>generalization from the touchscreen to familiar classroom settings and activities, category sorting was assessed in a table top task prior to any computer-based training. Furthermore, generalization to other, untrained, category members was measured after training and testing with a table top task.</p>	
<p>Regarding the second point, only when reading the method was it clear to me that the authors assessed generalization in their post-testing (i.e., their "additional" items). I think the manuscript could be strengthened by discussing this issue in their intro, helping readers understand the value of this experiment. Without it, I am uncertain as to whether there is enough here to warrant publication. (R1)</p>	<p>Authors: As above the use of table top procedures and card sort tests (printed pictorial stimuli) was designed to assess generalisation to other material/settings.</p> <p>This has been clarified on p.3</p> <p>Furthermore, EBI leads to the acquisition of additional untrained relations by design, while conventional classroom instruction through exemplars and rules does not necessarily have this generative outcome.</p> <p>And p. 5 Line 11</p> <p>To ensure generalization from the touchscreen to familiar classroom settings and activities, category sorting was assessed in a table top task prior to any computer-based training. Furthermore, generalization to other, untrained, category members was measured after training and testing with a table top task</p> <p>and p. 14 Line.1</p> <p>or to a fourth irrelevant category. This exercise therefore tested for the generalization to untrained and untested stimuli.</p>	<p>p. 3</p> <p>And p. 5 Line 11</p> <p>p. 14 Line 1</p>

Method:

Revision Suggestion	Modification Made	Page/ Reference
<p>Editor: Methodology Reviewer 4 has a particularly provocative question regarding your relatively complex EBI procedures. The reviewer acknowledges that EBI is not his/her special area of research. Notwithstanding, I believe his/her question should be asked regarding any complex training</p>	<p>Authors: This issue has been addressed in the Introduction (P. 3) and also Paragraph 2 of the Discussion, (p. 13) and highlighted as an area for future investigation. It has been highlighted that the generative nature of equivalence responding is of critical importance. Furthermore, proposing the use of abstract rules such as this is “an object worn in the hair” is a complicated issue. Rules such as that may</p>	<p>Addressed at several points throughout the manuscript Introductio</p>

<p>procedures that employ normal functioning participants:</p> <p>I kept wondering while reading the relatively complex methods of the EBI procedure and reviewing the number of trials necessary whether this type of procedure should be used to teach the type of target skills in the current study, particularly in teaching typically developing preschool children that have good receptive skills. Might providing the participants with some rules regarding the categories and stimuli increased the efficiency of acquisition and generalization? That is, might an explanation of the category and description of things that fit in that category suffice? In addition, I wondered whether the degree to which a participant did not know what a particular stimulus was (e.g., that a scrunchie was worn in the hair) influenced the results, which could have been addressed with some simple rules. I guess I'm asking whether relations need to be derived in the current way if they can be taught using rules. Again, I understand this area of research is outside of my area of expertise, but with respect to a larger audience, the authors might consider addressing this in their discussion section.</p> <p>Forgive my laboring this particularly thorny issue; however, it is not unusual for behavior analytic researchers to develop intricate EBI strategies and simply neglect the possibility that their participants may learn relations more efficiently by way of rules. Many researchers have refrained from appealing to rule governance to explain the formation of stimulus relations or even to develop training procedures that employ instructions. However, as a practical matter, preschool teachers (or designers of computer-assisted software for preschoolers) would find it untenable to expect children to become skilled at pre-academic/abstract tasks without first introducing them to any rules that might facilitate their "understanding" of the relations among stimuli.</p>	<p>actually comprise functional equivalence training but certainly require a degree of derivation and the same processes may underlie this as more traditional matching related derivation (See Smyth, S., Barnes-Holmes, D., & Barnes-Holmes, Y. (2008). Acquired equivalence in human discrimination learning: The role of propositional knowledge. <i>Journal of Experimental Psychology: Animal Behavior Processes</i>, 34(1), 167.) It is not clear to us that instruction through the use of rules could be extended to make it generative without making it too difficult for this age group. A limited examination of the literature on category learning in young children suggests that other techniques involving presentation of exemplars may work but not the type of rule governance discussed by this reviewer. While additional mention of this and of issues such as the insensitivity effect in rule governed behaviour could be addressed, it appears to us that this is not central to the current research.</p>	<p>n p.3</p> <p>Paragraph 2 of the Discussion , (p. 13)</p>
--	---	---

<p>I share Reviewer 4’s fundamental concern regarding EBI training procedures that fail to employ (or even acknowledge) the potential value of rule governance in the training of stimulus relations. A revised version of your paper should carefully speak to the issue of EBI training absent rule governance. I suggest that this critically important issue should be addressed within the introduction and discussion sections of the revised manuscript.</p>		
<p>Editor: Reviewer 3 expressed some difficulty following your methods stating: In the Settings and Materials section, the authors described trial blocks. I had a hard time following this because the authors hadn’t yet gone into detail on the overall procedures of the study. I think it would be helpful to remove references to specific procedures from the Settings and Materials section.</p>	<p>Authors: This has been addressed by removing mention of block before they have been described in the procedure (pp. 4-5).</p>	<p>Removal pp. 4-5</p>
<p>Editor: Along similar lines, this reviewer indicates that it would be helpful if your paper provided terminology clarifications: In the Method section, the authors noted that all participants had, “good receptive language” (p. 3, line 28). I think it would be helpful if the authors gave a brief definition of what this means, as many readers, both those who are new to the field and those who don’t work with young children, may have difficulty conceptualizing what this entails.</p>	<p>Authors This has been addressed by outlining participant performance on the language assessment rather than referring to the level of receptive language.(p. 4) .</p>	<p>p.4</p>
<p>Editor: Software Your paper focuses on current technological applications to equivalence-based instruction. Thus, readers will be interested in the origin and availability of your EBI application. On p. 5 of your submission you provide a detailed description of your touch-screen interactive procedures:</p> <p>During each MTS pre-training trial, a black shape (triangle, square or circle) appeared in the top center of the screen.</p>	<p>Authors: The programme was written in Visual Basic by one of the authors. In order to programme for touchscreen in Visual Basic 6.0, a mouse click is coded as the input. The programme will work on any windows operating system. This information has been added to the manuscript (p. 5)</p>	<p>End of p-4 and start of p. 5</p>

The participant was required to touch this shape. It immediately disappeared from the screen and an array of three comparison shapes (triangle, square & circle) appeared in the lower area of the screen, one to the right, one to the center and one to the left. The location of the comparison stimuli was randomized across trials. The experimenter issued the instruction ‘Match’ at the beginning of each trial (appearance of the sample) and if no response was made after 5 s, the instruction was reissued. A correct response was recorded by the computer if the participant used one or more of their fingers to touch the screen and select the corresponding picture from the lower portion of screen (e.g., if the sample was circle, select circle from the array of three comparison shapes). [Related interactive training details continue]

The above description represents a fairly complex automated MTS application. Such computer-interactivity usually entails custom MTS software written in any one of several popular programming languages (e.g., Visual Basic, Visual C#, ActionScript, etc.); however, I could not locate any reference to the custom software designation or how this software system was developed.

The computer employed in your paper is described as an 8.9-inch Asus Eee PC T91 Touchscreen Netbook. This is a rather dated piece of hardware “usually” operating on Microsoft Windows XP. I understand that the computer comes with tablet-specific software ("touchsuite") that presents images and has touch-enabled browsing with an on-screen keyboard. Still, it is not at all clear as to how this particular computer performed the activities involved in the above training protocol. Was some type of training template built into the Asus Netbook, or did you employ a form of customized software?

A revised version of this paper should provide a clearer description of the EBI touch-screen program’s development and its current availability for readers. Based

<p>on your description of the program I cannot be certain, but it is possible that your study may have been conducted on a rather dated operating system. Readers would certainly benefit from information regarding how they might be able to obtain similar interactive functionality (i.e., existing MTS touch-screen protocols that are similar or the same as the system employed in your study). It will be critically important to know if the software used in your particular study is available (perhaps downloadable) for more current operating systems. In the absence of EBI program availability for readers, it will be especially difficult to recommend your paper for publication.</p>		
<p>Editor: Testing Procedures Reviewer 5 was particularly enthusiastic about your strategies and results. Nevertheless, this reviewer provides a series of challenging questions (and phraseology edits) all of which should be addressed and resolved in a revised version of your current submission. In particular, I share this reviewer’s concerns regarding the readability /clarity of your testing procedures: For the Pre experimental Category Sort Test, it might have been helpful to see some example pictures. The children seemingly remembered the locations of the category labels spoken to them. Perhaps it should be made clearer that the experimenters were looking for categories of which the children’s choices indicated they seemingly had no knowledge. Presumably for every child at least 3 such categories were found, and these were then used in the MTS phase. I am not quite clear, given that there was a total of nine stimuli for each category, how three of each were allocated to the MTS tests.</p>	<p>Authors: Additional information has been added to p. 5 and p. 6 in order to clarify this point.</p> <p>Inclusion of the pictures is not possible due to copyrighting, however the source has been cited and the stimuli name in Table 1 so they may be sourced by the reader.</p> <p>All picture stimuli used for either category sorting, training or computerized phases of the experiment were obtained from Picture This 3.0 Professional Edition (CD©; Silver Lining Multimedia, Inc, 1999).</p>	<p style="text-align: right;">p. 5 and p. 6</p> <p style="text-align: right;">p. 4</p>
<p>Reviewer 1: The second and third revisions are similar in that both the method and (especially) results sections need to be condensed. For example, the description of the MTS pre-training in the method section seems unnecessarily long,</p>	<p>Authors: have made effort to reduce/condense where possible.</p>	<p>Changes to reduce/condense have been made throughout the method and results</p>

<p>and the descriptions of Figures 1 and 2 in the results section can be reduced considerably.</p>		<p>sections of the manuscript</p>
<p>Reviewer 3: In the Method section, the authors noted that all participants had, “good receptive language” (p. 3, line 28). I think it would be helpful if the authors gave a brief definition of what this means, as many readers, both those who are new to the field and those who don’t work with young children, may have difficulty conceptualizing what this entails.</p>	<p>Addressed (see response to Editor’s comments)</p>	

Results:

Revision Suggestion	Modification Made	Page/Reference
<p>Editor: Results Reviewer 3 provides a series of questions and recommendations regarding outcomes. In particular, this reviewer had difficulty tracking your description of results relative to your illustrations.</p> <p>In the Results section, when discussing each pattern of responding, I think it would be helpful if the authors directed the reader to where each participant’s data can be found in each Figure. For example, in the MTS Training and testing section, when discussing the data from participant 1, the authors could simply state, “Participant 1 (top, left panel)....”</p>	<p>Authors: The results section has been considerably condensed making the tracking of results from the text less necessary. However where relevant, the reader is now directed to the relevant portion of the figure.</p>	<p>Addressed throughout manuscript</p>
<p>Figure 2 present percent correct as a function of "stimulus type" (i.e., pre-test, trained, and additional). There were no significant differences as a function of category type such that the data do not need to be separated in that way. Presenting the accuracies on these different trial types would be less cumbersome to readers. (R1)</p>	<p>Figure 2 has been amended to simplify it. The accompanying text has also been changed.</p>	<p>p. 10 and throughout the results section</p>
<p>In the Results section, when discussing each pattern of responding, I think it would be helpful if the authors directed the reader to where each participant’s data can be found in each Figure. For example,</p>	<p>Addressed (see response to Editor’s comments)</p> <p>This has been addressed by outlining participant performance on the language assessment rather than</p>	<p>p.4</p>

<p>in the MTS Training and testing section, when discussing the data from participant 1, the authors could simply state, “Participant 1 (top, left panel)....” (R3)</p>	<p>referring to the level of receptive language.(p. 4) .</p>	
<p>In Figure 1, I think the y-axis label may have been inadvertently cropped, as what is depicted appears to me the number of trials with a correct response. (R3)</p>	<p>The y-axis may have been inadvertently cropped. I have been unable to resolve this issue as the files as set to publishers recommendations, however, direct links are provided to uploaded images or could be amended in a word file.</p>	<p style="text-align: center;">Results section</p>
<p>I had a bit of a hard time following the MTS Training and testing section. Given that the main focus of the study is on the derived category relations, I think it would be helpful reduce the length of this section to a paragraph. In doing so, the authors could focus on cross-participant patterns, such as describing the range of trial blocks necessary to obtain criterion-level performance and the fact that the number of blocks necessary decreased as training progressed. (R3)</p>	<p>This section has been reduced in length and greater focus has been placed on participant performance.</p>	<p style="text-align: center;">Throughout results section</p>
<p>. I also had a hard time following the Post-Experimental Category Sort Testing section. I think it would be helpful if the section was re-organized to focus on the quantitative results. The graphs in Figure 2 adequately denote the different categories used for each participant. Rather than going into such detail on the categories, I think it would be helpful if the authors gave the specific quantitative values for each participant across the three phases, pretest, training, and post-test. (R3)</p>	<p>Addressed (see response to Editor’s comments)</p> <p>The figures have been amended to focus on quantitative values.</p>	

Discussion:

Revision Suggestion	Modification Made	Page/Reference
<p>Editor: Discussion As noted by Reviewer 5, in large part your study is a replication of many earlier investigations in the area of EBI for teaching skills to children. What makes your paper most distinctive is your special emphasis on individualized training by way of computer touchscreen technology. Reviewer 1 has concerns regarding the degree to which your submission makes contact with a wider range of theoretical issues within the field of EBI instruction:</p> <p>...the discussion section needs to be more impactful to warrant publication. I understand that a goal of the manuscript is to consider how to implement EBI under these types of conditions, but there needs to be greater contact with conceptual and/or theoretical issues in the field to justify publication.</p>	<p>Authors: Considerable revisions to the Discussion (in particular pp. 14-15) have addressed this issue. Further information about taxonomy, and generalization were a particular focus of this revision.</p>	<p>Added changes throughout p. 14-15</p>
<p>Finally, the discussion section needs to be more impactful to warrant publication. I understand that a goal of the manuscript is to consider how to implement EBI under these types of conditions, but there needs to be greater contact with conceptual and/or theoretical issues in the field to justify publication. (R1)</p>	<p>The discussion section has been amended in parts to address impact of the research. Additionally, theoretical issues have been discussed in more detail.</p>	<p>pp. 12 - 15</p>
<p>I found the Discussion section tricky to follow. 1. I think it would be helpful to include more quantitative, summary values. For example, the authors stated, “also most of the additional untrained stimuli” (p. 11, lines 39-40). I think it would be helpful to simply give a quantitative value so the reader can make his or her own decision on how to classify this pattern. (R3)</p>	<p>Specific quantitative values have been added to the discussion.</p> <p>Furthermore, at post-testing all participants accurately categorized the stimuli that were pre-experimentally known, the stimuli that were trained during match-to-sample phases and all participants successfully matched at least 15 of 18 additional untrained stimuli (Participant 1 matched 15 stimuli, while Participant 5 matched all 18 and the other three participants matching 17 of the 18).</p>	<p>p. 12</p>
<p>I think the paragraph on the use of “portable touch-screen devices” (p. 11, line 42) could be reduced in length.</p>	<p>The study focus was on both the use of touchscreen and the real-world individualised categories. This has been</p>	<p>Throughout manuscript</p>

<p>Although I whole-heartedly agree that this study is a great demonstration of how these ubiquitous devices can be used for a variety of training purposes, this seemed a bit out of place given that the study was focused on an evaluation of EBI for real-world categories.(R3)</p>	<p>clarified throughout.</p>	
<p>I didn't understand exactly what the authors were saying in the paragraph on page 12, lines 42-48. (R3)</p>	<p>This point may have been addressed in the rewrite of the discussion section.</p>	<p>Discussion section pp. 12 - 15</p>

Minor Edits:

Revision Suggestion	Modification Made	Page/ Reference
<p>Reviewer 1: minor issues (in order of appearance in the manuscript): 1. In the first paragraph, it should be noted that the A-C trials test transitivity and the C-A trials test symmetry plus transitivity, or equivalence.</p>	<p>Addressed on p.1</p>	
<p>Reviewer 1: 2. More details are needed in the description of Haegele et al. (2011). For example, did it involve MTS? Also, here and below, I think the phrase "taught to pair" probably should be reworded to something along the lines of "taught to select in the presence of." Or, at least, tell the reader what is meant by the phrase "taught to pair."</p>	<p>This section has been removed from the introduction and additional information on Haegele et al. (2011) has been added.</p>	<p>See p 2 for revisions</p>
<p>Reviewer 1 3. In the A-B training section in the method, state explicitly what happened after an incorrect response. Presumably it was comparable to the MTS pre-training such that the experimenter did nothing?</p>	<p>This has been clarified and expanded prior to outlining the MTS training phases.</p> <p>As well as the visual feedback following a correct response, the experimenter delivered a token and verbal feedback (e.g., 'good matching', 'nice matching', or 'super matching'). No other feedback was given for incorrect responses. Each MTS pre-training block consisted of 12 trials during which each of the three shapes was presented as a sample four times. At the end of the block the word "finished" appeared on the</p>	<p>p. 7</p>

	<p>screen. The experimenter pointed to the screen and stated, ‘matching is finished’. A minimum break of 5 min occurred between each block of trials, this was inclusive of time with access to toys earned.</p> <p>Participants were required to make 11 correct responses in a 12-trial block in order to proceed to the first experimental phase.</p>	
<p>Reviewer 1: The results presented at the end of the results section about pre-training should appear earlier. In other words, the results should be presented in the order in which they were obtained.</p>	<p>This has been moved to the start of the results</p>	<p>p. 10</p>
<p>Reviewer 1: Given that a stated goal of the research was to explore the utility of touch-screen technology in this research area, it seems somewhat odd that the category (pre and post) testing did not occur on the touchscreen. This issue needs to be discussed, at least briefly, in the discussion section.</p>	<p>Authors: The use of table top was designed to assess generalisation to other settings. This has been clarified on p.3</p> <p>Furthermore, EBI leads to the acquisition of additional untrained relations by design, while conventional classroom instruction through exemplars and rules does not necessarily have this generative outcome.</p> <p>And p. 5 Line 11</p> <p>To ensure generalization from the touchscreen to familiar classroom settings and activities, category sorting was assessed in a table top task prior to any computer-based training. Furthermore, generalization to other, untrained, category members was measured after training and testing with a table top task</p> <p>and p. 14 Line.1</p> <p>or to a fourth irrelevant category. This exercise therefore tested for the generalization to untrained and untested stimuli.</p>	<p>p. 3</p> <p>And p. 5 Line 11</p> <p>p. 14 Line 1</p>
<p>Reviewer 3 I think the authors meant the word “assent” rather than “ascent” when describing the children’s agreement to participate (p. 3, line 37).</p>	<p>Changed to assent</p>	<p>p. 4</p>
<p>I think the manuscript would benefit from close look-over for APA style issues.</p>	<p>Addressed</p>	<p>Throughout manuscript</p>
<p>On p. 3, line 45, “minutes” should be “min.”</p>	<p>Changed to min</p>	<p>p. 4</p>

<p>When including units, there should be a space between the numeral and the unit</p>	<p>Changed</p>	<p>Throughout manuscript</p>
<p>The headings should be revised (using bolding and including italics for only levels 4 and 5)</p>	<p>Changed</p>	<p>Throughout manuscript</p>
<p>In the Settings and Materials section, the authors discussed the use of preferred stimuli. Throughout this section, the authors used the term “reinforcers” and “secondary reinforcers.” Based on the text, it appears the authors used items identified via a preference assessment but didn’t explicitly test the reinforcing efficacy of the items. If that is the case, I think it would be more conceptually consistent to refer to the items as “high preference” and “moderate preference” rather than as reinforcers.</p>	<p>This has been clarified.</p> <p>Prior to any other testing or training, a two-choice preference assessment (Fisher et al., 1992) was conducted using a variety of developmentally appropriate toys. Pictures of the toys were affixed to a board later presented as choices for participants following an experimental training block. Access to highly preferred toys was contingent upon performance during training, with each token earning 30 s of play following completion of a block, and, in order to maintain motivation, less preferred toys were presented for 60 s if no tokens had been earned in that block.</p> <p>Additionally, all the participants engaged with the researcher using the visual/textual instructions which made specific reference to play during training sessions.</p>	<p>p. 4</p> <p>And instructions p. 5</p>
<p>In the Settings and Materials section, the authors described trial blocks. I had a hard time following this because the authors hadn’t yet gone into detail on the overall procedures of the study. I think it would be helpful to remove references to specific procedures from the Settings and Materials section.</p> <p>The Phase 1 and Phase 2 sections, the authors stated there was remediation if a certain pattern of responding didn’t occur. However, the authors indicated the remediation was never needed. I think it would aid in clarity if the references to remediation were removed.</p>	<p>Addressed, removed (see response to Editor’s comments)</p>	<p>Settings and materials section</p>
<p>In the descriptions of each phase, I think it would be helpful to remind the reader of the criterion for progression (maybe it was just me, but I kept returning to the “A-B” section to remind myself). I think it could be something as simple as, “If the 11/12 correct trials criterion was met....”</p>	<p>Addressed criterion reference at frequent points for MTS phases (see response to Editor’s comments)</p>	<p>Throughout method section</p>
<p>I think the description of the Pre experimental Category Sort Test could be re-arranged. Specifically, the authors reference “meeting criteria” (p. 4, line 51) in the first paragraph but don’t</p>	<p>This has been rearranged:</p> <p>Categories for the computerized stages of the study were chosen based on the results of the pre-experimental category sort test. Category exclusion</p>	<p>p. 5 and p. 6</p>

<p>define the criteria until the final paragraph. I think it would be helpful to include the specific inclusion and exclusion criteria first and then describe the specific procedure used to identify categories to include.</p>	<p>occurred when a participant placed more than three of the nine pictures from the same category set into the same container, regardless of the container's category label (ie even if they miscategorised 3 or more pictures together). This conservative strategy was employed to minimize the possibility of prior knowledge of a to-be-taught category. Participants were tested across several categories until three sets had been identified as meeting the criterion outlined. Those categories identified were finally tested together to ensure that they still met criteria and ensure minimal overlap in topography of images The inclusion criterion for the exact stimuli to be trained and tested using the computer-based programme was that none of the three chosen should have been sorted in the same container (See Table 1 for the categories and stimuli chosen for each participant).</p>	
<p>I wasn't quite sure why the authors set the criterion as 1 trial block with 11/12 correct trials. Specifically, I was curious as to whether the authors any other within-block criteria, such as no incorrect responses within the last X trials. Given that only one block was required, I think it would be important to demonstrate stability of correct responses because the stimulus-stimulus relations are just being trained. I think it would be helpful if the authors briefly explained why they chose this criterion and if there were any within-block criteria.</p>	<p>There were no other in-block criteria. Given that the participants had separate AB, BC and then mixed training, one block was deemed sufficient. This also ensured short sessions given the young ages of the children, and to attempt to reduce fatigue and maintain engagement.</p>	<p>Not address specifically in manuscript.</p>
<p>I wondered why the authors didn't consult the category training on the mobile device. Given the relative ease with which they were able to train the A, B, and C relations without using prompts, I wasn't sure why they did not do the same for the category training. I think it would be helpful to describe why they made this decision and include some brief discussion of the limitations of this approach. For example, this required an additional equivalence between the digital images and printed images, this required attending to two sets of stimuli (i.e., a person stating the name and the pictures rather than just the mobile device), etc.</p>	<p>Authors: The use of table top was designed to assess generalisation to other settings. This has been clarified on p.3</p> <p>Furthermore, EBI leads to the acquisition of additional untrained relations by design, while conventional classroom instruction through exemplars and rules does not necessarily have this generative outcome.</p> <p>And p. 5 Line 11</p> <p>To ensure generalization from the touchscreen to familiar classroom settings and activities, category sorting was assessed in a table top task prior to any computer-based training. Furthermore, generalization to other, untrained, category members was measured after training and testing with a table top task</p> <p>and p. 14 Line.1</p>	<p>p. 3</p> <p>And p. 5 Line 11</p> <p>p. 14 Line 1</p>

	<p>or to a fourth irrelevant category. This exercise therefore tested for the generalization to untrained and untested stimuli.</p>	
<p>Reviewer 4</p> <p>The authors did a nice job explaining somewhat difficult and complex procedures. However, there are some procedures that should be clarified to increase the technological aspect of the paper.</p> <p>a. What was the purpose of the "break" card (p. 3)?</p>	<p>As described in the visual and textual instructions this acted as a means for the children to request a break, other than verbally requesting one.</p>	<p>p. 6</p>
<p>The use of the token economy, how the tokens were exchanged and for what stimuli, and how the tokens were delivered (and for what) is unclear throughout the method section. I suggest moving all the description of the token economy to the procedures/session description section rather than discussing some of these procedures in the materials section and some in other sections. It will make it easier for the reader to follow what the experimenters did, when, and why.</p>	<p>The authors have moved all of the description to the start of the procedure section. However, reference is still made in individual sections where for example in testing phases no tokens were provided.</p>	<p>p. 4</p>
<p>Why were low to moderately preferred stimuli provided for 1 min after training sessions in which no tokens were earned (p. 4)?</p>	<p>This has been clarified:</p> <p>Access to highly preferred toys was contingent upon performance during training, with each token earning 30 s of play following completion of a block, and, in order to maintain motivation, less preferred toys were presented for 60 s if no tokens had been earned in that block.</p>	<p>p. 4</p>
<p>d. What is the visual instruction board and what was the purpose of this board (p. 5)?</p>	<p>As described on p. 4. A visual instruction boards (described further in the MTS pre-training section) contained written instructions for the computerized experiment accompanied by corresponding pictures.</p> <p>Given that the children were all in pre-school and did</p>	<p>p. 4</p>

	<p>not have reading comprehension this acted as a means to link the instructions with picture samples.</p>	
<p>e. What was the purpose of the green and red symbols? Were the participants told what these symbols meant? Were these symbols presented in all training and testing conditions? (p. 6)</p>	<p>Yes, these symbols were shown and explained in the visual and textual assent form (not included) which were developed for the children. As previously mentioned the children were not of reading age and therefore the use of pictures and descriptors were used to explain the research process to them. E.g. if you match the correct picture – was accompanied by the green symbol. These symbols were present for all training phases only.</p>	<p>p. 6</p>
<p>f. Why wasn't reinforcement provided in testing conditions (p.6)? Also, might this have been a problem (extinction effects)?</p>	<p>Reinforcement is never given in equivalence testing. Such tests consist of probe trials designed to determine if additional relations (e.g., symmetry, transitivity etc) which have not been directly trained can be derived. Offering reinforcement means that the trials would in effect function as training trials and it would not be possible to measure derivation.</p>	<p>p. 6</p>
<p>g. It is unclear how the most-to-least prompt hierarchy was implemented (p. 8). First, it is unclear what the different prompts look liked. Second, it is unclear how the prompts were faded. For example, were they faded within or across sessions? Furthermore, based on the description, what is graphed in Figure 1 is correct responding according to the programmed hierarchy for that trial rather than correct, independent responses. Graphing the latter would be most appropriate for determining acquisition curves regarding the target responses.</p>	<p>The procured has been added to clarify. Initially, participants were not expected to have category knowledge, and a most-to-least prompt regime was put in place to facilitate learning (Libby, Weiss, Bancroft & Ahearn, 2008). The most-to-least prompting hierarchy (MTL) included four prompting levels: full physical - hand-overhand, light physical - light touch or shadow by the elbow, gestural – instructor pointing at correct card and independent responding - no prompt. For each trial, the experimenter noted the level of prompt required. Each block consisted of 12 trials during which each of the three C stimuli were presented four times in a quasi-random order. Prompting was faded across blocks and the criterion for reducing the level of prompt was 11/12</p>	<p>p. 9</p>

	<p>correct trials (92%) over one block. A correct response to a category training trial was recorded if the participant selected the corresponding picture at the designated prompt level, or unprompted within 5 s of the instruction.</p> <p>Corrective action was taken within trial when an incorrect response was made. Incorrect selections were blocked and the experimenter stepped back a level of intrusiveness in the prompt hierarchy while reissuing the instruction. An incorrect response was recorded if a participant selected an incorrect picture card and the experimenter prompted at a level higher than designated, or if the participant made no attempt to select a card within 5 s of the instruction. If a participant achieved less than 6/12 correct trials over three blocks, the experimenter moved back a step in the prompt hierarchy intrusiveness. This phase of the study ended when the participant met the criteria of 11 out of 12 trials correct independently – no prompt, and participants moved to final phase of the study, post-experimental category sort test.</p>	
<p>Although I appreciate the inclusion of the training data and the replication of effects across the different training and testing phases for the different stimuli (Figure 1), it would have been helpful if the experimenters had conducted pre-tests with respect to the different relations that were tested following training. This would have allowed for pre-post comparisons of each of the tested relations.</p>	<p>The participants were tested using the same open sort test for all 27 stimuli.</p> <p>As previously outlined, it not recommended to test for derived relations (symmetry, transitivity and equivalence) prior to training as this may inadvertently provide participants with training/prior knowledge of those relation. Such tests consist of probe trials designed to determine if additional relations (e.g., symmetry, transitivity etc) which have not been directly trained can be derived.</p>	<p style="text-align: center;">n/a</p>
<p>3. It is unclear to me why the additional</p>	<p>This generalization to novel untrained stimuli is of interest.</p>	<p style="text-align: center;">n/a</p>

<p>stimuli were correctly sorted after EBI training (Figure 2). One could look at this in two ways. This could be viewed as a positive result in that stimuli that have certain similarities with trained stimuli came along for the ride. On the other hand, these results may be viewed as a negative result in that the additional stimuli that were not included in any of the training should have served as control stimuli by which to compare the effects of EBI training with test stimuli. Regardless, the authors should spend considerable more time in the discussion section describing these two different views and why these results may have been obtained.</p>	<p>Future research needs to explore possible reasons behind it.</p>	
<p>4. I kept wondering while reading the relatively complex methods of the EBI procedure and reviewing the number of trials necessary whether this type of procedure should be used to teach the type of target skills in the current study, particularly in teaching typically developing preschool children that have good receptive skills. Might providing the participants with some rules regarding the categories and stimuli increasd the efficiency of acquisition and generalization? That is, might an explanation of the category and description of things that fit in that category suffice? In addition, I wondered whether the degree to which a participant did not know what a particular stimulus was (e.g., that a scrunchie was worn in the hair) influenced the results, which could have been addressed with some simple rules. I guess I'm asking whether relations need to be derived in the current way if they can be taught using rules. Again, I understand this area of research is outside of my area of expertise, but with respect to a larger audience, the authors might consider addressing this in their discussion section.</p>	<p>Addressed (see response to Editor's comments)</p>	<p>n/a</p>
<p>5. I suggest the authors provide some discussion of the limitations of their study and expand on pertinent areas of future research in their discussion</p>	<p>Addressed (see response to Editor's comments)</p>	<p>n/a</p>

<p>section.</p>		
<p>Reviewer 5: The second paragraph under “Settings and Materials” is virtually incomprehensible, and of doubtful relevance.</p>	<p>This has been addressed paragraph reworded.</p>	<p>Method: Setting and Materials p. 4</p>
<p>It should be made clearer that “reinforcement” during the MTS phases consisted of tokens which were then cashed in for opportunities for 30s of play. What play? How did the children behave? Seemingly these play interludes were scheduled to occur during a break after a block of 12 trials, which would have taken a minute or more if responding was prompt. I think a more detailed account of all the events and outcomes during MTS should be spelled out so that it could be more easily grasped. This would include the so-called “secondary” reinforcement. (However this was done, the actual performance of the children was seemingly under very good control, judging from the fairly clear account accompanying the Figures)</p>	<p>This has been clarified.</p> <p>Prior to any other testing or training, a two-choice preference assessment (Fisher et al., 1992) was conducted using a variety of developmentally appropriate toys. Pictures of the toys were affixed to a board later presented as choices for participants following an experimental training block. Access to highly preferred toys was contingent upon performance during training, with each token earning 30 s of play following completion of a block, and, in order to maintain motivation, less preferred toys were presented for 60 s if no tokens had been earned in that block.</p> <p>Additionally, all the participants engaged with the researcher using the visual/textual instructions which made specific reference to play during training sessions.</p>	<p>p. 4</p> <p>And instructions p. 5</p>
<p>Prior to MTS the Pre experimental Language Assessment sounds very thorough but I couldn’t find any further mention of it. Presumably these children were entirely competent for their ages. I couldn’t find Table 1 for participant demographics.</p>	<p>Clarification on normative language age has been made. Table 1 was removed due to submission constraints and the information is now provided in text.</p> <p>The standardized and norm referenced assessment of language, the Preschool Language Scale – Fourth Edition (PLS – 4; Zimmerman, Steiner, & Pond, 2005), was used to assess baseline levels of language ability. These assessments showed that the children had language skills closely matching norms for their chronological ages. The ages and PLS-4 age equivalents (each in years: months) were as follows for Participants 1 to 5 respectively: 4:3 and 4:2, 4:6 and 4:3; 4:2 and 3:10, 4:8 and 4:0, 4:2 and 3:10.</p>	<p>Top of p. 4</p>
<p>For the Pre experimental Category Sort Test it might have been helpful to see some example pictures. The children seemingly remembered the locations of the category labels spoken to them. Perhaps it should be made clearer that the experimenters were looking for categories of which the childrens’ choices indicated they seemingly had no knowledge. Presumably for every child</p>	<p>Authors: Additional information has been added to in order to clarify this point. Inclusion of the pictures is not possible due to copyrighting, however the source has been cited and the stimuli name in Table 1 so they may be sourced by the reader.</p> <p>All picture stimuli used for either category sorting, training or computerized phases of the experiment were obtained from Picture This 3.0 Professional Edition (CD©; Silver Lining Multimedia, Inc, 1999).</p>	<p>p. 4</p> <p>and</p> <p>Table 1. p.</p>

<p>at least 3 such categories were found, and these were then used in the MTS phase. I am not quite clear, given that there was a total of nine stimuli for each category, how three of each were allocated to the MTS tests.</p> <p>Overall this is an interesting study which I think should be published. A re-write which steers the reader more clearly through the various stages, explaining their purpose as it goes, and perhaps giving some examples more clearly evocative of the actual conduct of the experiments, would greatly improve it however.</p>		6
<p>A central problem however is the nature of the resemblance between “artificial” stimulus equivalence classes and people’s concepts of real world categories. The individual instances of the latter almost always have some stimulus features in common, unlike the arbitrary assemblages arranged for the former, and unlike also most of the relations between words and their referents. If a pigeon can be taught to recognize oak leaves as a subdivision of leaves in general (Cerella, 1979) the success of the children in this study might be explained without recourse to the concept of stimulus equivalence. It would be like having a number of laboratory equivalence classes in which all of the stimuli in a particular class were given a distinctive colour peculiar to that class. It would have been of interest if an extra, non-natural (ecologically outrageously invalid!) set had been taught in parallel with the natural ones. Would children uniquely fail the usual tests of derived relations specifically with these stimuli only, or perhaps pass?</p>	<p>Authors: Inclusion of the pictures is not possible due to copyrighting, however the source has been cited and the stimuli name in Table 1 so they may be sourced by the reader.</p> <p>All picture stimuli used for either category sorting, training or computerized phases of the experiment were obtained from Picture This 3.0 Professional Edition (CD©; Silver Lining Multimedia, Inc, 1999).</p>	p. 4 and Table 1. p. 6
<p>I think all the actual stimuli used in the children should be presented, even if this requires a coloured figure, or an electronic file with these, and all the data should accompany the publication.</p>	<p>Authors: Inclusion of the pictures is not possible due to copyrighting, however the source has been cited and the stimuli name in Table 1 so they may be sourced by the reader.</p> <p>All picture stimuli used for either category sorting, training or computerized phases of the experiment were</p>	p. 4 and Table 1. p. 6

	obtained from Picture This 3.0 Professional Edition (CD©; Silver Lining Multimedia, Inc, 1999).	
And here's a silly question: were the children able to name the stimuli?	This was not tested in the pilot study but is a question which has been examined in another study by the authors.	n/a
Page 2, 16 analogue not quite the right word – laboratory simulations perhaps	Analogue has been changed to laboratory	p. 2 and p. 16

Introduction:

<u>Revision Suggestion</u>	<u>Modification Made/Comment</u>	<u>Page/Reference</u>
<u>Reviewer 1: Fifth, I think it would be useful for the authors to elaborate on the issues raised at the start of the paragraph referencing the work of Quinn and colleagues (i.e., to provide greater context with the existing literature, broadly speaking).</u>	<u>Further detail on these issues have been added.</u> <u>These issues have been further detailed throughout two paragraphs.</u>	<u>Pg. 2-3</u>

Method:

<u>Revision Suggestion</u>	<u>Modification Made</u>	<u>Page/Reference</u>
<u>Reviewer 1: First, in the methods section, the description of Phase 7 needs to be altered slightly such that it is clearer that AC involves transitivity testing and CA involves symmetry + transitivity (or, equivalence) testing.</u>	<u>Description has been updated.</u> <i>Phase 7. A-C and C-A transitivity and combined symmetry and transitivity (equivalence) tests.</i>	<u>Pg. 9</u>

Results:

<u>Revision Suggestion</u>	<u>Modification Made</u>	<u>Page/Reference</u>
<u>Reviewer 1: Second, Figure 2 in the current version does not show the data it is said to display.</u>	<u>There may be some error perhaps. Figure 2 had not been amended from original submission.</u> <u>New amendment is to the descriptor to include 'equivalence'.</u> <u>Figure 2. Participant 1-5 results for MTS Pre-training, A-B Training, B-A Testing, B-C Training, C-B Testing, Mixed A-B, B-C Training, Mixed B-A, C-B Testing and A-C, C-A Symmetry Transitivity (Equivalence) Testing.</u>	<u>N/A</u> <u>Figure 2 title</u>
<u>Reviewer 1: Third, the description of Figure 3 in the results section still needs to be condensed.</u>	<u>This description has been further condensed.</u>	<u>Pg. 11</u>

Discussion:

<u>Revision Suggestion</u>	<u>Modification Made</u>	<u>Page/Reference</u>
<u>Reviewer 1: Fourth, I think it would be useful in the discussion for the authors to compare and contrast their usage of "transfer of function" and "generalization," given what they did. That is, I think most investigators would describe the successful (Phase 9) sorting of the A and B stimuli as transfer of function in that the function trained to C spread to A and B through the derived stimulus relations.</u>	<u>Attempts to clarify and expand these issues in the discussion have been made.</u>	<u>Pg. 14-15</u>

<p><u>On the other hand, the successful (Phase 9) sorting of the stimuli from the pre-testing seems to involve both transfer of function and an additional (generalization) process. However the authors decide to cast their findings, it certainly is the case that these issues need to be laid out more carefully in the discussion.</u></p>		
--	--	--

Minor Edits:

<u>Revision Suggestion</u>	<u>Modification Made</u>	<u>Page/Reference</u>
<p><u>Reviewer 5: Note there is a reference you forgot to include in your bibliography which, along with Galizio et al, I feel deserve a little more coverage in your account.</u></p> <p><u>Add this to reference list: Libby, M.E., Weiss, J.S., Bancroft, S., Ahearn, W.H., 2008. A Comparison of Most-to-Least and Least-to-Most Prompting on the Acquisition of Solitary Play Skills. Behav. Anal. Pract. 1, 37–43. doi:10.1007/BF03391719</u></p>	<p><u>Reference has been added</u></p>	<p><u>Pg. 18</u></p>
<p><u>please do try to get all the actual stimuli incorporated.</u></p>	<p><u>This is a copyrighted commercial product available for purchase. The stimuli are not permitted for use outside of the research application process. Permission was sought for use of the stimuli as this is copyrighted software, publication of the stimuli would breach copyright law and the agreement made with the developers. However, as outlined in our previous response to Reviewer 5. This is an affordable cd package available for purchase. Picture This 3.0 Professional Edition (CD©; Silver Lining Multimedia, Inc, 1999)</u></p>	<p><u>Pg. 5</u></p>
<p><u>I've interpolated the abstract here for my own convenience. Shouldn't a penultimate sentence be inserted to include the interesting results of the successful extension of correct categorization to virtually all of the untrained stimuli on the second test?</u></p>	<p><u>The abstract has been modified.</u></p> <p><u>All participants demonstrated categorization of the directly trained class members and further generalization to addition unknown stimuli.</u></p>	<p><u>Abstract</u></p>
<p><u>It would be very helpful to see examples of the stimuli here, and to refer to Table 2. On further thought all the stimuli should be available for the reader to view in supplementary materials (perhaps they are) which would facilitate REPLICATION</u></p>	<p><u>As above it is against copyright agreement to publish or make available the actual stimuli. Replication is facilitated through the provision of the stimuli software package. Picture This 3.0 Professional Edition (CD©; Silver Lining Multimedia, Inc,</u></p>	<p><u>Pg. 5</u></p>

	<u>1999)</u>	
<u>Phase 1. A-B baseline training. Participants were directly trained to match the A and B stimuli (A1-B1, A2-B2, and A3-B3)(for their individual stimuli see Table 2) Reviewer 5: comment, I think this has to be made crystal clear</u>	<u>This has been amended.</u> <u>Phase 1. A-B baseline training. For their individual stimuli see Table 2.</u>	<u>Pg. 7</u>
<u>An incorrect response was recorded (a) if a participant selected an incorrect picture card and the experimenter ?prompted at a level higher than designated, or (b) if the participant made no attempt</u> <u>Reviewer 5 Comment: Have I got the sense of this with my a and b?</u> <u>Before prompted: Should “thereafter” or “had” be inserted here, before “prompted”?</u>	<u>This has been amended</u> <u>An incorrect response was recorded if a) a participant selected an incorrect picture card and the experimenter had prompted at a level higher than designated, or b) if the participant made no attempt to select a card within 5 s of the instruction.</u>	<u>Pg. 5</u>
<u>This phase of the study ended when the participant met the criteria of 11 out of 12 trials correct independently – with no prompt, and participants were then moved to the final phase of the study, post-experimental category sort test.</u> <u>Reviewer 5 Comment: So this was equivalent to establishing each C stimulus as a discriminative stimulus, as in functional equivalence?</u>	<u>Essentially, that the auditory taxonomy label or category name would transfer not from the directly trained C stimulus to the other members trained during MTS. Further generalized to previously known and unknown class members was then tested.</u>	<u>N/A</u>
<u>Post-experimental category sort test. A category sort test was used to test for the generalized derived transfer of the categories trained in Phase 8. This phase involved the participants sorting the stimuli that had been tested pre-experimentally. It included the three stimuli chosen for training during the experiment and an additional six from each of the three categories 27 stimuli in total, to one of the three categories established for the C stimuli or to a fourth irrelevant category. This exercise therefore tested for the generalization to untrained and untested stimuli. The procedure and the instructions used were the same as in the pre-experimental category sort test, each of the stimuli were presented once. Reviewer 5 Comment: This is interesting. Some untaught comprehension of semantic categories is being sought, notwithstanding that this was seemingly ruled out before training with other stimuli from that category?</u>	<u>Yes, essentially, that the auditory taxonomy label or category name would transfer not from the directly trained C stimulus to the other members trained during MTS. Further generalized to previously known and unknown class members was then tested.</u>	<u>N/A</u>
<u>Post-Experimental Category Sort Testing</u> <u>The results of the post-experimental sort test for all</u>	<u>Three figures were uploaded in the resubmission of the manuscript, there may have been an error somewhere. Pre</u>	<u>Figures 1, 2 and 3.</u>

<p>participants can be seen in Figure 3. Nine stimuli were used for each of the three categories (27 stimuli in total). These consisted of the three stimuli in each category that had been trained and tested in the MTS procedure, and an additional six category members that not been used in the intervention but that had been tested in the pre-experimental category sort test.</p> <p><u>Reviewer 5: I am unclear since I have 2 figures, a Fig.2 which distinguishes 3 components in each column, including a middle “pretest” components, and a Fig 3 where the columns are of identical height and pre-test and “additional” untrained stimuli are seemingly lumped together. Was Fig.2 and earlier and inaccurate graph, since all the untrained stimuli had been tried out earlier, or is there some other reason for not distinguishing those that were categorized correctly both before and after??</u></p>	<p><u>and post category sort tests were separated as Figure 1. pretest and Figure 2. posttest.</u></p>	
<p><u>As a result of this pre-experimental pairing, it was expected that category membership would generalize from the stimuli chosen for training to those that had been paired pre-experimentally but not trained (Galizio et al., 2004). Indeed, at post-experimental testing all participants did correctly sort additional category members which had been tested pre-experimentally but had not participated in the equivalence classes.</u></p> <p><u>Reviewer 5: More could perhaps be made of this interestingly relevant Galizio et al study</u></p>	<p><u>Expansion has been made to the introduction and discussion section to explore these issue more in-depth. Reference to Roche & Mervin Prototype theory and typicality effects as seen in stimulus discrimination and generalization has been included in order to give a more comprehensive overview of these issues.</u></p>	<p><u>Introduction</u> <u>Pg. 2-3</u></p> <p><u>Discussion</u> <u>Pg. 14-15</u></p>
<p><u>However, the current methodology uniquely included an additional aspect whereby the category name was receptively trained to the C stimuli.</u></p> <p><u>Reviewer 5: The little word “name” perhaps deserves more emphasis, and comment. A reference to Horne, P.J., Lowe, C.F., 1996. On the origins of naming and other symbolic behavior. J. Exp. Anal. Behav. 65, 183–353, might be a good idea.</u></p>	<p><u>Naming theory has been briefly introduced in the introduction and a paragraph has been added to the discussion section.</u></p>	<p><u>Introduction</u> <u>pg. 1</u></p> <p><u>Discussion pg.</u> <u>16</u></p>
<p><u>Minor edits had been suggested by Reviewer 5 in terms of grammar. Sentence structuring.</u></p>	<p><u>These suggestions have been incorporated into the reviewed manuscript.</u></p>	<p><u>Throughout manuscript</u></p>