MASTER'S THESIS
Master in Learning in Complex Systems
June 2018

Stimulus Equivalence and Instructions

Annika Poulsen

Faculty of Health Sciences
Department of Behavioral Sciences

OsloMet – Oslo Metropolitan University

Acknowledgements

First, I would like to thank my supervisor Professor Erik Arntzen for helping me planning the experiment and providing me with helpful comments on the text. I would also show my gratitude to former and current members of the Experimental Studies of Complex Human Behavior Research Group. The comments and discussions have been an important contribution to this work. Also, thank you Aleksander, for your patience, support, and helpful discussions during data collection and writing. Finally, thank you Max, your curiosity about life, in all of its aspects, reminds me of why I have science close to heart.

Table of Contents

Overview of Tables and Figures

Article 1

The Investigation of Instructions in Stimulus Equivalence Research: A Literature Review

Article 2

Effects of Instructions on Training of Conditional Discriminations and the Formation of Equivalence Classes

The Investigation of Instructions in Stimulus Equivalence Research: A Literature Review

Annika Poulsen

Oslo Metropolitan University

Abstract

The current review aimed to investigate different type of instructions that have been used in stimulus equivalence research in the time periods 1988–1992 (i.e., first time period) and 2013–2017 (i.e., second time period). Instructions were categorized as either general instructions (GEN), specific instructions (SPEC) or not task specific instructions (NTS). The main findings which covered 118 experiments showed that the most commonly used instruction during stimulus equivalence experiments was NTS instructions for both time periods with an increase in the second time period (from 48,84% to 60,53%). It was also found a decrease in GEN instructions from the first to the second time period (from 41,86% to 28,95%). The present review highlights that there is much variability in the parameters used in stimulus equivalence research. Even if this might indicate that stimulus equivalence is a robust phenomenon, it can also lead to difficulties when interpreting results across studies. Thus, the finding shows that the usage of instructions varies and that future researchers might consider a more parsimonious approach to instructions when conducting stimulus equivalence research.

Keywords. general instructions, specific instructions, not task specific instructions, stimulus equivalence, matching-to-sample

Stimulus equivalence has for over 40 years been a central and active research field within behavior analysis (Arntzen, 2010). The increased attention over the years may be explained by the fact that stimulus equivalence relations have been used to describe many forms of complex human behavior (e.g., Fields, Arntzen, & Moksness, 2014). Subsequently, the equivalence research field have had an extensive amount of publications (Pilgrim, 2016) and given important contributions to societal questions that we can connect to areas such as memory, language and problem solving. Studies have, for example, shown how equivalence classes can impact everyday life such as in teaching sequences connected to reading in the educational system, or the learning of concepts for individuals with severe learning disabilities (Sidman, 1994).

Conducting a quick citation search, 9th of June 2018, at Google Scholar citing "*stimulus equivalence*" between the years 1988–1992 and 2013–2017 yielded a search result of 347 and 1620 hits, respectively. Considering the increased citations and published articles over the last five years, a literature review ought to be of great interest.

In the behavior analytic field it has been conducted many reviews that have reported central and useful information. Two examples are the citation analyses by McPherson, Bonem, Green, and Osborne (1984) and Dymond, O'Hora, Whelan, and O'Donovan (2006) which investigated the impact of Skinner's (1957) *Verbal Behavior* by assessing number of articles and the prevalence of verbal operants in the behavior analytic literature. Results showed, among other things, that *Verbal Behavior* still "make an important contribution to the psychological literature and that the number of citations have remained steady across the past 20 years" (Dymond et al., 2006, p. 83). Conducting summaries of literatures therefore seems like a good approach to ensure that relevant findings reaches a wider audience (Pautasso, 2013).

So far, there are to the author's knowledge no conducted reviews within the stimulus equivalence research field investigating the scope of instructions. The objective of the present article was, therefore, to investigate the phenomenon of stimulus equivalence and assess the use of instructions under experimental settings with humans as participants.

Stimulus equivalence can as Green and Saunders (1998) wrote broadly be defined as stimulus substitutability, that is, treating unrelated stimuli as if they were the same. This substitutability is most often shown experimentally with the use of a matching to sample (MTS) format. In MTS procedures conditional discriminations are first trained (four-term contingencies), followed by a test for emergent properties. For example, if a young child learns to point at picture of a cat and not a picture of a dog when the mother says cat while looking in a picture book, next learn to point at the correct printed text word [cat] and not the text word [dog] when the mother says cat, and afterwards point to the correct printed text word [cat] when a picture of a cat and a dog is presented, it can be concluded that the stimulus class "cat", with its three members, has been established.

Sidman and Tailby (1982) defined stimulus equivalence as responding in accordance with the properties of reflexivity, symmetry, and transitivity. The definition is inspired by mathematical set theory and all properties must be shown to conclude that an individual is responding in accordance with stimulus equivalence (Dickins, 2015; Sidman, 1994). To better illustrate the phenomenon of stimulus equivalence a coding system can be used. For example, in a stimulus equivalence experiment with three 3-member stimulus classes, the classes can be referred to as class 1, 2, and 3. Furthermore, the stimuli can be coded by the use of letters (e.g., A, B, and C). Thus, the A stimulus in the first class can be referred to as A1, the B stimulus as B1, and the C stimulus as C1. The same coding system goes for the other classes (e.g., A2, B2, C2, A3, B3 and C3).

The property of reflexivity means that the stimulus must have a conditional relation to itself. Hence, if training A to B and B to C, then if an individual selects A in the presence of A, B in the presence of B and C in the presence of C it can be concluded that the individual responded in accordance with reflexivity. The property of symmetry refers to a bidirectionality among two stimuli. When training A to B and B to C, a successful test of symmetry would be that the individual selects B in the presence of A and C in the presence of B. Further, to show the property of transitivity after training A to B and B to C, the individual selects A in the presence of C (Green & Saunders, 1998). Training and testing for stimulus equivalence can, therefore, demonstrate how an individual relates stimuli in new ways, leading to the emergence of novel behavior without direct reinforcement (Sidman, 1992). Hence, when it is said that an individual has responded in accordance with stimulus equivalence, it is reinforcement alone, of a few trained relations, that has led to the emergence of new behavior (Sidman, 2000).

Stimulus equivalence has shown to be a robust phenomenon, since responding in accordance with stimulus equivalence have been found across populations and different parameters. The phenomenon is being studied within both experimental (e.g., Sidman, 1994) and applied research (e.g., Arntzen, Halstadtro, Bjerke, Wittner, & Kristiansen, 2014; Sidman, 1971). However, comparing the experimental and applied research approaches, it appears that the growing number of empirical studies have been published in experimental research (Arntzen, 2012). To understand what types of conditions that can make individuals respond in accordance with stimulus equivalence, have researchers primary been interested in investigating parameters that can affect the outcome, such as training and testing protocols (e.g., Arntzen, Grondahl, & Eilifsen, 2010; Imam, 2006), class size (e.g., Johnson, Meleshkevich, & Dube, 2014) and class members (e.g., R. R. Saunders, Wachter, & Spradlin, 1988), and meaningfulness of stimuli (e.g., Arntzen & Lian, 2010; Fields, Arntzen, Nartey, &

Eilifsen, 2012). Until present date, as mentioned earlier, systematical summaries of parameters that affect stimulus equivalence responding are to the author's knowledge not existing. However, gathering an overview of empirical evidence, assessing parameters and review the scope of all or some of them can give a better understanding of the phenomenon.

A variable that has been discussed but not investigated to a great extent experimentally within the stimulus equivalence field is instructions. One of the first conducted experiments on this topic was published by Green, Sigurdardottir, and Saunders (1991). The researchers manipulated this variable in two different experiments, where typical adults as participants either got full or minimal information about the MTS task and training contingencies. Another study by K. J. Saunders, Saunders, Williams, and Spradlin (1993) investigated different types of instruction with atypical children as participants. The few studies conducted with instructions in modern time has mainly investigated the establishment of conditional discrimination with typical children as participants (Arntzen, Vaidya, & Halstadtro, 2008; Pilgrim, Jackson, & Galizio, 2000).

When viewing the early conducted stimulus equivalence research of Sidman (Bush, Sidman, & de Rose, 1989; Sidman, Wynne, Maguire, & Barnes, 1989), one can notice that the implementation of instructions in the procedures often was limited and kept to a minimum. An explanation for this may be that the result of the effects of instruction on stimulus equivalence research is, accordingly to Sidman (1992, 1994), seemingly apparent with regards to a person's learning history. Sidman (1992) stated for example that:

> "…until we have answered the question of whether rules give rise to equivalence, or equivalence makes rules possible, we are going to have to be careful about our experimental procedures in investigations of stimulus equivalence. If we tell our subjects that stimuli 'go with' each other (or that they 'match each other,' 'belong together,' 'are the same,' 'go first' or 'go second' etc.), the data mat then tell more

about the subject's verbal history than about the effects of current experimental operations." (pp. 21–21).

Thus, if a participant has a history with an instruction such as *goes with*, and if that instruction is used in an experiment, it may enough for having the participant to respond in accordance with stimulus equivalence (Sidman, 1992). The participant might then become less sensitive to the contingencies set up by the experimenter as a result of instruction-following (Catania, 2013). However, it is important to note that even if instructions can help an individual to respond in accordance with stimulus equivalence, it is not sure if instructions is a necessary variable (Sidman, 1992). Hence, not using instructions at all in equivalence procedures or being careful with instructions gives researchers reason to believe that reinforcement alone during training conditions is sufficient for the individuals' performance (i.e., selecting correct comparison stimuli) during testing (Sidman & Tailby, 1982).

Studies like those by Arntzen et al. (2008), Green et al. (1991) and K. J. Saunders et al. (1993) and discussions concerning instructions as presented by these authors and Sidman (1992, 1994) are important in two ways. First, problems can arise in the context of interpreting the results of experiments (Dymond & Rehfeldt, 2000). That is, investigating the same variables when different kinds of instructions (i.e., very or little informative) are presented to the participants across studies. Second, it can be argued that there are two different ways of establishing equivalence relations. Either by following the contingency shaped procedure (e.g., MTS format) as proposed by Sidman or by using instructions that potentially can influence the behavior of participants during the stimulus equivalence experiment, indicating rule-governed behavior (de Rose, McIlvane, Dube, Galpin, & Stoddard, 1988; Silveira, Mackay, & de Rose, 2017). Hence, investigating the choice of instruction in stimulus equivalence research would give an overview of these two procedures, reflect the sensitivity to the contingencies and what type of equivalence responding (i.e., contingency shaped or rule-governed) that will be shown in the results.

The primary focus of this review was to shred light of the extent of instructions and to systematically investigate if instructions differ by looking at two different time periods. The assessment started with the time period 1988–1992 and was followed by a second time period in 2013–2017. With almost 25 years apart it is plausible to say that the carefulness may have changed. Sidman's concerns were not disseminated until the beginning of the 90s (Sidman, 1992, 1994). Namely, it can be hypothesized that the majority of conducted studies in the late eighties did not consider the issue with instructions to the same extent. On the contrary, it has been a long time since the topic has been discussed, and with regards to contemporary conducted studies, the considerations of using instructions in stimulus equivalence research may have been forgotten.

Method

Qualification Criteria

To be included in the present review the articles had to meet criteria such as being published in peer-reviewed journal between 1988–1992 (first time period) and 2013–2017 (second time period), written in English and, include the search terms *matching-to-sample or matching to sample* and *stimulus equivalence*. The search excluded book chapters and other paper and poster presentations.

Search Strategy

The search was limited to the electronic websites of the following journals; *Journal of the Experimental Analysis of Behavior* (JEAB), *Journal of Applied Behavior Analysis* (JABA), *The Psychological Record* (TPR), *Learning & Behavior* (LB), *European Journal of Behavior Analysis* (EJOBA), *Behavioural Processes* (BP), *the Quarterly Journal of Experimental Psychology* (QJEP), and *Psicologia: Relexão e Crítica* (PRC). The aforementioned journals were included in the analysis because of their central role in the behavior analytic field. The Experimental Analysis of Human Behavior Bulletin was excluded from the search since the journal was canceled for nine months, and no articles were to be retrieved at the date of search and at the time of writing.

Three types of searching strategies identified articles. The first strategy was to use the journals search engine and search for the aforementioned search terms. Afterwards, a limitation was made to find articles published between the predetermined years. If the journal did not have a functioning search engine, the second strategy was to manually go through each issue and look for the search terms in the heading, abstract or keywords in the articles. Due to changes with regard to publishers for the TPR journal in the time period 1988–1992, was Google Scholar used as a third source.

The search for 1988–1992 was conducted January 6th, 2018 for JEAB, JABA, TPR and LB. The search for the remaining journals in 1988–1992: BP, QJEP and PRC was conducted January 7th, 2018. In the time period 2013–2017 was the search conducted December 27th, 2017 for the journals EJOBA and JEAB, December, 28th, 2017 for JABA and December 29th, 2017 for LB, BP, QJEP and PRC.

Screening and Selection of Articles

All full-texted articles, including potentially eligible articles, which contained the search terms were retrieved. In the screening, a reviewer assessed the articles by looking for stimulus equivalence studies containing instructions about stimulus equivalence and the MTS format. Articles were selected if they met the search criteria. Articles that did not include an instruction, that used other training- and testing procedures than MTS, did not test during extinction conditions or not tested stimulus equivalence properties were excluded from the review. Articles that were considered as equivalence based instructions (EBI) were also excluded due to the difficulties with including such studies using instructions as stimuli (e.g., Reyes-Giordano & Fienup, 2015). If articles contained more than one experiment and matched the inclusion criteria, the experiments were considered as separate studies. This meant that the number of experiments, therefore, would be higher than the total number of articles found in the citation search. For instance, if an article contained two experiments with changes in the parameters, such as using different types of participants (i.e., typical and atypical developed), it was to be considered as two separate experiments and plotted accordingly.

Uncertainty about the eligibility of articles was discussed and resolved by a second screening, using a second reviewer with behavior analytic background and knowledge about stimulus equivalence. The second screening was carried out after the first assessor had gone through all articles (see reliability further down).

Scoring Information and Categorization

Of the selected articles, the assessor scored information such as authors names, article title, year of publication, type of journal, type of instruction and experimental parameters (see subcategorization) in a standardized form. The assessor started by scoring general article information and then categorizing the instruction as either general instruction (GEN), specific instruction (SPEC), or not task-specific instruction (NTS). Articles with GEN instructions included at least one citation where a directive such as *goes with*, *matches*, *belongs together* were used. For example, Doughty and Best (2017) investigated transformation of function by training and testing with a MTS format by giving the instruction “click on the mouse over the additional item that you think *goes with* the one in the center” (p.5). Instructions that gave away specific information about the stimulus relations in the MTS format such as “when this stimulus appears, choose this one” were to be considered as specific. The specific instructions could also include stimuli with pre-determined names by the experimenter or participants and include general instructions. de Almeida and de Rose (2015) provided an example of specific instruction by informing the participants “when this picture is here, …. choose this” (p. 454) by pointing to the correct comparison stimulus in a study examining meaningfulness of abstract stimuli in a delayed matching-to-sample procedure. The remaining articles including instructions that were considered as not having an implication on MTS performance were categorized and assigned to the NTS instruction group. For example, Arntzen and Vie (2013) used the instruction “a stimulus will appear in the middle of the screen….three other stimuli will then appear. Choose one of these…” when investigating the effects of distractors on equivalence responding.

Next, the categorizing continued with trying to classify the articles as either experimental, applied or translational research, where the experimental studies had the primary purpose to investigate the phenomenon of stimulus equivalence. An example of an article that qualified as an experimental was the study of Lian and Arntzen (2013) who investigated training structures with a delayed matching-to-sample procedure. Categorizing articles as applied proved to be difficult because these studies can be considered as translational by investigating social significant behavior and answering experimental questions (e.g., Critchfield, 2011). Such example was demonstrated in the study by Cowley, Green, and Braunling-McMorrow (1992) who trained persons with brain injuries to match dictated names to photos of their therapists and later tested for new and untrained relations between those photos and written names of the therapists.

The analysis of the experiments continued by assessing parameters in the participant demography category. The subcategorization covered information such the gender (female, male or mixed female and male) type (i.e., typical or atypical developed) and age (i.e., adults or children), where a label or description of participants functioning in the articles decided if they were considered as atypical. All participants under the age of 18 years were defined as children. Information about the presentation of instructions, the use of computer or table top, training structures and training and testing protocols were also assessed during the subcategorization.

Reliability

Inter-rater agreement (IOA) was carried out on 20% of the experiments and randomly picked out by the first assessor. The IOA was taken on the assignment of instructions and type of research. An agreement was defined as when both assessors assigned the articles to an identical type category. The percentage of agreement was calculated by dividing the number of agreements by the total sum of agreements and disagreements and then multiplying it with 100. The inter-rater agreement resulted in 90,9% for the instruction category and 81,8% for the type of research category.

Random samples were also taken in both time periods for the included journals (the year 1991 and 2015 in the first and second time period) to strengthen the reliability for the conducted search. The samples were manually reviewed in the journals electronically browsing volume and issue catalogue by the first reviewer with the purpose of ensuring that the original search had covered all relevant articles. The control, which meant that the reviewer assessed all issues for each respective years, showed that the chosen method for searching articles in this present review and manually going through each issue had a 90,1% correspondence.

Results

The search identified a total of 265 articles (Figure 1) within the two time periods 1988–1992 (first time period) and 2013–2017 (second time period); JEAB *n* = 96, JABA *n* = 33, TPR *n* = 75, LB *n* = 19, EJOBA *n* = 24, *BP* *n* = 14, QJEP *n* = 3, and PRC *n* = 1. After applying the inclusion criteria were 177 articles excluded, consequently, 147 articles remained for screening. Of the remaining articles were 37 of them assigned the first time period and 110 articles the second time period. Instructions were described in 88 articles or

118 experiments (an article could include more than one experiment) of the 147 articles. Twenty-one articles (4,5%) did not describe or mention instructions; these articles were part of the 177 excluded articles. Further, seven experiments investigated other phenomena than stimulus equivalence. In these seven experiments the authors pointed out that the participant had to pass the equivalence testing. Of these experiments 43% were given GEN instructions, 28,5% SPEC instructions, and 28,5% the NTS instruction. The remaining analysis will further focus on conducted experiments.

Of the 118 experiments 90,5% were categorized as experimental research, 7% as translational and 2,5% as applied. Further, the assessment assigned the experiments to one of the three groups independent of time periods (Figure 2); GEN instructions (n = 40; 33,90%), SPEC instructions (n = 12; 10,17%), and NTS instruction (n = 66; 55,93%). The experiments were next categorized into the two aforementioned time periods.

Assessed parameters from the experiments were after that subcategorized into different categories and analyzed for to the time periods. Table 2 shows an overview of this analysis by presenting the prevalence of the type of parameter and its specification in the two separate time periods and the combination of these two. Further, Table 3 shows a detailed analysis of both time periods with respect to the use of the specific type of instruction for selected parameters.

Distribution of Instructions in the First and Second Time Period

The percent distribution of different type of instructions is shown in Figure 3. Forty-two of the 118 experiments describing instructions were assigned the first time period and then subcategorized into either the GEN instruction group (n = 18; 42,86%), the SPEC instruction group (n = 4; 9,52%) or the NTS instruction group (n = 21; 47,62%). The results of the remaining 76 experiments in the second time period showed that 22 experiments

(28,95%) were assigned the GEN instructions group, eight experiments (10,53%) the SPEC instruction group, and 46 experiments (60,53%) the NTS instruction group.

As for mentioning instructions during stimulus equivalence research, the results in Figure 3 show that there has been a decrease in the GEN instruction with 13,91% from the first to the second time period. A minor change was seen with regards to the use of SPEC instructions with a 1,23% decrease from the first to the second time period. The NTS instruction increased with 12,91% from the first to the second time period.

Presentation of Instructions

An analysis of the presentation of instructions in the first time period showed that 17% of the instructions were presented on screen, 64% told, 12% written, and 2,5% modified, that is, using combinations of different presentations. The results after analyzing the three experimental groups (Table 3) showed that 5,5% were presented on screen, 78% told, and 16,5% written in the GEN group (n = 18). For the SPEC group (n = 4) were 25% presented on screen, 50% told and, 25% modified. Lastly, for the NTS instruction group (n = 20) were 25% presented on screen, 55% told, 10% written down, and 10 % not specified.

Conversely, the analysis for the second time period showed (Table 2) that 62% of the experiments used screen presentation, 22% told, 5% written, 3% modified, and 8% not specified. This show a significant change in the presentation of instruction with regards to screen presentation (from 17% to 62%) versus told (from 64% to 22%), depending on the given time periods. The result from the detailed analysis (Table 3) of the three instruction groups showed that 36% used screen presentation, 41% were told, 5 % written, and 18% not specified in the GEN group (n = 22). Thirty-seven point five percent of the instructions in the SPEC instruction group (n = 8), were presented on screen, 37,5% told, 12,5% written, and 12,5% not specified. In the NTS instruction group (n = 46) were 78% of the instructions presented on screen, 11% told, 4,5% written, 4,5% not specified, and 2% modified.

Of the 42 experiments in the first time period 67% of the experiments use a computer and 31% a table top in the matching-to-sample procedure. Two percent of the experiments did not specify the use of the author's knowledge. The results for the second time period (*n* = 76) showed that 87% used a computer and 8% a table top, 5 % used a combination of computer and table top. the findings show an increase of 20% with regard to the use of computer and a 23% decrease with respect to the use of table top from the first to the second time period.

Population Demography Analysis

The overall analysis of demography of populations for all included experiments (*n* = 118) showed that 68,5% of the participants were students, 18% adults, 12% children, and 1,5% a combination of adults and children. Of the assessed experiments did 7% include females, 10% males and 56% include a mix of female and male participants. Twenty-seven percent of the experiments did not specify participants genders. The distribution of gender in both time periods showed an increase with mixed genders (both female and male) as participants from the first (48%) to the second time period (60,5%). Also, the inclusion of only females or males as participants decreased from 12% to 4% for females and 14% to 8% for males from the first to the second time period.

In the first time period 66,5% of the participants were students, 24% adults, and 9,5% children. Seventy percent students, 14% adults, 13% children, and 3% combination of adults and children in the second time period, respectively. An analysis of the GEN instructions and the NTS instruction group in both time periods showed that the use of students were equally distributed (i.e., 67% in the first and 68% in the second time period for the GEN instruction; 75% in the first and 74% in the second time period for the NTS instruction). There were however, differences when children were included as participants for two instructional groups, across the two time periods. The results showed that there was an increase in the participation of children with GEN instructions from the first to the second time period (11%

to 32%). Similar results were seen in the SPEC instructions group (0% to 37,5%). With adults as participants was a decrease seen with the use of both GEN and SPEC instruction from the first to the second time period, 22% to 0% and 75% to 12,5%, respectively. There was, however, an increase in the usage of SPEC instruction and students as participants from 25% to 50%, and the NTS instruction with adults as participants, from 15% to 22% from the first to the second time period. NTS instruction were used with combined adults and children as participants in 4% of the experiments in the second time period.

The overall analysis (Table 2) of the experiments showed that the majority of participants included were typical developed students (68,5%) or adults (13,5%). In 7% of the experiments were typically developed children included as participants. Nevertheless, 4% of the experiments included both atypical developed adults and 4% atypical developed children. In the remaining experiments were participants were mixed with combinations of children and adults both typical and atypical developed (1-2%).

A closer analysis of GEN instructions (n = 18) in the first time period (Table 3) shows that 67% of the experiments included typical students, 16,5% typical adults, 11% atypical children, and 5,5% atypical adults. For the SPEC instructions (n = 4) included the experiments 75% atypical adults and 25% typical students. The distribution of NTS instruction (n = 20) included 75% typical students, 15% typical adults, 5% typical children, and 5% atypical children as participants.

An analysis of the second time period and GEN instructions (n = 22) showed a result of 68% with typical students, 23% with atypical children, 4,5 % typical children, and 4,5% children mixed atypical and typical developed as participants. With SPEC instructions (n = 8) in the same time period included 50% typical students, 37,5% typical children, and 12,5% atypical adults as participants. Seventy-four percent were typical students, 22% typical adults, and 4% adults and children with mixed atypical and typical participants in the NTS instruction group.

These main results show that the use of GEN instructions remained unchanged when typical students were included as participants. However, when atypical children were included as participants GEN instructions increase with 12% from the first to the second time period. Furthermore, the SPEC instruction was more commonly used in the first time period with atypical adults as participants than in the second time period where participants such as typical students and children got this kind of information prior the experimental session.

Other Parameters

Training structures. The most commonly used training structure (Table 2) in all experiments was the One to Many (OTM) training structure (41,5%) followed by the Linear Series (LS) training structure (37,5%). The Many to One (MTO) training structure were used in 9 % of the experiments and remained unchanged between the two time periods. Twelve percent used combinations of training structures or modified training structures. The analysis of the two time periods showed that there was a 35,5% decrease for the OTM training structure (from 64,5% in the first time period to 29% in the second time period). However, the use of LS increased with 47% from the first to the second time period (from 7% to 54%).

Protocols. Further, the Simultaneous (SIM) training and testing protocol was the most commonly used (66%), followed by the Simple to Complex (STC) protocol (18%), and lastly the Complex to Simple (CTS) protocol (8%) when analyzing all experiments independent of time periods (Table 2). The remaining experiments (8%) which only tested for one equivalence property were assigned to a separate category. A 36% increase was shown with the SIM protocol (from 43% in the first to 79% in the second time period). Moreover, decreases were shown as for the rest of protocols from the first to the second time period; 9,5

% for the STC protocol, 8,5% in the CTS protocol, and 18% in those experiments which only tested for one equivalence property.

Combination of training structures and protocols. Since some combinations of training structures and protocols can influence the participants' performance in training and testing for stimulus equivalence (e.g., Arntzen, 2012), an additional analysis investigated how these two types of parameters were combined with GEN, SPEC and NTS instructions. Noteworthy results (Table 3) in the second time period showed that GEN instructions were given in 50% of the experiments when combining the LS training structure and SIM protocol, 36% when combining the OTM training structure and SIM protocol. As for the use of specific instruction, in the same time period, was the most commonly used combination the LS training structure and STC protocol, and the OTM training structure and the SIM protocol with a 25% distribution each. When including the NTS instruction in the experiments showed the analysis a 39% usage of the combination of the LS training structure and SIM protocol, and 22% for the OTM training structure and SIM protocol. As for the first time period and the usage of GEN instruction was the most used combination the OTM training structure and SIM protocol (33%). For the SPEC instructions was a mix of all training structures and a mix protocols the most commonly used (75%). The results for the NTS instruction showed that the most commonly used combination was the OTM training structure and SIM protocol, followed by a mix of training structures and protocols (25%).

Location of Stimulus Equivalence Research

The worldwide locations of stimulus equivalence articles from the present literature review search are shown in Figure 4. The upper panel shows the number of articles in the first time period and the lower panel the second time period. The categorization was based on the affiliation of the first author. Of the total of 28 articles in the first time period were 26 of them originated from the United States of America, one from Brazil and one from Ireland. Of the

60 articles in the second time period were 17 articles originated from the United States of America, 16 from Norway, 11 from Brazil, seven from Ireland, two from New Zealand, two from the UK, one from Colombia, Northern Ireland, Netherlands and Spain, respectively.

Discussion

The main purpose of the current literature review was to investigate types of instructions given during stimulus equivalence research between the time periods 1988–1992 and 2013–2017. Instructions were either categorized as GEN, SPEC or NTS instructions. The primary results (Figure 3) showed that the most commonly used instruction for both time periods was the NTS instruction, followed by the GEN instruction. The SPEC instruction was the least used instruction where the occurrence remained unchanged in both time periods. For the 21 articles which were eligible in the screening but did not mention instructions, it is plausible to think that the group of participants chosen were not responsive to instructions (de Rose, McIlvane, Dube, & Stoddard, 1988). Other procedures might, therefore, have been implemented to show how the task was built up.

As for the NTS instruction a 14% increase was found in usage from the first to the second time period. This might indicate that researchers take a more careful approach regarding instructions used in equivalence research today. Additionally, the use of GEN instructions turned out to be higher in the first time period (42,86%) compared to the second time period (28,95%). It may be the case that the reduction of GEN instructions in the second time period can be accounted for by Sidman's (1992, 1994) concerns regarding instructions and equivalence research. However, even if the use of GEN instructions decreased from the first to the second time period, GEN instructions is still often used. This may be credited to the type of research conducted (i.e., applied or translational), choice of participants (i.e., typical or atypical developed) or if the purpose of the study was to investigate phenomena that require participants to form equivalence classes. Since, in these type of studies, it can be beneficial to include instructions that can facilitate the establishment of baseline relations and increase the likelihood of participants responding in accordance with stimulus equivalence. For example, findings showed that when analyzing participant demography that instructions categorized as GEN or SPEC often were used in studies including children as participants. Giving information such as *goes with* or specifying names of the stimuli included in the experiment may be considered as alternatives, as for making the task easier. Previous research with typically developing children has, for example, shown that these type of participants may have difficulties to acquire conditional discriminations and to respond in accordance with stimulus equivalence (see Arntzen et al., 2008; Pilgrim et al., 2000). The issue with difficulties of acquiring stimulus class formation should also be reasonable to argue for, with respect to atypical developed participants. These participants may have similar difficulties under training and testing for equivalence relations (e.g., Kenny, Barnes-Holmes, & Stewart, 2014; Zygmont, Lazar, Dube, & Mcilvane, 1992) and would therefore benefit, if possible, from instructions that can facilitate the task, which also was confirmed from the results of this present review.

Procedure Materials

The analysis further revealed that there was a change in the delivery of instructions between both time periods. In the first time period, the majority of instructions were told in contrary to the second time period where the majority of the instructions were presented on screen. This might be explained by the decrease in the table top and increase in usage of computers when comparing the two time periods.

Training Structures and Protocols

It was also showed that OTM and SIM were the most commonly used training structure and protocol. Furthermore, a more refined analysis when combining training structures and protocols revealed that 50% of the experiments combined the LS training structure, SIM protocol and GEN instructions in the second time period. This is an interesting finding, since, the combination of LS training structure and SIM protocol have shown to produce lower yields of equivalence responding, compared to other combinations (Arntzen, Norbom, & Fields, 2015; Imam & Warner, 2014). It can be proposed that LS and SIM in combination is a good choice when one wants to test sensitive parameters which other training structures or protocols, in combination, would not have been able to show. For example, the MTO training structure has shown to produce a high yield of equivalence class formation (e.g., Arntzen et al., 2010), indicating that this type of training structure could lead to ceiling effects when investigating sensitive parameters. Hence, not show small effects at all. Further, the finding raises interesting questions about why the combination of GEN instruction, LS training structure and SIM protocol have been used. The first apparent answer would be that the purpose of these studies was not to investigate the emergence of stimulus equivalence formation. A second answer could be the choice of participants included in the experiment, as earlier discussed or third, that the experimenters have overlooked this issue.

Type of Participants

The results showed that the majority of the experiments used students as participants. This might be accounted for that most stimulus equivalence research, as shown in the current review, are experimental research which often is carried out at universities. A small decrease in the use of adults as participants was also found, there is, however, the possibility that adults as participants in the second time period have fallen into the category of students.

A small increase was found from the first to the second time periods with respect to children as participants. This finding can be viewed as a positive trend in the application of equivalence procedures since there is need for more documented research for this population in both the applied and translational field. Hence, future research should focus more on social significant behavior (Sidman, 1994) and the usage of efficient teaching techniques based on finding from the laboratories. These techniques do not have to be as stringent as in the experimental research, where the purpose is to get a better understanding of the phenomenon by manipulation of different parameters. Since research already has shown that the procedures work, it is possible to adjust the these by for example not having the same focus on experimental control or being more eclectic in the method application. Effective procedures may then reach populations and its staff, independent if being schooled in behavior analysis or not (Sidman, 1994).

Localization of Stimulus Equivalence Articles

It is fair to say that the stimulus equivalence field has changed when comparing the two time periods against each other. The identification of affiliation of the first author of the articles that met the inclusion criteria (Figure 4) showed that the number of published articles had an overall increase from the first to the second time period. The notion of change within the field is also strengthened by the results from the distribution of publications between the included journals as seen in Table 1. This present review showed that *JEAB* was the biggest outlet in the first time period, and *TPR* the biggest outlet in the second time period. Table 1 does not only show a change in publications of journals between the two time periods but also a low number of published articles in the journal *JABA*. This may be attributed to the low prevalence of applied articles and translational research. Further, the choice of assessing peer-viewed journals instead of using electronic bibliographic databases were to limit unnecessary omissions of stimulus equivalence citations. The use of databases in the procedure would on the other side have given the opportunity to uncover articles published in non-behavioral analytic journals.

Limitations and Future Research

As for the methodological setup of this present review, a few issues need to be accounted for. First, it should be stressed that all the articles included are published in

English, making the United States and Europe the most prominent contributors. This denotes that articles in other languages focusing on stimulus equivalence research were excluded from the present review and that the assessment might not have had the width as desired. For future research, it may be suggested that studies written in other languages are included.

Secondly, conducting a review requires that criteria for inclusion are set up for the work not to be overwhelming. In this study were articles limited to two different time periods. This can be seen as a limitation and future studies should focus on extending the investigation by either including more years or different time periods. For example, it may have been interesting to investigate one more time period straight after the writing of Sidman (1992, 1994) to see if the compliance became higher with respect to the mentioned considerations. Further, due to the amount of work, it was only taken IOA of two parameters. This can be considered as a limitation. However, it is important to note that the main focus of the review is instructions and that the results from the IOA on type of instructions were high, which may be argued to strengthen the reliability of the assessed parameters. Also, the outcome from the sample control indicates that the assessment, by using the aforementioned search strategies, included relevant articles for the two time periods.

Third, this review included MTS as a training and testing format. In future reviews, it could be interesting to include procedures like the go/no-go which in recent years has become more implemented as a procedure in stimulus equivalence research (e.g., Debert, Matos, & McIlvane, 2007). The choice of not including go/no-go was however based on the main purpose of the review, mainly, investigating two time periods and assessing the same type of variables. As for the first time period, which covered part of the beginning of stimulus equivalence research, experiments most often use the MTS format. Including the go/no-go procedure might then have produced difficulties in inferring parallels between each respective time periods.

Ultimately, future reviews might consider the inclusion of a separate and fourth instruction category, suggested as combined instructions. In this study were instructions that specified (i.e., SPEC instruction) and described the relationship between the stimuli (i.e., GEN instruction) in the same instruction assigned to the SPEC group. This combination of SPEC and GEN instructions might have given an incorrect picture of the occurrences of the specific instruction. Thus, separating and categorizing instructions as either SPEC or combined instructions might be more beneficial as for the variation of the instruction types.

Questions That Needs Answers

Overall, the results of the present review raise several questions. Such as why the use of instructions is varied across studies despite the fact that the use of instructions have been discussed by several authors (e.g., Arntzen, 2012; Arntzen et al., 2008; de Rose, McIlvane, Dube, Galpin, et al., 1988; Dymond & Rehfeldt, 2000; Fienup, Wright, & Fields, 2015; Sidman, 1992, 1994). Another question is why instructions that give away too much information about the task are used when such a variable is assumed to affect stimulus equivalence responding. Further, the review also raises the question about the uncertainty of what the stimulus equivalence outcomes are a product of (i.e., contingency shaped or rule-governed behavior) by including extensive pre-experimental instructions in the procedure (e.g., de Rose, McIlvane, Dube, Galpin, et al., 1988). Further investigations are therefore needed, to get a better understanding of the effects of instruction on stimulus equivalence outcome that can affect the establishment of stimulus equivalence.

Conclusion

The present study shred light of the prevalence of instruction given during stimulus equivalence research with the use of MTS format in the time periods 1988–1992 and 2013–-2017. This type of overview seems to be an important contribution with respect to the number of parameters employed in stimulus equivalence research. The results do not only provide information about the use of different parameters but also guides researchers in terms of methodological issues that warrant further empirical investigations and procedural adjustments. Instruction as a parameter can easily be modified, and if we are to follow the considerations mentioned by Sidman (1992, 1994), it would be an idea for researchers to consider the extent of instructions participants get prior the MTS training and also reduce variability with respect to the content of instructions. Such adjustments would make it easier to compare subsequent results across studies and different labs.

References

Arntzen, E. (2010). Om stimulusekvivalens: teoretiske betraktninger, oppsummering av en del empiri og noen praktiske implikasjoner. In S. Eikeseth & F. Svartdal (Eds.), *Anvendt atferdsanalyse: teori og praksis* (pp. 100–138). Oslo: Gyldendal akademisk, 2010.

Arntzen, E. (2012). Training and testing parameters in formation of stimulus equivalence: Methodological issues. *European Journal of Behavior Analysis, 13*(1), 123–135. doi:10.1080/15021149.2012.11434412

Arntzen, E., Grondahl, T., & Eilifsen, C. (2010). The effects of different training structures in the establishment of conditional discriminations and subsequent performance on tests for stimulus equivalence. *The Psychological Record, 60*(3), 437–462. doi:10.1007/bf03395720

Arntzen, E., Halstadtro, L. B., Bjerke, E., Wittner, K. J., & Kristiansen, A. (2014). On the sequential and concurrent presentation of trials establishing prerequisites for emergent relations. *The Behavior Analyst Today, 14*(1–2), 1–8. doi:10.1037/h0101280

Arntzen, E., & Lian, T. (2010). Trained and derived relations with pictures versus abstract stimuli as nodes. *The Psychological Record, 60*(4), 659–678. doi:10.1007/bf03395738

Arntzen, E., Norbom, A., & Fields, L. (2015). Sorting: an alternative measure of class formation? *The Psychological Record, 65*(4), 615–625. doi:10.1007/s40732-015-0132-5

Arntzen, E., Vaidya, M., & Halstadtro, L. B. (2008). Brief report: On the role of instruction in conditional discrimination training. *Expermental Analysis of Human Behavior Bulletin, 29*, 17–24. Retrieved from http://www.eahb.org/

Arntzen, E., & Vie, A. (2013). The expression of equivalence classes influenced by distractors during DMTS test trials. *European Journal of Behavior Analysis, 14*(1), 151–164. doi:10.1080/15021149.2013.11434453

Bush, K. M., Sidman, M., & de Rose, T. (1989). Contextual control of emergent equivalence relations. *Journal of the Experimental Analysis of Behavior, 51*(1), 29–45. doi:10.1901/jeab.1989.51-29

Catania, A. C. (2013). *Learning* (5th ed.). New York, NY: Sloan Publishing.

Cowley, B. J., Green, G., & Braunling-McMorrow, D. (1992). Using stimulus equivalence procedures to tech name-face matching to adults with brain injuries *Journal of Applied Behavior Analysis, 25*(2), 461–475. doi:10.1901/jaba.1992.25-461

Critchfield, T. S. (2011). To a young basic scientist, about to embark on a program of translational research. *The Behavior Analyst, 34*(2), 137–148. doi:10.1007/bf03392245

de Almeida, J. H., & de Rose, J. C. (2015). Changing the meaningfulness of abstract stimuli by the reorganization of equivalence classes: effects of delayed matching. *The Psychological Record, 65*(3), 451–461. doi:10.1007/s40732-015-0120-9

de Rose, J. C., McIlvane, W. J., Dube, W. V., Galpin, V. C., & Stoddard, L. T. (1988). Emergent simple discrimination established by indirect relation to differential consequences. *Journal of the Experimental Analysis of Behavior, 50*(1), 1–20. doi:10.1901/jeab.1988.50-1

de Rose, J. C., McIlvane, W. J., Dube, W. V., & Stoddard, L. T. (1988). Stimulus class formation and functional equivalence in moderately retarded individuals' conditional discrimination. *Behavioural Processes, 17*(2), 167–175. doi:10.1016/0376-6357(88)90033-2

Debert, P., Matos, M. A., & McIlvane, W. J. (2007). Conditional relations with compound abstract stimuli using a go/no-go procedure *Journal of the Experimental Analysis of Behavior, 87*(1), 89–96. doi:10.1901/jeab.2007.46-05

Dickins, D. (2015). Vocalizing phonologically correct non-word stimuli during equivalence class formation. *European Journal of Behavior Analysis, 16*(2), 248–278. doi:10.1080/15021149.2015.1083284

Doughty, A. H., & Best, L. (2017). Transfer of function and prior derived-relations testing. *Behavioural Processes, 143*, 4–6. doi:10.1016/j.beproc.2017.07.010

Dymond, S., O'Hora, D., Whelan, R., & O'Donovan, A. (2006). Citation analysis of Skinner's verbal behavior: 1984–2004. *The Behavior Analyst, 29*(1), 75–88. doi:10.1007/bf03392118

Dymond, S., & Rehfeldt, R. A. (2000). Understanding complex behavior: The transformation of stimulus functions. *The Behavior Analyst, 23*(2), 239–254. doi:10.1007/bf03392013

Fields, L., Arntzen, E., & Moksness, M. (2014). Stimulus sorting: A quick and sensitive index of equivalence class formation. *The Psychological Record, 64*(3), 487–498. doi:10.1007/s40732-014-0034-y

Fields, L., Arntzen, E., Nartey, R. K., & Eilifsen, C. (2012). Effects of a meaningful, a discriminative, and a meaningless stimulus on equivalence class formation. *Journal of the Experimental Analysis of Behavior, 97*(2), 163–181. doi:10.1901/jeab.2012.97-163

Fienup, D. M., Wright, N. A., & Fields, L. (2015). Optimizing equivalence-based instruction: Effects of training protocols on equivalence class formation. *Journal of Applied Behavior Analysis, 48*(3), 613–631. doi:10.1002/jaba.234

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

Green, G., Sigurdardottir, Z. G., & Saunders, R. R. (1991). The role of instructions in the transfer of ordinal functions through equivalence classes. *Journal of the Experimental Analysis of Behavior, 55*(3), 287–304. doi:10.1901/jeab.1991.55-287

Imam, A. A. (2006). Experimental control of nodality via equal presentations of conditional discriminations in different equivalence protocols under speed and no-speed conditions *Journal of the Experimental Analysis of Behavior, 85*(1), 107–124.

Imam, A. A., & Warner, T. A. (2014). Test order effects in simultaneous protocols. *Learning & Behavior, 42*(1), 93–103. doi:10.3758/s13420-013-0128-4

Johnson, C., Meleshkevich, O., & Dube, W. V. (2014). Merging separately established stimulus classes with outcome-specific reinforcement. *Journal of the Experimental Analysis of Behavior, 101*(1), 38–50. doi:10.1002/jeab.61

Kenny, N., Barnes-Holmes, D., & Stewart, I. (2014). Competing arbitrary and non-arbitrary relational responding in normally developing children and children diagnosed with autism. *The Psychological Record, 64*(4), 755–768. doi:10.1007/s40732-014-0042-y

Lian, T., & Arntzen, E. (2013). Delayed matching-to-sample and linear series training structures. *The Psychological Record, 63*(3), 545–561. doi:10.11133/j.tpr.2013.63.3.010

McPherson, A., Bonem, M., Green, G., & Osborne, J. G. (1984). A citation analysis of the influence on research of Skinner's Verbal Behavior. *The Behavior Analyst, 7*(2), 157–167. doi:10.1007/bf03391898

Pautasso, M. (2013). Ten simple rules for writing a literature review. *PLoS Computational Biology, 9*(7), 1–4.

Pilgrim, C. (2016). Considering definitions of stimulus equivalence. *European Journal of Behavior Analysis, 17*(1), 105–114. doi:10.1080/15021149.2016.1156312

Pilgrim, C., Jackson, J., & Galizio, M. (2000). Acquisition of arbitrarily conditional discriminations by young normally developing children. *Journal of Experimental Analysis of Behavior, 73*(2), 177–193. doi: 10.1901/jeab.2000.73-177

Reyes-Giordano, K., & Fienup, D. M. (2015). Emergence of topographical responding following equivalence-based neuroanatomy instruction. *The Psychological Record, 65*(3), 495–507. doi:10.1007/s40732-015-0125-4

Saunders, K. J., Saunders, R. R., Williams, D. C., & Spradlin, J. E. (1993). An interaction of instructions and training design on stimulus class formation: Extending the analysis of equivalence. *The Psychological Record, 43*(4), 725–744. doi:10.1007/BF03395909

Saunders, R. R., Wachter, J., & Spradlin, J. E. (1988). Establishing auditory stimulus control over an eight-member equivalence class via conditional discrimination procedures. *Journal of the Experimental Analysis of Behavior, 49*(1), 95–115.

Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech and Hearing Research, 14*(1), 5–13. doi:10.1044/jshr.1401.05

Sidman, M. (1992). Equivalence relations: Some basic considerations. In S. C. Hayes & L. J. Hayes (Eds.), *Understanding verbal relations* (pp. 15–27). Reno, NV: Context Press.

Sidman, M. (1994). *Equivalence relations and behavior: a research story*. Boston: Authors cooperative.

Sidman, M. (2000). Equivalence relations and the reinforcement contingency. *Journal of the Experimental Analysis of Behavior, 74*(1), 127–146. doi:10.1901/jeab.2000.74-127

Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior, 37*(1), 5–22.

Sidman, M., Wynne, C. K., Maguire, R. W., & Barnes, T. (1989). Functional classes and equivalence relations. *Journal of the Experimental Analysis of Behavior, 52*(3), 261–274. doi:10.1901/jeab.1989.52-261

Silveira, M. V., Mackay, H. A., & de Rose, J. C. (2017). Measuring the "transfer of meaning" through members of equivalence classes merged via a class-specific reinforcement procedure. *Learning & Behavior*, 1–14. doi:10.3758/s13420-017-0298-6

Zygmont, D. M., Lazar, R. M., Dube, W. V., & Mcilvane, W. J. (1992). Teaching arbitrary matching via sample stimulus-control shaping to young children and mentally retarded individuals: a methodological note. *Journal of Experimental Analysis of Behavior, 57*(1), 109–117. doi:10.1901/jeab.1992.57-109

Table 1

Number of experiments categorized in each journal for both time periods

Journal	1998–1992	2013–2017	Total
The Psychological Record	6	41	47
Journal of the Experimental Analysis of Behavior	33	9	42
Journal of Applied Behavior Analysis	1	4	5
Behavioural Processes	2	2	4
European Journal of Behavior Analysis	–	9	9
Learning & Behavior	–	9	9
Psicologia: Relexão e Crítica	–	1	1
The Quarterly Journal of Experimental Psychology	–	1	1

Note. The table shows an overview of included journals and number of published articles for the two time periods and in total.

Table 2
Overview of parameter analysis

Parameters	Specified	Combined time periods % (*n* = 118)	1988–1992 % (*n* = 42)	2013–2017 % (*n* = 76)
Categorization	Basic	90,5	95	88
	Applied	2,5	-	4
	Translational	7	5	8
Presentation of instruction	Screen	46	17	62
	Told	37	64	22
	Written	7,5	12	5
	Modified	2,5	2,5	3
	Not specified	7	4,5	8
Use of material for MTS procedure	Computer	-	67	87
	Table top	-	31	8
	Combination	-	-	5
	Not specified	-	2	-
Population demography	Students	68,5	66,5	70
	Adults	18	24	14
	Children	12	9,5	13
	Combined*	1,5	-	3
	Female	7	12	4
	Male	10	14	8
	Mixed	56	48	60,5
	Not specified	27	26	27,5
	Typical students	68,5	67	70
	Typical adults	13,5	14	13
	Typical children	7	2	5
	Atypical students	-	-	-
	Atypical adults	4	10	1
	Atypical children	4	7	7

Note. The table shows an analysis of different parameters for combined and individual time periods. *Combination of adults and children.

Table 2

Continued overview of parameter analysis

Parameters	Specified	Combined time periods % (*n* = 118)	1988–1992 % (*n* = 42)	2013–2017 % (*n* = 76)
Population demography	Mixed children typ. & atyp.	1	-	1
	Mixed adult & children typ. & atyp.	2	-	3
Training structure	Linear Series	37,5	7	54
	Many to One	9	9,5	9
	One to Many	41,5	64,5	29
	Combinations or modified training structures	12	19	8
Training & testing protocols	Simultaneous	66	43	79
	Simple to Complex	18	24	14,5
	Complex to Simple	8	14	5,5
	Testing only one equivalence property	8	19	1

Note. The table shows an analysis of different parameters for combined and individual time periods. Typ. = typical, atyp. = atypical

Table 3

Detailed analysis of the time periods, type of instructions and different parameters

Time periods		1988–1992			2013–2017		
Parameter	Specified	GEN %	SPEC %	NTS %	GEN %	SPEC %	NTS %
Presentation of instructions	Screen	5,5	25	25	36	37,5	78
	Told	78	50	55	41	37,5	11
	Written	16,5	-	10	5	12,5	4,5
	Modified	-	25	-	-	-	2
	Not specified	-		10	18	12,5	4,5
Participant demography	Students	67	25	75	68	50	74
	Adults	22	75	15	-	12,5	22
	Children	11	-	10	32	37,5	-
	Combined*	-	-	-	-	-	4
	Typical students	67	25	75	68	50	74
	Typical adults	16,5	-	15	-	-	22
	Typical children	-	-	5	4,5	37,5	-
	Atypical students	-	-	-	-	-	-
	Atypical adults	5,5	75	-	-	12,5	-
	Atypical children	11	-	5	23	-	-
	Mixed children typ. & atyp.	-	-	-	4,5	-	-
	Mixed adult & children typ. & atyp.	-	-	-	-	-	4
Other	Experiments wanting participants to pass equivalence testing (n = 7)	-	-	-	43	28,5	28,5
Combination of Training Structure and Protocol	Linear Series & Simultaneous	6	25	5	50	13	39
	Linear Series & Simple to Complex	-	-	-	9	25	4
	Linear Series & Complex to Simple	-	-	-	-	-	9

Note. The table shows an detailed analysis of different parameters for the individual time periods with each respective type of instruction. Typ. = typical, atyp. = atypical. GEN = general instructions, SPEC = specific instructions, NTS = not task specific instruction. *Combination of adults and children.

Table 3

Continued overview of detailed analysis of the time periods, type of instructions and different parameters

Time periods		1988–1992			2013–2017		
Parameter	Specified	GEN %	SPEC %	NTS %	GEN %	SPEC %	NTS %
Combination of Training Structure and Protocol	Many to One & Simultaneous	-	-	5	-	-	15
	Many to One & Simple to Complex	-	-	-	-	-	-
	Many to One & Complex to Simple	6	-	5	-	-	-
	One to Many & Simultaneous	28	-	35	36	25	22
	One to Many & Simple to Complex	33	-	20	5	-	2
	One to Many & Complex to Simple	11	-	5	-	-	-
	Mixed training structures and protocols	28	75	25	-	38	9

Note. The table shows an detailed analysis of different parameters for the individual time periods with each respective type of instruction. GEN = general instructions, SPEC = specific instructions, NTS = not task specific instruction.

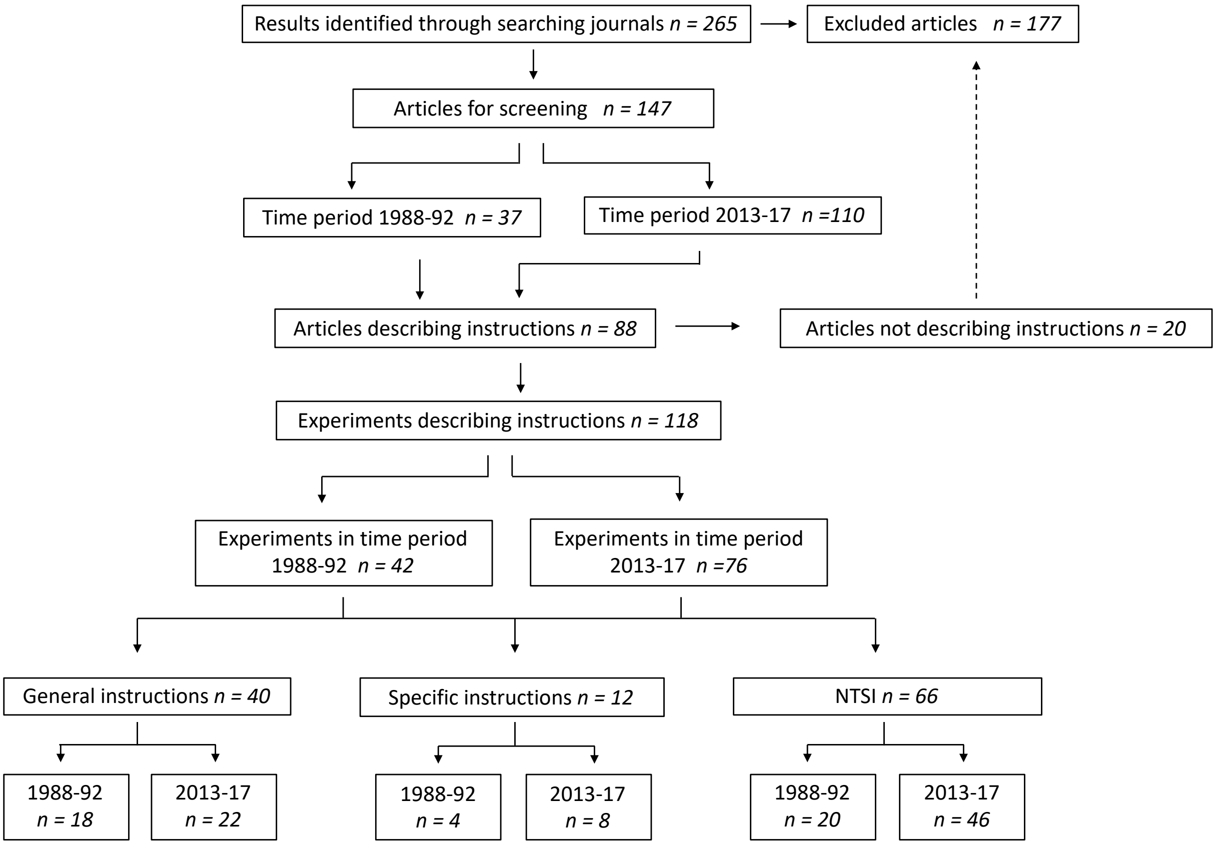

Figure 1. Flowchart illustrating the search conducted in the literature review

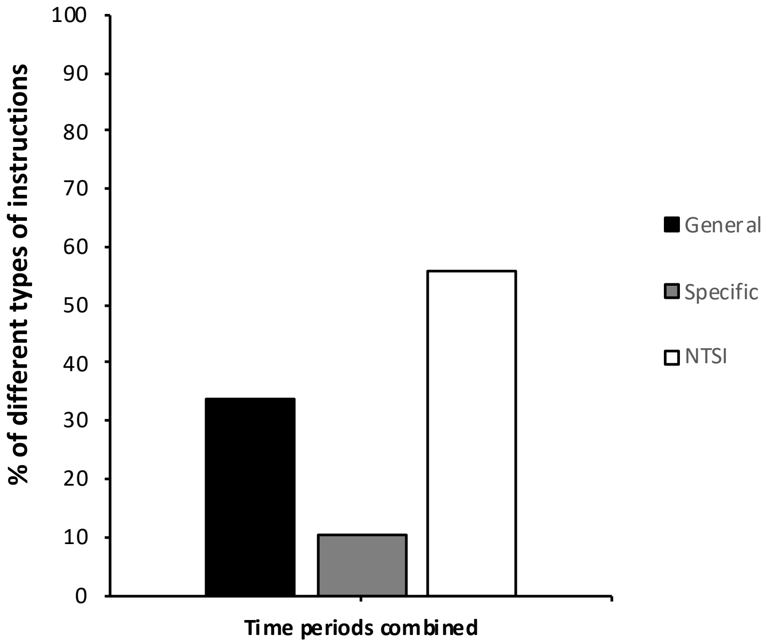

Figure 2. Percentage of type of instructions in the combined time periods 1988–1992 and 2013–2017.

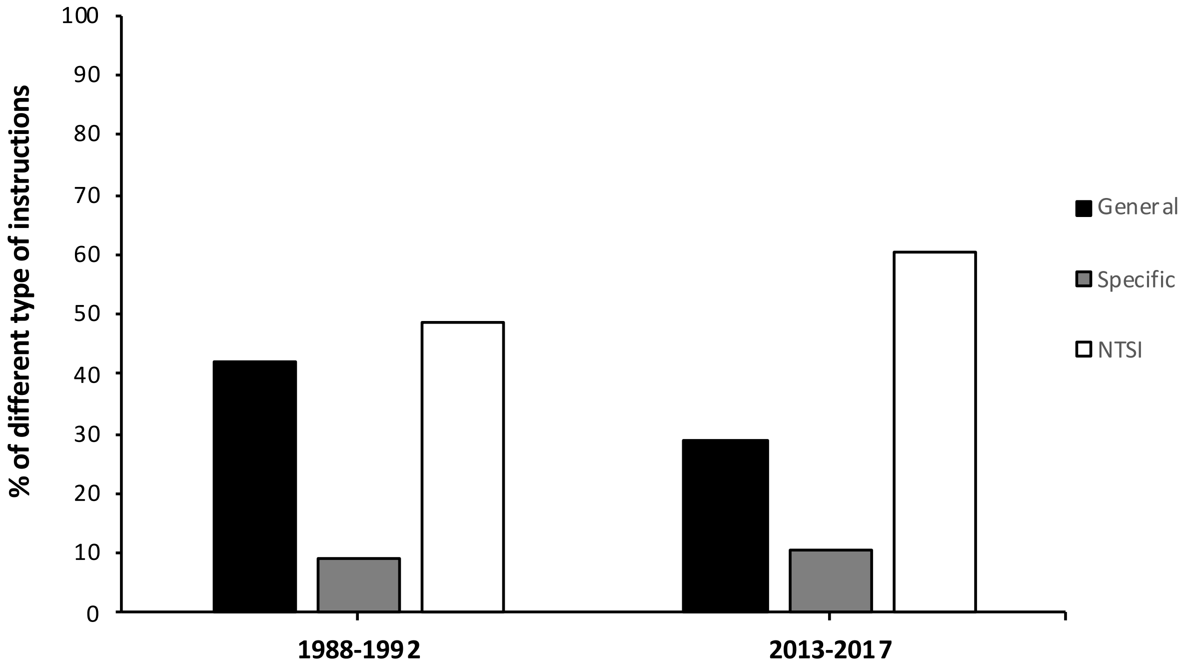

Figure 3. Percentage of type of instructions for the first and second time period.

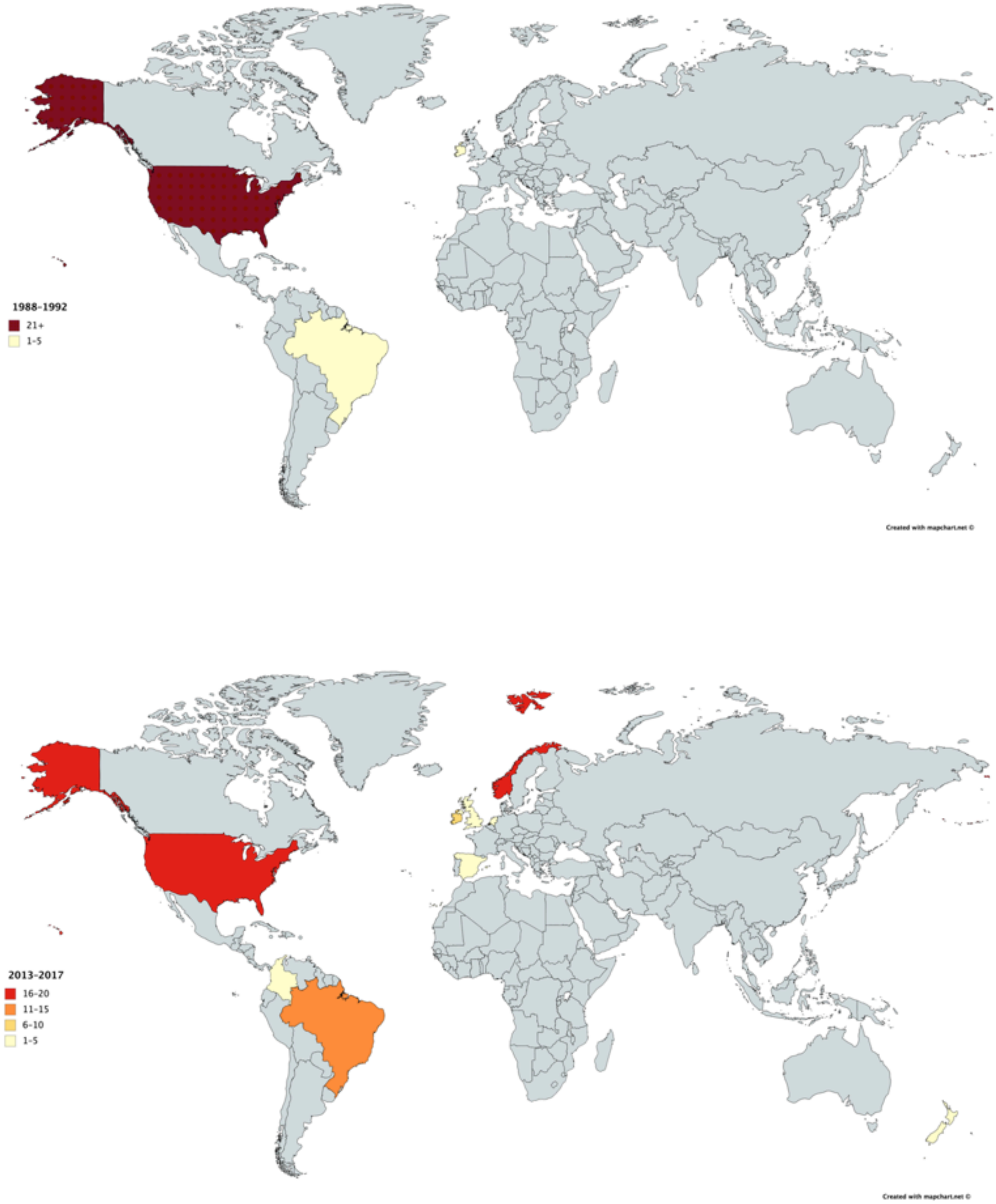

Figure 4. The figure show a worldwide localization of articles included in the review. The time period 1998–1992 is represented in the upper panel and the time period 2013–2017 in the lower panel.

Effects of Instructions on Training of Conditional Discriminations and the Formation of Equivalence Classes

Annika Poulsen

Oslo Metropolitan University

Abstract

The role of instructions in stimulus equivalence research has been widely discussed. However, only a few studies have empirically investigated the effects of instructions. Hence, in the current experiment, two different instructions were given preceding the experimental task to investigate potential effects of the instructions on stimulus equivalence outcome. Participants were assigned to either one of three groups: general instructions (e.g., “your task is to find out which stimuli that belongs together”), specific instructions (e.g., “this stimulus is a fish and belongs together with the car which is this stimulus”) or a control group which did not get any instructions about the task. The study used a linear series training structure with a stimulus set consisting of potentially three 3-member classes. The depended measures were number of training trials and equivalence responding. The results showed that the condition with general instructions gave a higher yield of stimulus class formation compared to the other two conditions.

Keywords: stimulus equivalence, instructions, matching-to-sample, conditional discriminations, adults

In the behavior analytic terminology, many instructions can be described as rules which according to Skinner (1969) is defined as a contingency specified stimulus. This definition infers that the rule itself is a discriminative stimulus describing the occasion, response and the consequence of the behavior. The behavior of the person that follows the rule or instruction is then often termed as rule-governed behavior and used as an explanation when trying to describe responses that have not been in direct contact with its particular contingencies, for example in trial and error exposures (Catania, 2013).

Considering the stringent definition to Skinner (1969) it can be argued that rules often is used without actually qualifying to the Skinner's definition. Examples of instructions that are not contingency specifying are "matches" or "goes together", while an example of a rule which is contingency specifying is "click on the picture two times and you will get a point". Thus, instructions can be complete or incomplete by describing the full – or only part of the contingency. Everyday examples can be when a person gives you a complete instruction categorized as advice "it is raining, bring the umbrella, or you will get wet" or when you get an incomplete instruction categorized as a warning "do not touch the hot stove". These examples incline that instructions appear to take different types of forms depending on how it is formulated – and that we as instruction givers can formulate information in different ways by for example using different topographies and arranging special conditions which may lead to the same type of behavior.

Another description of rules or instructions suggested by Schlinger (1993) and Schlinger and Blakely (1987) was that instead of considering the functioning of an instruction as a discriminative stimulus for a response, it would rather be seen as the altering of functions of other stimuli by operating as a verbal antecedent. This would, in turn, change the listener's behavior when natural contingencies were slowly effective or not effective at all. Although, a prerequisite for that would be that the contingencies of the instruction were more potent than the actual natural contingencies themselves (Catania, 2013).

Despite the different views with respect to instructions Skinner (1969) wrote that if people are to follow instructions, then, a learning history with listener behavior is a prerequisite. This means that the listener behavior has been reinforced in the contact of similar verbal stimuli at previous occasions. This can be understood in contrast to the contingency-shaped response where a person is responding “in a given way and with a given probability because the response has been followed by a given kind of consequence in the past” (Skinner, 1969, p. 147).

Since the majority of human beings, from the early age, have a functional language and also the possibility to give and follow instructions has the topic of rule-governed behavior drawn the attention of many research fields within behavior analysis. Consequently, experiments have been conducted to see if and how instructions could influence behavior (e.g., Catania, Shimoff, & Matthews, 1989; Hayes, 1989). One of those fields has been stimulus equivalence where the primary focus is to describe emergent responding. Instructions are in this type of experiments important since it can have a critical effect on the learners responding.

Stimulus Equivalence

In everyday language, stimulus equivalence can be one description of how organisms treat dissimilar events as the same even if they have not been trained directly (Green & Saunders, 1998). In the training and testing for equivalence class formation, conditioned discrimination has been the most common procedure in the form of matching to sample (MTS) arrangements, where three-term contingencies come under the control of conditional stimuli (Sidman, 2000).

The concept of stimulus equivalence implies that new relations, which never have been directly reinforced, can emerge from earlier trained conditioned discriminations. Thus, after a participant have been trained to match stimulus A to stimulus B, and stimulus B to stimulus C they often show the three properties of reflexivity, symmetry, and transitivity which define stimulus equivalence (Green & Saunders, 1998) under extinction conditions. Reflexivity refers to the matching of the stimulus itself, for example by choosing the A stimulus in the presence of stimulus A. Responding in accordance to symmetry means that there is interchangeability between the sample and the correct comparison. Thus, the participant will select stimulus B when stimulus A is presented when earlier training has consisted of matching stimulus A with stimulus B. Responding in accordance to transitivity means that when stimulus A is presented stimulus C will be selected. Finally, the last test which is often used, but is not a part of the definition is global equivalence, which means that the participant will select stimulus A as a comparison when stimulus C is presented (Sidman, 1992, 2000).

The investigation of emergent stimulus control has come to be a significant research area, especially in the understanding of the development of arbitrarily conditional discriminations. Studies have among other been investigating a number of conditions with respect to training procedures and different experimental setups (Pilgrim, Jackson, & Galizio, 2000). However, the role of instructional effects of conditional discrimination training and equivalence outcomes has not been examined to a great extent. Sidman (1992, 1994) discussed the implications of instructions during stimulus equivalence research and wrote that it is evident that verbal instructions can facilitate the establishment or be the reason for equivalence relations. Participants might also get less sensitive to the natural contingencies (e.g., Catania et al., 1989) in the conditional discrimination arrangement if instructions that gives away the purpose of the task are given in advance of the experiment. A variable such as the persons learning history with instructions like *goes with* may then be enough to establish equivalence relations. However, what not is clear is if verbal rules are required for a person to establish equivalence classes just because they can give rise to them (Sidman, 1992).

Until we can answer the question, that is, if verbal rules give rise to equivalence or if equivalence makes rules possible, Sidman proposes that researchers have to be careful with the instructions they are giving during experimental procedures that concern equivalence testing. Of course, it should not be forgotten that there is the possibility that the participant generates rules when nonverbal contingencies are used during an experiment, but this may rather be attributed to the actual experimental manipulations.

In equivalence research, different kinds of instructions have commonly been used. Provided what Sidman (1992, 1994) states about the instructional effects on the equivalence outcomes it can lead to difficulties in making conclusions when comparing studies that are using extensive instructions that may give away too much information (e.g., de Almeida & de Rose, 2015; Devany, Hayes, & Nelson, 1986) versus little information (e.g., Bush, Sidman, & de Rose, 1989; Sidman, Wynne, Maguire, & Barnes, 1989) about the experimental task – given that the experiments investigate the same kinds of variables.

There are, as mentioned, few published studies on the topic of instruction and its effects of instruction on conditional discrimination training and equivalence responding. So far, the investigations mainly focused on the effects of instructions with typical developed adults and children as participants. For example, in Experiment 1 in a study by Green, Sigurdardottir, and Saunders (1991) adults were exposed to full information (i.e., touch the figure that you think goes with the first one that appeared) about the MTS arrangement. The results showed that all participants formed four equivalence classes. Hence, indicating that the instructions could have facilitated the training and testing for stimulus equivalence class formation.

In another study by Pilgrim et al. (2000), which consisted of two experiments, the purpose was to see if different types of instructions could establish conditional discrimination with abstract arbitrarily related. The first experiment, which included typically developed children as participants in the age span three and six years, either gave the participants general instructions, i.e. “Look at this one. This will tell you where the prize is” or specific instructions, i.e. “When this one is in the middle pick this one” followed by holding up the correct comparison. The instructions were given in a new session prior to the first five trials or if the participant made an incorrect response during training. In some of the conditions a naming procedure were superimposed in combination with the specific instructions, where the participant had to name the samples themselves. The results showed that the instructions alone did not improve arbitrarily conditional discriminations, but the results of the combination of specific instruction and naming appeared promising. The second experiment which used the same type of participants included a naming condition together with the specific instruction where the samples were either named by the participants or predetermined by the experimenter. When the names were predetermined, five of seven participants showed rapid mastery of the arbitrary discriminations compared to the five participants who named the samples themselves and failed to meet the criterion (Pilgrim et al., 2000).

In a study by Arntzen, Vaidya, and Halstadtro (2008) the effects of general instructions, i.e. “Some of the stimuli belong together, and your task is to find out which stimuli go together” (p. 20) were investigated, when children as participants failed to acquire conditional discriminations during differential reinforcement contingencies alone. One of the purposes of the study was, therefore, to investigate the effectiveness of timing the instructions during the experiment, arguing that it might be more helpful to present the instruction during the ongoing experiment instead of giving it before the task had started when the participants were unfamiliar to the arrangement. The findings showed an increased accuracy of responding after the instructions were given compared to prior trials without instructions. All participants reached criterion for testing and passed the test for symmetry, two of nine met the criteria for the equivalence relation.

Due to the absence of the investigation of instructional effects on equivalence responding with adults as participants, and the warrant for such studies (e.g., see further discussions in Dymond & Rehfeldt, 2000; Silveira, Mackay, & de Rose, 2017), the purpose of this study was to investigate if two different types of instructions could facilitate training of conditional discriminations and the responding in accordance with stimulus equivalence. The study which included abstract stimuli that were arbitrarily related also implemented a linear series training structure together with the simultaneous (SIM) protocol. Since earlier research have shown that this combination produces lower yield of equivalence responding (Arntzen, Nartey, & Fields, 2015; Arntzen, Norbom, & Fields, 2015; Imam & Warner, 2014) it was hypothesized to make it more difficult for the participants to pass the testing condition, thus, potentially demonstrate the effects of the instructions.

Method

Participants

Thirty-six adults served as participants in this experiment. The participants were 26 females and four males ranging from 19 to 51 years, with an average age of 25.33 years, recruited from the Oslo Metropolitan University or through personal contacts. None of the participants had any experience with the stimuli employed in the experiment or formal knowledge about stimulus equivalence. Also, two participants chose not to finish the experimental session or were cancelled due to personal health. Prior to the experiment, the participants were given an informed consent form with general information such as that the experiment was carried out on a computer and that the participant’s task was to choose stimuli presented on the screen. They were also told about their anonymity and that they could withdraw from the experiment without adverse consequences. Further, they were informed that no necessary computer skills were needed and that the experiment would last approximately one hour. After the experimental session, all participants were thanked for their participation, debriefed about their results and informed about stimulus equivalence. All the participants were included in a lottery for a smart tablet as compensation for participating.

Apparatus and Setting

The experiment was conducted in cubicles with an average size of 1 m2 containing a table and chair. The participants were seated on a chair in front of the table, which was placed against a wall. All experimental tasks were conducted on an HP ProBook 655 G1 laptop with a 15.5-inch screen running Windows 7 Enterprise. A custom-made MTS and sorting program were used in the experiment.

Stimuli

A set of nine abstract visual stimuli (Figure 1), which included a variety of Arabic, Chinese, Cyrillic, Japanese, and Hebrew characters, were used as potential members of the equivalence classes. The stimuli were displayed in black color on a white background with an average size of 3.5 × 3 cm. The sample was always presented in the center of the screen and the comparisons in the corners of the screen with an alternated position for each trial. The diagonal distance between the sample stimulus and the comparisons were 12.5 cm. The vertical and horizontal distance between comparisons were 10 cm and 25.5 cm, respectively.

Procedure

Experimental design. The participants (*n* = 36) were randomly assigned to one of three conditions that differed in the type of instructions provided preceding the MTS training: Control group (CTRL) with no instructions, General instructions (GEN), or Specific instructions (SPEC). For the GEN and SPEC group did the experiment begin with the presentation of instructions and an individual quiz. After that, the participants attempted to establish on the computer three equivalence classes with three members. Each of which had a linear series training structure (A→B→C). Following the completion of the equivalence class formation, the participants were exposed to a sorting task on the computer, where the task was to sort the pictures, as they preferred.

Instructions. After reading and signing the informed consent, instructions were given to participants in the GEN and SPEC group. The instructions were typed with 36-point Calibri font, printed and laminated on A4-paper. For the SPEC group, the instruction came with two additional pictures which visually demonstrated the information (Figure 2). The control group did not get any instruction or quiz and started the experimental session after signing the informed consent. For the GEN group with the general instruction the following instructions were given:

> Some stimuli **goes together** with other stimuli. Your task is to find out which stimuli **goes together** with one another. In addition, the stimuli **that go together** do not need to have physical similarities. Furthermore, the position or order of the stimuli is irrelevant.

For the SPEC group with the specific instruction the following instruction was given together with the pictures as seen in Figure 2:

> This stimulus is a book and **belongs together** with the car, which is this stimulus. The position or order of the stimuli is irrelevant.

Participants in the GEN and SPEC group were asked to read the instructions out loud while the researcher was present in the experimental setting. When the participants in the SPEC group read the specific instruction the researcher simultaneous pointed to the correct stimuli on the picture (see the upper panel in Figure 2). Additionally, another picture was presented afterwards with new stimuli and empty thought bubbles (see lower panel in Figure 2), the participants were then asked to self-generate names for the new stimuli.

Questions about the instructions were limited to translation of words or giving synonyms to unfamiliar terms such as “symbols” for “stimuli”. After the instructions, the participants were presented with quizzes that were custom-made for the instructions. That is, only containing questions about the information given. The instruction and pictures (for the SPEC group) remained on the desk beside the computer throughout the experiment.

Quiz. Two quizzes were prepared to assure that the participants comprehended the instructions. The two different quizzes were customized for the GEN and SPEC instructions, respectively, and contained questions only to the instructions given. If the participant did not answer 100 % correct on the quiz, the instructions and quiz were reintroduced until the participants answered correctly on all questions.

After the instructions and quizzes were completed the participants were provided with the following instructions about the MTS task. For the control group the experimental session directly began with the MTS instructions, which was presented on the computer screen in Norwegian:

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

Sorting stimuli. A sorting task was conducted after testing for stimulus equivalence class formation. The participant was asked to sort the stimuli that were used in the present experiment, by moving stimuli and drawing relationships among them with the computer mouse. Prior to the sorting, the participants were instructed in English, regardless of the participant's first language, as follows:

> Sort the pictures as you want. When you have sorted the pictures as you want, please mark the sorting by holding down the left mouse button and draw by moving the mouse. The stimuli will be placed on top of each other, you have to drag them to an optional location on the screen. By moving one of the stimuli, you can undo the drawn markings.

The researcher was present at the beginning of the sorting task and only referred to the instructions given if the participant had questions about the content. If no questions were asked or when potential questions were answered the researcher left the experimental setting. The sorting task was presented twice and used as an estimate of the equivalence class formation of emergent relations.

Trial format and contingencies. For each trial, a sample stimulus was presented in the middle of the screen. After a mouse-click on the sample, three comparison stimuli appeared in three of four corners of the screen while the sample still was present. The positions of the comparison were randomized by the software program and remained on the screen until the participant made a selection of one comparison. During baseline training, the selection of correct comparison was followed by programmed consequences such as "correct" and synonyms like "excellent" and "very good". Incorrectly selected comparison stimuli were followed by the word "wrong". Programmed consequences after selection of a comparison, if presented, were shown in the middle of the screen for 500 ms and followed by a 1000 ms intertrial interval displayed as a blank screen before the next trial.

Acquisition and maintenance of baseline relations. The baseline relations were trained in a concurrent order (AB/BC) with thinning of programmed consequences, all with mastery criterion of 100%. The order of training is shown in Table 1. The blocks involved training AB and BC relations for each class with a total of 18 trials (three presentations each of the six trial types) and was coded as followed: A1/B1B2B3, A2/B1B2B3, A3/B1B2B3, B1/C1C2C3, B2/C1C2C3, B3/C1C2C3 (the members coded as letters and the classes as numbers where the first letter and number is the sample followed by three comparisons. The correct comparison is underlined). Programmed consequences were at the beginning provided for comparison selection of every trial and repeated until the participant responded correctly on all trials. The probability of programmed consequences for each trial was then thinned in the following order: 75%, 50%, and 0%. During the thinning of the programmed consequences, the blocks were repeated until the participant selected the correct comparison for all trial types.

Emergent relations test blocks. The emergent relations were tested and consisted of 108 randomly mixed trials (six trials for each trial type) without programmed consequences. The tests consisted of previously directly trained baseline trials: A1/B1B2B3, A2/B1B2B3, A3/B1B2B3, B1/C1C2C3, B2/C1C2C3, B3/C1C2C3, together with the test for emerged symmetry relations: B1/A1A2A3, B2/A1A2A3, B3/A1A2A3, C1/B1B2B3, C2/B1B2B3, C3/B1B2B3, those for emerged transitivity relations: A1/C1C2C3, A2/C1C2C3, A3/C1C2C3 and finally the emerged equivalence relations: C1/A1A2A3, C2/A1A2A3, C3/A1A2A3.

Behavior recorded. In this experiment, the target response was defined as mouse-clicking on stimuli. The criterion for responding in accordance with equivalence was set to 90% where 33 of 36 correct responses were needed for the baseline and symmetry trials and 17 of 18 correct responses for transitivity and equivalence trials.

Results

Individual data of all the participants are summarized in Table 2 and shows the main results such as participant number, number of training trials, number of errors, responding during testing (baseline, symmetry, transitivity, equivalence relations) and post-class formation sorting. Additionally, reaction time (RT) and class categorization were added as supplementary measures. The median value was used to calculate the number of training trials to criterion and mean of median values for RT due to high scores and time, respectively, for some of the participants.

Training

The main findings show that the GEN group had 261 as a median number of trials to criterion compared to the SPEC group with 279, and the CTRL group with 306 trials. During the acquisition of baseline relations (until reached criterion for correct trials in a test block during 100% programmed consequences) the GEN group had 198 median of trials with a median of 70,5 errors. The SPEC group 216 trials with 78,5 errors and the CTRL group 225 trials with 76,5 errors. During the maintenance (fading the programmed consequences to 75, 50, and 0 %) all the groups had the identical sum of median trials of 18 and zero error trials. Two exceptions were seen with P15535 and P15753 in the CTRL group, those participants had 108 and 90 training trials, respectively, during the 75% programmed consequence phase.

Trials to criterion, that is, the total of trials before testing for stimulus equivalence for each group (CTRL, GEN and SPEC) showed no statistically significant difference when conducting a one-way ANOVA test, $F(2, 33) = .687$, $p = .510$ nor when conducting a U test between participants who responded in accordance with equivalence and those participants who did not, independent of group assignment ($U = 82.5$, $p = .153$).

Equivalence Class Formation

As seen in Table 2, the highest outcome of stimulus class formation can be seen in the GEN group. Seven of 12 participants responded in accordance with stimulus equivalence, thus responded over the 90 % criteria for baseline, symmetry, transitivity and equivalence relations. Nine of 12 participants had both the baseline and symmetry relations intact. Moreover, eight of 12 participants established the transitivity and seven of 12 participants the equivalence relations. In the CTRL group, responded one participant in accordance with stimulus equivalence. Eight of 12 participants established baseline relations. Six of 12 participants had the symmetry relations intact. Further, three and two of the 12 participants had the transitivity respectively the equivalence relations intact. In the SPEC group responded one of 12 participants in accordance with stimulus equivalence. Eleven of 12 participants formed classes for the baseline relations and eight of 12 participants the symmetry relations. Further, three of 12 participants established transitivity relations and two of 12 participants the equivalence relations.

A Fisher's Exact Test (FET) was conducted to see if instructions had an effect on class formation. The results showed a statistically significant difference between the CTRL group and the GEN group ($p = .027$). Additionally, a statistically significant difference was shown between the SPEC group and the GEN group ($p = .027$). There was no significant difference between the CTRL and the SPEC groups ($p = 1.0$).

Sorting

The right columns in Table 2 show the post-sorting for all participants in each respective group. The two columns named 1 and 2 refer the two separated sorting tasks. The table shows the grouping of stimuli (horizontally) by the participants for each class. Further, the stimuli in the equivalence classes, are sorted by the numbers 1, 2, and 3. As an example, the number 300 indicates that the participant has grouped three stimuli from Class 1 and none from Class 2 or 3. This was considered as correctly sorted according to the experimenter-defined classes (i.e., three 3-member classes). For Class 2 and 3, respectively, the grouping of 030 and 003 were considered as correct. All the correct sorted classes are marked in bold text.

For the CTRL group, which had one participant showing equivalence class formation, sorted 10 of 12 participants correctly and according to the three experimenter-defined classes in both phases. P15546 and P15754 did not sort the classes correctly even though P15546 had the baseline relation intact in the test. In the GEN group sorted 11 of 12 participants correctly and according to the experimenter defined classes in both phases. Seven of the participants responded in accordance with stimulus equivalence during the test. Of the 12 participants, did four participants (P15549, P15758, P15759, and P15769) not show equivalence class formation. P15536 did not form classes nor sort the stimuli correct. For the SPEC group showed the results that the participant who responded in accordance with stimulus equivalence also sorted correctly. Additional ten of 12 participants sorted correctly during sorting in both phases in the SPEC group even if they did not form equivalence classes. P15550, P15756 and P15757 had three of four properties intact in the test. A FET test showed no statistically significant difference for the three groups in the two sorting conditions ($p = 1.0$).

Response Patterns During Testing

An analysis of response patterns (Figure 3) was conducted, for each group, to see if the participants who did not form equivalence classes, self-generated own stimulus classes (i.e., participant-defined classes) or had an indeterminate response pattern during the MTS test.

In experimenter-defined classes were the criteria set to include five or more correct relations occurring in the test block, the identical criteria were set for the participant-defined classes. This meant that an 83% or higher responding was considered enough to meet the criteria. Four or fewer relations occurring in the test block was considered as an indeterminate response pattern by the participant. For the participant-defined and indeterminate classes, the participants had to pick different stimulus relations from the correctly defined experimenter-defined classes (A1-B2 or C3-B1, as an example).

As seen in Figure 3, the CTRL group produced 66,05%, the GEN group 79,40%, and, the SPEC group 70,45% trial responses in accordance with the experimenter-defined classes.

Additionally, the indeterminate responding by the participants were calculated to 28,94%, 16,90%, and 31,57% for each group, respectively. Finally, 5,02%, 1,85%, and, 2,31% of the trials were participant-defined trial responses for the CTRL, GEN, and SPEC group, respectively.

With regards to participant-defined trial responses, it was observed that four participants in the CTRL group had defined at least one own class in testing, that is, choosing five or more trial responses of a not experimenter-defined comparison in the presence of a sample stimulus. For example, P15766 chose the comparison C2 in the presence of the sample A3 six out of six times. Also, the participant chose A1 in the presence of C3 as sample. The same was seen for the latter relation in the responding of P15761. For P15766, were all baseline and the CB symmetry relations intact, there were indeterminate responding in the B-A relations. P15761 had all baseline and symmetry relations intact. P15754 chose B3 in the presence of A2 as sample and then A2 in the presence of the B3 as sample. Additionally, A3 was chosen in the presence of C2 when testing for equivalence relations. P15542 had A2-C3, and C3-A2 as participant-defined classes.

In the GEN group did P15758 and P15769 chose A3 when C2 was presented as sample. For these participants were baseline relations intact, however, symmetry relations were analyzed as indeterminate. P15541 from the SPEC group generated two participant-defined classes. A3 was chosen in the presence of both C1 and C2 as samples. Baseline relations for this participant were intact for all relations, and three of six symmetry relations were to be considered as indeterminate. Similar results were observed for the transitivity relations.

Reaction Time

The RT for the different groups are shown in Figures 4 and 5. The analysis has been distributed between the participant assigned groups and additionally among participants who was categorized as responding or not responding in accordance with stimulus equivalence, independent of group assignment. As seen, the time for the different groups to respond to the different equivalence relations is divided into two phases, the last five training trials before testing and the first five trials in the test. The RT for responding was considered as the ongoing time from when the participant clicked on the sample and made a choice of the presented comparisons. A criterion of minimum five correct responses for each relations was set to ensure the possibility of calculating the RT. Participants who did not achieve five correct responses during the test were therefore excluded from the analysis. For the analysis across the different assigned groups (i.e., CTRL, GEN and SPEC) this resulted in one participant (P15542) in the CTRL group, two participants in the GEN group (P15769 and P15536) and one participant in the SPEC group (P15541). The RT was then calculated as the mean of median in milliseconds (ms) represented in bars. Error bars show the range of RT within the group for the different trial types.

Between participants assigned groups. Across all groups, there was a consistent pattern of increased RT in baseline, symmetry, transitivity and equivalence trials in the first five trials of testing compared to the training trials. As seen in the three panels in Figure 4, the most significant increase was observed in the transitivity relation for the control group and general instructions group with a mean of median of 6630 and, 5517 ms respectively. For the participants in the specific instruction was the mean of median RT in the transitivity relation 5628 ms, and the equivalence relation 7051 ms. The RT for the equivalence relations for the other two groups were 6010 ms (CTRL) and, 5541 ms (GEN). The mean of median RT for symmetry relations was 2716 ms for the CTRL group, 2603 ms for the GEN group and, 3873 ms for the SPEC group. Furthermore, the lowest increase was seen between the baseline relations in all groups between the last five training trials and first five testing trials with 2072 ms in training compared to 2716 ms in testing for the CTRL group, 2262 ms in training and

2603 ms in testing for the GEN instructions group and, 2464 ms in training compared to 3873 ms in testing for the SPEC group.

Responding or not responding in accordance with stimulus equivalence. As seen in Figure 5 was the RT analyzed on the basis of participants responding in testing, independent of group assignment. The participants who responded in accordance with stimulus equivalence (upper panel), showed an increased RT in the five first test trials for symmetry, transitivity and equivalence compared with the baseline trials during testing. The first five baseline trials during testing showed a small increase (2425 ms) in RT compared to the last five training trials (2608 ms).

For the participants that did not respond in accordance with stimulus equivalence (lower panel), was an increase of RT shown from the five last training trials (2204 ms) to the first five testing trials for all relations. Baseline showed an increase to 3263 ms, symmetry to 3263 ms, transitivity to 5990 ms and the equivalence relations to 6672 ms.

Overall, there is a lower RT in the baseline, transitivity and equivalence relations in the first five testing trials for the participants responding in accordance with stimulus equivalence compared to the participants who did not. However, the mean of median RT of symmetry relations was higher for the participants which responded in accordance with stimulus equivalence compared to those that did not form stimulus equivalence classes.

Discussion

The purpose of the present study was to investigate if two different types of instructions would affect equivalence responding with typical adults as participants. The main result showed that participants assigned the GEN group had the highest yield of equivalence class formation compared to the CTRL and SPEC group. The outcome of the GEN group is in accordance with the findings from Experiment 1 in the study conducted by Green et al. (1991), who used general instructions in training and testing for equivalence class formation with adults as participants. Hence, an overall interpretation of the present results suggests that the participants who received the GEN instruction were more likely to respond in accordance with stimulus equivalence compared to the other two groups (i.e., CTRL and SPEC).

Further, the result from the present study showed that the GEN group had less number of training trials to criterion followed by the SPEC and lastly the CTRL group which had the most number of training trials to criterion. Even though the results did not prove to be statistically significant between the three groups, or when the participants were categorized as responding or not responding in accordance to stimulus equivalence, are the results, with respect to number of training trials to criterion, in line with the earlier conducted studies by Arntzen et al. (2008) and Pilgrim et al. (2000). These studies investigated the acquisition of arbitrarily conditional discriminations by using general or specific instructions with typical children as participants. The main focus in those studies was, however, on the establishment and facilitation of conditional discriminations and not stimulus equivalence class formation.

At present, it seems like the results are in line with Sidman (1992, 1994) and his view regarding instructions, since, giving a general instruction, in this study, facilitated responding in accordance with stimulus equivalence for a fair number of participants. The difference between the groups was also shown to be statistically significant when conducted a FET test between the CTRL and GEN, and the SPEC and GEN group. Hence, raising questions regarding the use of instructions in the formation of equivalence classes should be considered as relevant.

Even if instructions appears to be a critical variable in stimulus equivalence research one should not forget that there are experiments where the purpose is not to find variables that affect the responding in accordance with stimulus equivalence, but rather investigates other phenomena, as for example transfer of function (eg., Hayes, Kohlenberg, & Hayes, 1991; Munnelly, Martin, Dack, Zedginidze, & McHugh, 2014). In these type of studies, the most effective procedures are often wanted. For example, by reducing the amount of time used to establish conditional discriminations. Another example might be within applied work were the effectiveness of a procedure is more important than finding the exact independent variables which control the equivalence outcome (Sidman, 1994). In these settings, giving instructions should not be considered as a problem since the verbal rules most often can short-cut sequences that would have been extensive in the first place. This could be beneficial in school settings, which are in great need of empirically based interventions, especially considering individuals with learning disabilities (Sidman, 1994).

Nevertheless, one should be careful in concluding that instructions are necessary for successful stimulus class formation with adults as participants. Instructions can rather facilitate stimulus equivalence responding if the participants can follow rules, which also was proposed as a necessity by Skinner (1969). The effects are, therefore, dependent upon the verbal competence of the participants. Moreover, even if it is tempting to conclude that instructions affect stimulus equivalence responding it must be remembered that studies investigating the emergence of derived relations also includes other parameters, results of different outcomes are, thus, connected to how the experiment has been arranged (see Arntzen, 2012 for an overview). Even though the widespread use of parameters demonstrates that stimulus equivalence is a robust phenomenon, it can lead to difficulties when we compare results between studies. Especially when investigating sensitive parameters. If the verbal history of the participants is involved in the outcome of stimulus equivalence testing, as Sidman (1992, 1994) proposes, it can be critical that experimenters use minimal or the same type of instructions across all stimulus equivalence research. If possible, no instructions at all. This would, in turn, lead to a reduction of the high use of parameter variability and interpretative difficulties between experiments when studying emergent behavior (Dymond & Rehfeldt, 2000).

General Instructions

As for the results of the GEN group it might be that the general instructions functioned as a general rule that generalized to novel stimuli in the conditional discrimination training and the stimulus equivalence testing phase. Such an interpretation was also suggested by Arntzen et al. (2008) and later argued for as a critical procedural variable in Arntzen (2012). The issue of potential rule-governed behavior during discrimination procedures was also discussed by de Rose, McIlvane, Dube, Galpin, and Stoddard (1988) who wrote about the relevance of instructions with respect to behavior that emerges without direct training, as seen in stimulus equivalence research. If such rule-governed behavior is demonstrated with adults during testing, it is then of interest to investigate if this type of emergent behavior would also be shown with atypical participants or children with a lower developed verbal repertoire. This empirical question was, therefore, investigated in three experiments by using different populations and general instructions that could be expected to develop such rules during the experimental task. The outcome showed that adults demonstrated emergent relations and that typical and atypical children to a larger degree showed emergent relations after MTS training (de Rose et al., 1988). Further studies should, however, explore whether such rules arise when using general instructions.

The interpretations of the effects of rule-governed behavior can furthermore be considered as in line with verbal behavior and its role in human action (Catania, 1995). In the article Catania (1995) discusses that instructions can be viewed as a high-order class of behavior where the verbal antecedents rather than the particular consequences are controlling the behavior. In this way, the author suggests that response classes of rule-following are established where the contingencies maintaining rule-following presumably are different from natural contingencies for each specific instance of behavior. This might be an alternative terminology of how people in general, or as the participants in the present study, demonstrated specific emergent behavior without direct training.

Specific Instructions

The choice of including the specific instruction in the experiment was based on the procedure of the earlier conducted study by Pilgrim et al. (2000). In that study, the names of stimuli were pre-determined and presented together with specific instructions. The procedure demonstrated facilitation of conditional discrimination training with children as participants. In this present study experimenter-defined names were given to one exemplar, which suggested that there was a relationship between the stimuli, and perhaps encourage anticipatory naming. Additionally, was the intention of the combination of specific instructions and pictures to make the instruction clearer than just providing it text-based. As for the specific instruction, where only one of 12 participants responded in according to stimulus equivalence, it can be argued that the instruction not worked as intended or that the effectiveness of the instruction was limited by the number of examples given. That is, the participants might have thought that the instruction was only for the current conditional relations presented in the instruction, and might, therefore, not have generated rules for the MTS task as a whole (as the speculated process in the GEN group). It could be argued that presenting more exemplars with experimenter-defined names, that was done with the SPEC group in this present study, could have facilitated the conditional discrimination training and testing by the process of naming (Dunvoll & Arntzen, under review). Although it is speculative, further studies are required to investigate this notion, which also is challenging to demonstrate experimentally, with respect to potentially covert verbal behavior. One possible solution would, however, be to implement talk aloud procedures (e.g., Arntzen, Halstadtro, & Halstadtro, 2009) which was demonstrated in a recent study by Vie and Arntzen (2017), who investigated what participants were saying aloud (i.e., thinking) during the conditional discrimination procedures and testing for stimulus equivalence.

Procedural issues. The chosen names in the specific instruction (upper panel in Figure 2) were to be considered as arbitrarily related. The car and book, as used in the specific instruction could, however, be considered as familiar depending on the learning history of the participants. It should, therefore, be added that it can be problematic to choose names which in some form is not considered as arbitrary related. A suggestion for future research may be to use nonsense words. Alternatively, include familiar related stimuli with the use of names that most typically developed participants would consider as belonging together, for example, a paper and a pen. If these suggestions facilitates the conditional discrimination training and the testing for equivalence formation are however empirical questions.

Another procedural issue with the specific instruction was that the information that was given could be considered as a combined instruction. Meaning, that the specific instruction in the current study both included a specific and general instruction, by naming the stimuli, and defining the relationship between those two. Future research might benefit of including an extra category such as a combined instruction, thus, isolating the specific instruction (e.g., when this stimulus is here, choose this [pointing to stimuli]) without specifying the relationship between the stimuli. Another alternative could be to adjust the specific instruction by for example keeping the instruction as is but following it with extra information such as “this is an example of how an exemplar in this experiment could look like” and then see if this added information would affect the equivalence outcome. If such information affects the outcome, it might be argued that the instructions works for other exemplars in training and testing for stimulus equivalence.

Quizzes. It is also plausible that the participants in the SPEC group did not read or comprehend the instructions, which could have had an impact on the results. However, the results of the quizzes showed that the participants scored correctly on all questions related to the instructions on the first or second trial. Thus, should not be seen as a significant threat to the study in general.

Control Group

As for the CTRL group where only one participant responded in accordance with stimulus equivalence, it could be argued that the combination of LS training structure and the SIM protocol were the reason of lower yields. These findings are in line with other studies which have shown or discussed that the combination of LS and SIM typically produces poor stimulus equivalence outcomes (Arntzen, Nartey, et al., 2015; Arntzen, Norbom, et al., 2015; Imam & Blanche, 2013; Imam & Warner, 2014). The training structure and protocol used, may also have been a contributing factor to the low equivalence responding in the SPEC group. Hence, the results in the GEN group can, therefore, be seen as strengthened. Future research should, therefore, include other combinations of training structures and protocols, and also include more class members to see if the effect of instructions is as apparent as in this present study.

Instructions and Timing

Even if the investigation of different types of instructions in stimulus equivalence research is of great importance, one should not forget another influential variable such as the timing. This was also considered in the study by Arntzen et al. (2008) who also discussed that timing the instruction after participants had failed to meet mastery criterion might make the instruction more effective. Another interesting idea would be to investigate how the instructions affect the results by programming the MTS software to present instructions at different times during the experiment. For example, in the middle of the experiment or after a certain amount of errors during training (C. Miguel, personal communication, April 27, 2018). The instructions in the present study were considered as start-up instructions, that is, presented before the MTS task started. Since all participants, in this study, passed the mastery criterion these potential effects would rather have been seen with respect to the number of training trials to criterion.

Sorting

The post-class formation sorting test was included in the analysis to examine a potential correspondence with the performance in MTS-formatted derived-testing, as have been shown in earlier conducted experiments (Arntzen, Norbom, et al., 2015; Fields, Arntzen, & Moksness, 2014; Nartey, Arntzen, & Fields, 2014). The results showed, for 13 participants, that when experimenter-defined classes were formed after the MTS testing, the post-class formation sorting test had identical class maintenance. It should be noted that eleven of these 13 participants were assigned the GEN group indicating that the instruction could have facilitated the sorting test in the same way as in the MTS test. The four participants who did not respond in accordance with stimulus equivalence did neither show the correct classes in the post-class formation test. However, for the 19 participants assigned to either the CTRL of SPEC group who did not form classes in the MTS testing, showed correct experimenter-defined class maintenance during the sorting test. This latter finding is not in line with earlier findings. Rather, the results raise questions if sorting can be set equal to class formation in an MTS-formatted test. The reason for the difference may be that this study included three-member classes and used an abstract stimulus set while other studies have used different procedural arrangements such as the inclusion of meaningful stimuli and more members (Fields et al., 2014; Nartey et al., 2014). An interpretation can be that smaller stimulus classes might be a reason why the participants managed to sort correctly in the post-class formation tests.

Mastery to Criterion

The issue of mastery to criterion for conditional discriminations and to meet the criterion of responding in accordance with stimulus equivalence was discussed by Arntzen (2012) and it appears that the use of criterion in both training and testing is widespread across type of research and experiments conducted. In the present study, the mastery criterion for training was set to 100% to ensure that the participants had well-established the nine conditional discriminations. A lower criterion would have risked that the participant got through the training with incorrect trials, for example by having one or two incorrect trial types through the training, which might have affected the performance in the testing condition. A 90% testing criterion was used in this study. This is considered as a customary criterion which makes it preferably easier to compare the results across conducted experiments (Arntzen, 2012). However, it should be noted that some of the participants in the study performed just below the criterion for defining the responding in accordance with stimulus equivalence. This was seen with P15535 and P15753 in the CTRL group, and P15757 in the SPEC group. These participants had one error from responding in accordance with the mastery criterion in the testing condition (i.e., 89% correct responding) in either the symmetry, transitivity or equivalence trials.

Reaction Time and Response Pattern

This study included RT as an additional measure since it can potentially lead to a more fine-grained analysis of the formation of stimulus equivalence classes (Holth & Arntzen, 2000; Spencer & Chase, 1996). The analysis showed that the RT for the five last trained relations were shorter than the five first derived relations at the beginning of the testing condition. These results are in line with earlier conducted stimulus equivalence studies, even those including variation of parameters in the experiments, as for example the investigation training structures or type of stimuli (Arntzen, Braaten, Lian, & Eilifsen, 2011; Arntzen, Grondahl, & Eilifsen, 2010; Arntzen & Lian, 2010). Further, have studies showed that increased RT in testing can be connected to the relational type tested (Spencer & Chase, 1996). That is, showing longer RT to TRA and EQ relations than the SYM and BL trials during stimulus equivalence testing. This finding was shown in this present study between the CTRL, GEN and SPEC group, and also when participants were grouped and categorized as responding in accordance with stimulus equivalence or not. Interestingly, results in the GEN group showed that the RT was slightly lower compared to the other two groups as for the BL, TRA and EQ trials in testing. This result may be attributed to the use of general instruction where a general rule might have affected the RT.

For participants that did not respond in accordance with stimulus equivalence (Figure 5), it is possible the outcome was a result of a conflict of the inconsistency of stimulus topographies. That is, the participants may have had problems with choosing the correct experimenter-defined topography (i.e., comparison) since there would be many stimulus control variations to choose from (e.g., McIlvane & Dube, 2003; Mensah & Arntzen, 2017). This notion was also strengthened by the results of the participant's response patterns (Figure 3), where 29% in the CTRL group and 32% in the SPEC group showed an indeterminate pattern of responding. Thus, not forming participant-defined nor experimenter-defined classes. The analysis further showed that the participants across the three groups had more errors with the TRA and EQ relations, which also can explain the increased RT. However, it can only be speculated why these relations produced more errors than the others. One answer can be that the relations are derived. Thus, the participant did not have direct contact with the relations in training. Another answer might be that the training was not sufficient for the participants to form the equivalence classes. Further, the result of experimenter-defined classes proved to be highest in the GEN group, followed by the SPEC, and lastly the CTRL group. Few participants showed participant-defined classes. These results are in line with the earlier conducted study by Mensah and Arntzen (2017).

Conclusion

This study demonstrated that general (GEN) instructions produced a higher yield of equivalence class formation, compared to specific (SPEC) or no instructions at all (CTRL), with adults as participants. These results support the notion by Sidman (1992, 1994) that one must be careful with the use of instructions when conducting this type of research, since instructions seem to affect equivalence outcome. Thus, it might be argued that researchers within the field of stimulus equivalence should have a more parsimonious approach to instructions – which may make it easier to compare results across studies and between different laboratories.

References

Arntzen, E. (2012). Training and testing parameters in formation of stimulus equivalence: Methodological issues. *European Journal of Behavior Analysis, 13*(1), 123–135. doi:10.1080/15021149.2012.11434412

Arntzen, E., Braaten, L. F., Lian, T., & Eilifsen, C. (2011). Response-to-sample requirements in conditional discrimination procedures. *European Journal of Behavior Analysis, 12*(2), 505–522. doi:10.1080/15021149.2011.11434398

Arntzen, E., Grondahl, T., & Eilifsen, C. (2010). The effects of different training structures in the establishment of conditional discriminations and subsequent performance on tests for stimulus equivalence. *The Psychological Record, 60*(3), 437–462. doi:10.1007/bf03395720

Arntzen, E., Halstadtro, L.-B., & Halstadtro, M. (2009). The "silent dog" method: Analyzing the impact of self-generated rules when teaching different computer chains to boys with autism. *The Analysis of Verbal Behavior, 25*(1), 51–66. doi:10.1007/bf03393069

Arntzen, E., & Lian, T. (2010). Trained and derived relations with pictures versus abstract stimuli as nodes. *The Psychological Record, 60*(4), 659–678. doi:10.1007/bf03395738

Arntzen, E., Nartey, R. K., & Fields, L. (2015). Enhanced equivalence class formation by the delay and relational functions of meaningful stimuli. *Journal of the Experimental Analysis of Behavior, 103*(3), 524–541. doi:10.1002/jeab.152

Arntzen, E., Norbom, A., & Fields, L. (2015). Sorting: an alternative measure of class formation? *The Psychological Record, 65*(4), 615–625. doi:10.1007/s40732-015-0132-5

Arntzen, E., Vaidya, M., & Halstadtro, L. B. (2008). Brief report: On the role of instruction in conditional discrimination training. *Expermental Analysis of Human Behavior Bulletin, 29*, 17–24. Retrieved from http://www.eahb.org/

Bush, K. M., Sidman, M., & de Rose, T. (1989). Contextual control of emergent equivalence relations. *Journal of the Experimental Analysis of Behavior, 51*(1), 29–45. doi:10.1901/jeab.1989.51-29

Catania, A. C. (1995). Higher-order behavior classes: Contingencies, beliefs, and verbal behavior. *Journal of Behavior Therapy and Experimental Psychiatry, 26*(3), 191–200. doi:10.1016/0005-7916(95)00033-v

Catania, A. C. (2013). *Learning* (5th ed.). New York, NY: Sloan Publishing.

Catania, A. C., Shimoff, E., & Matthews, B. A. (1989). An experimental analysis of rule-governed behavior In S. C. Hayes (Ed.), *Rule-governed behavior: cognition, contingencies and instructional control* (pp. 119–150). Reno, NV: Context Press.

de Almeida, J. H., & de Rose, J. C. (2015). Changing the meaningfulness of abstract stimuli by the reorganization of equivalence classes: effects of delayed matching. *The Psychological Record, 65*(3), 451–461. doi:10.1007/s40732-015-0120-9

de Rose, J. C., McIlvane, W. J., Dube, W. V., Galpin, V. C., & Stoddard, L. T. (1988). Emergent simple discrimination established by indirect relation to differential consequences. *Journal of the Experimental Analysis of Behavior, 50*(1), 1–20. doi:10.1901/jeab.1988.50-1

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:10.1901/jeab.1986.46-243

Dunvoll, G., G., & Arntzen, E. (under review). Effect of individual and common naming of stimuli in equivalence classes formation in children.

Dymond, S., & Rehfeldt, R. A. (2000). Understanding complex behavior: The transformation of stimulus functions. *The Behavior Analyst, 23*(2), 239–254. doi:10.1007/bf03392013

Fields, L., Arntzen, E., & Moksness, M. (2014). Stimulus sorting: A quick and sensitive index of equivalence class formation. *The Psychological Record, 64*(3), 487–498. doi:10.1007/s40732-014-0034-y

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

Green, G., Sigurdardottir, Z. G., & Saunders, R. R. (1991). The role of instructions in the transfer of ordinal functions through equivalence classes. *Journal of the Experimental Analysis of Behavior, 55*(3), 287–304. doi:10.1901/jeab.1991.55-287

Hayes, S. C. (1989). *Rule-governed behavior: cognition, contingencies, and instructional control*. New York: Plenum.

Hayes, S. C., Kohlenberg, B. S., & Hayes, L. J. (1991). The transfer of specific and general consequential functions through simple and conditional equivalence relations. *Journal of the Experimental Analysis of Behavior, 56*(1), 119–137. doi:10.1901/jeab.1991.56-119

Holth, P., & Arntzen, E. (2000). Reaction times and the emergence of class consistent responding: A case for precurrent responding? *The Psychological Record, 50*(2), 305–337. doi:10.1007/bf03395358

Imam, A. A., & Blanche, J. V. (2013). Effects of a CARO on stimulus equivalence formation: a systematic replication. *The Psychological Record, 63*(1), 141–156. doi:10.11133/j.tpr.2013.63.1.011

Imam, A. A., & Warner, T. A. (2014). Test order effects in simultaneous protocols. *Learning & Behavior, 42*(1), 93–103. doi:10.3758/s13420-013-0128-4

McIlvane, W. J., & Dube, W. V. (2003). Stimulus control topography coherence theory: Foundations and extensions. *The Behavior Analyst, 26*(2), 195–213. doi:10.1007/bf03392076

Mensah, J., & Arntzen, E. (2017). Effects of meaningful stimuli contained in different numbers of classes on equivalence class formation. *The Psychological Record, 67*(3), 325–336. doi:10.1007/s40732-016-0215-y

Munnelly, A., Martin, G., Dack, C., Zedginidze, A., & McHugh, L. (2014). The transfer of social exclusion and inclusion functions through derived stimulus relations. *Learning & Behavior, 42*(3), 270–280. doi:10.3758/s13420-014-0144-z

Nartey, R. K., Arntzen, E., & Fields, L. (2014). Two discriminative functions of meaningful stimuli that enhance equivalence class formation. *The Psychological Record, 64*(4), 777–789. doi:10.1007/s40732-014-0072-5

Pilgrim, C., Jackson, J., & Galizio, M. (2000). Acquisition of arbitrarily conditional discriminations by young normally developing children. *Journal of Experimental Analysis of Behavior, 73*(2), 177–193. doi: 10.1901/jeab.2000.73-177

Schlinger, H. D. (1993). Separating discriminative and function-altering effects of verbal stimuli. *The Behavior Analyst, 16*(1), 9–23. doi:10.1007/bf03392605

Schlinger, H. D., & Blakely, E. (1987). Function-altering effects of contingency-specifying stimuli. *The Behavior Analyst, 10*(1), 41–45. doi:10.1007/bf03392405

Sidman, M. (1992). Equivalence relations: Some basic considerations. In S. C. Hayes & L. J. Hayes (Eds.), *Understanding verbal relations* (pp. 15–27). Reno, NV: Context Press.

Sidman, M. (1994). *Equivalence relations and behavior: a research story*. Boston: Authors cooperative.

Sidman, M. (2000). Equivalence relations and the reinforcement contingency. *Journal of the Experimental Analysis of Behavior, 74*(1), 127–146. doi:10.1901/jeab.2000.74-127

Sidman, M., Wynne, C. K., Maguire, R. W., & Barnes, T. (1989). Functional classes and equivalence relations. *Journal of the Experimental Analysis of Behavior, 52*(3), 261–274. doi:10.1901/jeab.1989.52-261

Silveira, M. V., Mackay, H. A., & de Rose, J. C. (2017). Measuring the "transfer of meaning" through members of equivalence classes merged via a class-specific reinforcement procedure. *Learning & Behavior*, 1–14. doi:10.3758/s13420-017-0298-6

Skinner, B. F. (1969). *Contingencies of reinforcement: a theoretical analysis*. New York: Appleton-Century-Crofts.

Spencer, T. J., & Chase, P. N. (1996). Speed analyses of stimulus equivalence. *Journal of the Experimental Analysis of Behavior, 65*(3), 643–659. doi:10.1901/jeab.1996.65-643

Vie, A., & Arntzen, E. (2017). Talk-Aloud Protocols during Conditional Discrimination Training and Equivalence Class Formation. *The Analysis of Verbal Behavior, 33*(1), 80–97. doi:10.1007/s40616-017-0081-y

Table 1
Overview of the procedure

Condition	Trial types	PPC %	Criterion
Concurrent training	A1/B1B2B3, A2/B1B2B3, A3/B1B2B3, B1/C1C2C3, B2/C1C2C3, B3/C1C2C3	100	18/18
	A1/B1B2B3, A2/B1B2B3, A3/B1B2B3, B1/C1C2C3, B2/C1C2C3, B3/C1C2C3	75	18/18
	A1/B1B2B3, A2/B1B2B3, A3/B1B2B3, B1/C1C2C3, B2/C1C2C3, B3/C1C2C3	50	18/18
	A1/B1B2B3, A2/B1B2B3, A3/B1B2B3, B1/C1C2C3, B2/C1C2C3, B3/C1C2C3	0	
Mixed testing	A1/B1B2B3, A2/B1B2B3, A3/B1B2B3, (BL) B1/C1C2C3, B2/C1C2C3, B3/C1C2C3 B1/A1A2A3, B2/A1A2A3, B3/A1A2A3	0	
	C1/B1B2B3, C2/B1B2B3, C3/B1B2B3 (SYM) A1/C1C2C3, A2/C1C2C3, A3/C1C2C3 (TRA) C1/A1A2A3, C2/A1A2A3, C3/A1A2A3 (EQ)		

Note. The table shows different conditions and trial types, additionally the probability for programmed consequence (PPC) and mastery criterion for each condition. The test block consisted of 108 intermixed trials each (36 BL, 36 SYM, 18 TRA and 18 EQ trials). BL = baseline relations; SYM = symmetry; TRA = transitivity; EQ = equivalence.

Table 2

Overview of individual data for conditional discrimination training, equivalence testing, and post-class formation sorting

CTRL group

			Training							Test				Post-sorting					
Participant	**Gender**	**Age**	**Trials 100 %**	**Error**	**Trials 75 %**	**Error**	**Trials 50 %**	**Error**	**Total**	**BL**	**SYM**	**TRA**	**EQ**	**1**			**2**		
15543	F	51	288	66	18	0	18	0	342	**100%**	**100%**	**94%**	**94%**	**300**	**030**	**003**	**300**	**030**	**003**
15535	F	20	180	70	108	11	18	0	324	**100%**	89%	**94%**	**94%**	**300**	**030**	**003**	**300**	**030**	**003**
15753	F	42	432	146	90	10	18	0	558	**100%**	**97%**	**94%**	89%	**300**	**030**	**003**	**300**	**030**	**003**
15529	F	26	216	67	18	0	18	0	270	**97%**	**100%**	89%	83%	**300**	**030**	**003**	**300**	**030**	**003**
15752	F	35	234	100	18	0	18	0	288	**100%**	**94%**	83%	67%	**300**	**030**	**003**	**300**	**030**	**003**
15761	F	22	144	51	18	0	54	3	234	**100%**	**94%**	56%	56%	**300**	**030**	**003**	**300**	**030**	**003**
15766	M	22	126	35	18	0	18	0	180	**100%**	64%	61%	33%	**300**	**030**	**003**	**300**	**030**	**003**
15546	F	20	270	177	18	0	18	0	342	**97%**	83%	56%	39%	111	111	111	111	111	111
15547	F	20	252	137	36	1	18	0	324	89%	**100%**	56%	50%	**300**	**030**	**003**	**300**	**030**	**003**
15528	F	20	378	139	36	2	18	0	450	72%	61%	44%	39%	**300**	**030**	**003**	**300**	**030**	**003**
15542	F	21	198	83	18	0	18	0	252	89%	78%	39%	22%	**300**	**030**	**003**	**300**	**030**	**003**
15754	F	20	144	59	18	0	18	0	198	69%	69%	50%	33%	111	111	111	111	111	111
Median			**225**	**76.5**	**18**	**0**	**18**	**0**	**306**										

GEN group

			Training							Test				Post-sorting					
Participant	**Gender**	**Age**	**Trials 100 %**	**Error**	**Trials 75 %**	**Error**	**Trials 50 %**	**Error**	**Total**	**BL**	**SYM**	**TRA**	**EQ**	**1**			**2**		
15545	F	25	126	52	18	0	18	0	180	**100%**	**100%**	**100%**	**100%**	**300**	**030**	**003**	**300**	**030**	**003**
15762	F	35	126	39	18	0	18	0	180	**100%**	**100%**	**100%**	**100%**	**300**	**030**	**003**	**300**	**030**	**003**
15534	F	20	108	43	18	0	18	0	162	**100%**	**100%**	**100%**	**94%**	**300**	**030**	**003**	**300**	**030**	**003**
15531	F	20	252	90	18	0	18	0	306	**100%**	**100%**	**94%**	**100%**	**300**	**030**	**003**	**300**	**030**	**003**
15544	F	44	378	176	54	4	18	0	468	**100%**	**97%**	**94%**	**94%**	**300**	**030**	**003**	**300**	**030**	**003**
15539	M	22	198	79	18	0	18	0	252	**97%**	**97%**	**94%**	**94%**	**300**	**030**	**003**	**300**	**030**	**003**
15532	F	20	198	52	18	0	36	2	270	**94%**	**100%**	**100%**	**94%**	**300**	**030**	**003**	**300**	**030**	**003**
15549	F	22	252	62	18	0	18	0	306	**97%**	**97%**	**100%**	67%	**300**	**030**	**003**	**300**	**030**	**003**
15758	F	35	180	80	18	0	18	0	234	**100%**	**94%**	72%	78%	**300**	**030**	**003**	**300**	**030**	**003**
15769	F	21	270	97	90	5	18	0	396	**100 %**	53 %	22 %	11 %	**300**	**030**	**003**	**300**	**030**	**003**
15759	F	26	414	148	18	0	18	0	468	89%	81%	56%	33%	**300**	**030**	**003**	**300**	**030**	**003**
15536	M	22	162	54	18	0	18	0	216	83%	86%	17%	6%	113	020	200	213	020	100
Median			**198**	**70.5**	**18**	**0**	**18**	**0**	**261**										

SPEC group

			Training							Test				Post-sorting					
Participant	**Gender**	**Age**	**Trials 100%**	**Error**	**Trials 75%**	**Error**	**Trials 50%**	**Error**	**Total**	**BL**	**SYM**	**TRA**	**EQ**	**1**			**2**		
15768	M	23	108	40	36	1	50	0	216	**100%**	**100%**	**100%**	**100%**	**300**	**030**	**003**	**300**	**030**	**003**
15550	M	24	162	56	36	4	18	0	234	**100%**	**97%**	**94%**	83%	**300**	**030**	**003**	**300**	**030**	**003**
15756	F	24	270	144	72	3	18	0	378	**97%**	**100%**	**100%**	78%	**300**	**030**	**003**	**300**	**030**	**003**
15757	F	33	108	26	18	0	18	0	162	**92%**	**92%**	89%	**100%**	**300**	**030**	**003**	**300**	**030**	**003**
15540	F	19	198	52	18	0	18	0	252	**100%**	**100%**	61%	72%	**300**	**030**	**003**	**300**	**030**	**003**
15537	F	21	324	120	18	0	18	0	378	**100%**	**92%**	78%	67%	**300**	**030**	**003**	**300**	**030**	**003**
15767	F	24	882	483	18	0	36	3	972	**92%**	**100%**	62%	62%	**300**	**030**	**003**	**300**	**030**	**003**
15760	F	28	378	161	18	0	18	0	432	**94%**	**94%**	39%	50%	**300**	**030**	**003**	**300**	**030**	**003**
15541	F	19	216	100	18	0	18	0	288	**100%**	81%	17%	6%	**300**	**030**	**003**	**300**	**030**	**003**
15533	M	19	180	53	18	0	18	0	234	**97%**	89%	83%	89%	**300**	**030**	**003**	**300**	**030**	**003**
15530	F	19	216	57	18	0	18	0	270	**92%**	69%	72%	72%	310	020	**003**	310	020	**003**
15538	F	23	342	158	36	1	36	1	486	89%	83%	28%	39%	**300**	**030**	**003**	**300**	**030**	**003**
Median			**216**	**78.5**	**18**	**0**	**18**	**0**	**279**										

Note. BL = baseline relations; SYM = symmetry; TRA = transitivity; EQ = equivalence.

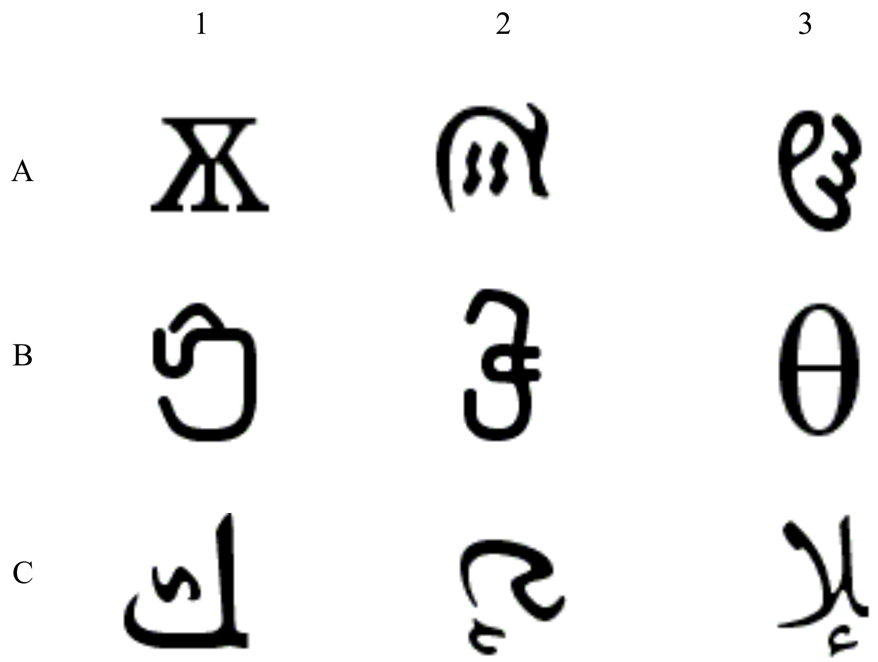

Figure 1. The stimulus set used in the experiment. The numbers refer to the experimenter-defined classes and the letters to the class members.

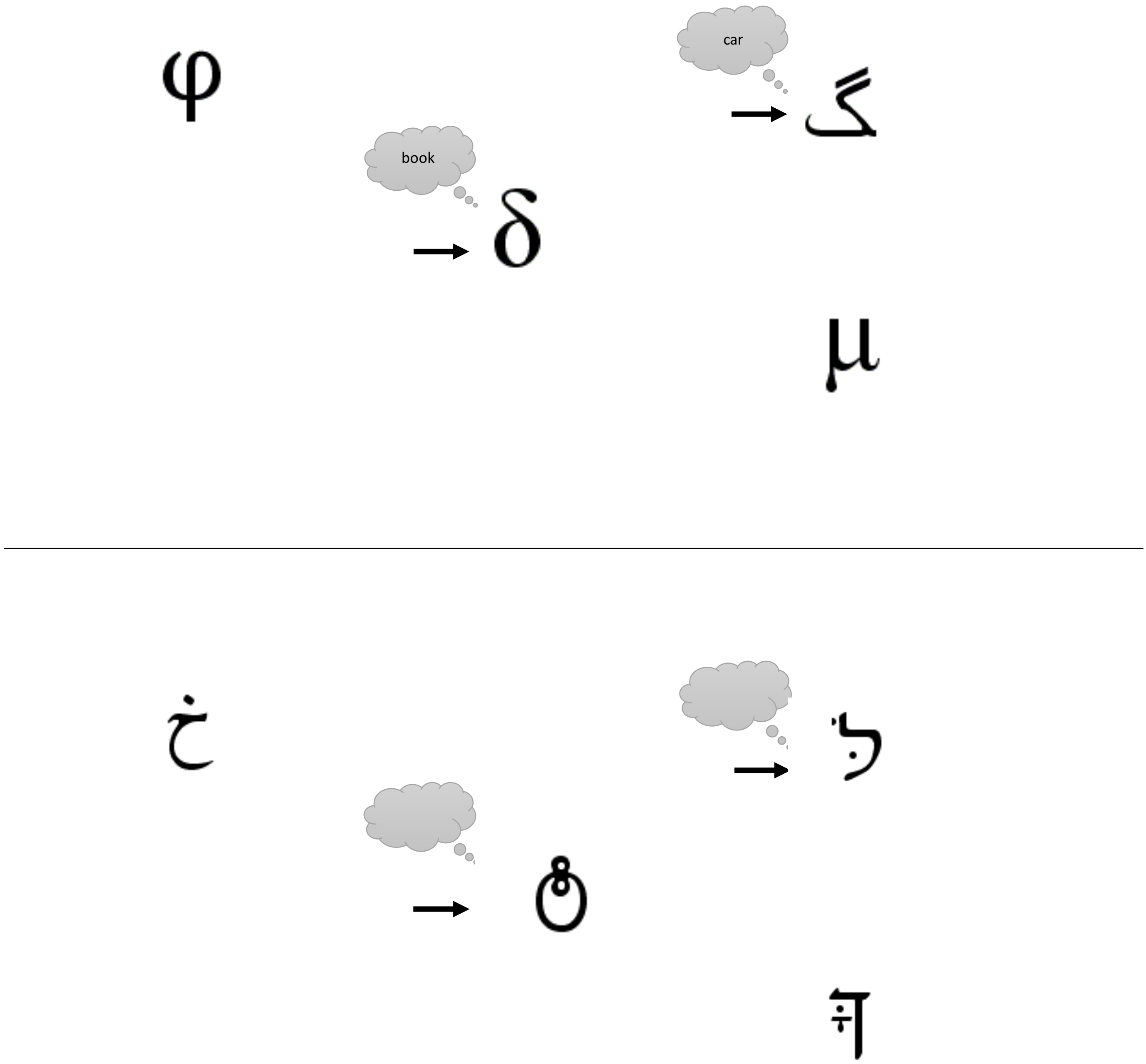

Figure 2. Presented pictures for the SPEC group. The upper panel was first presented and shows the already named stimuli. The lower panel illustrates the sheet where the participants were told to self-generate names for the stimuli.

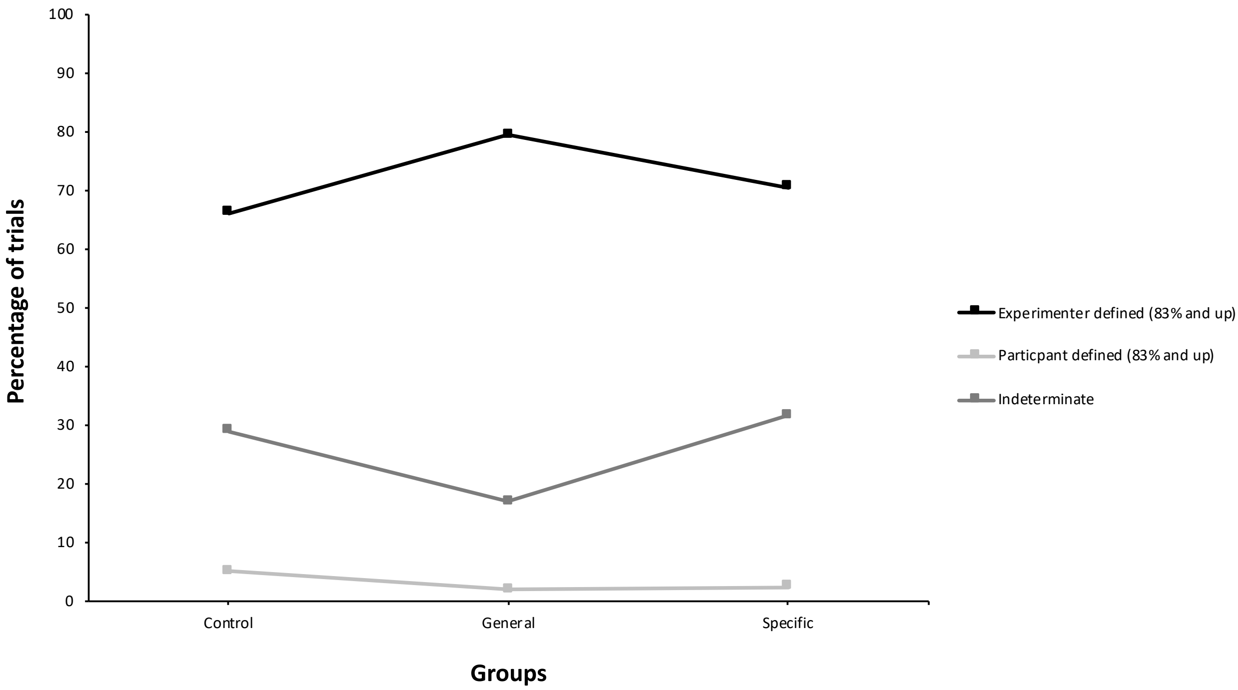

Figure 3. Results of type of defined stimulus equivalence classes by participants in each respective assigned group presented in percentage.

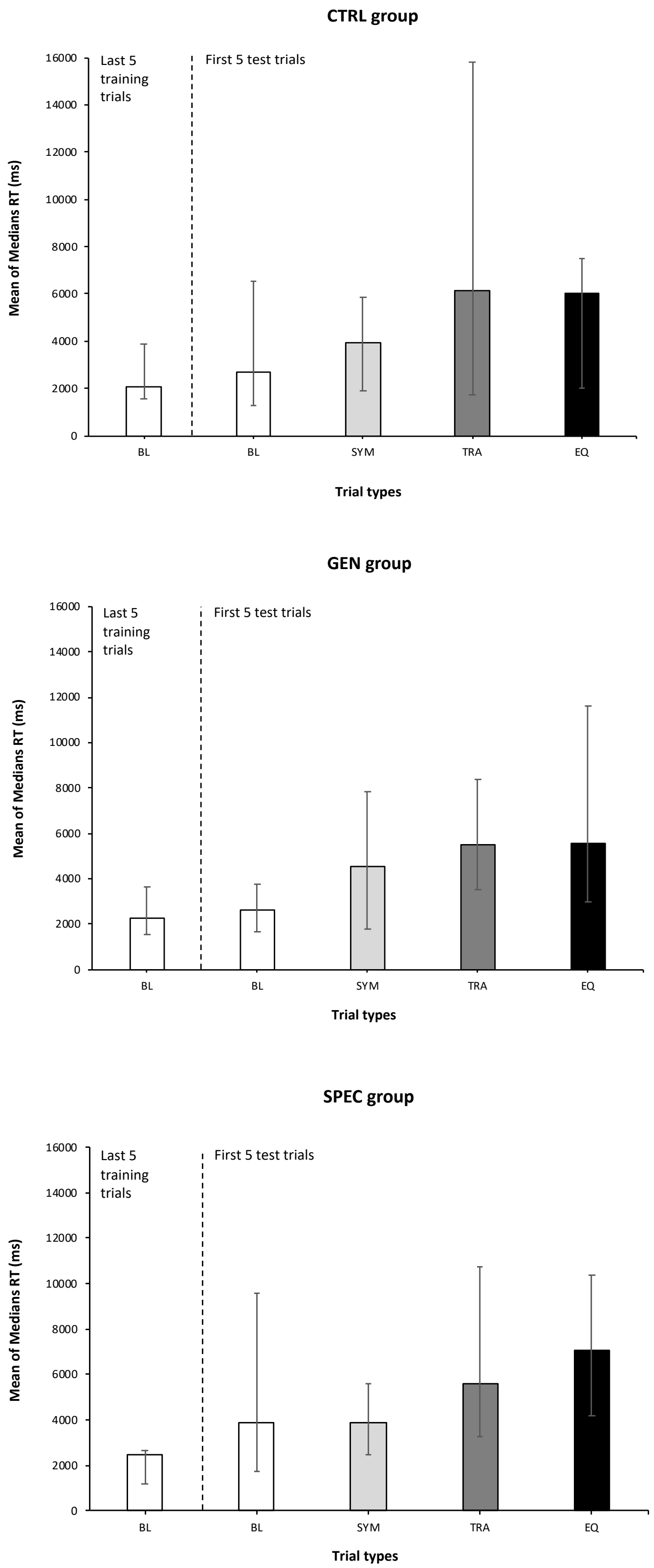

Figure 4. The figure shows mean of median reaction time (RT) in milliseconds (ms) for the last five training trials and five first test trials for each participant assigned group. Error bars show highest and lowest RT for each respective relation in the three groups. BL = baseline relations; SYM = symmetry; TRA = transitivity; EQ = equivalence.

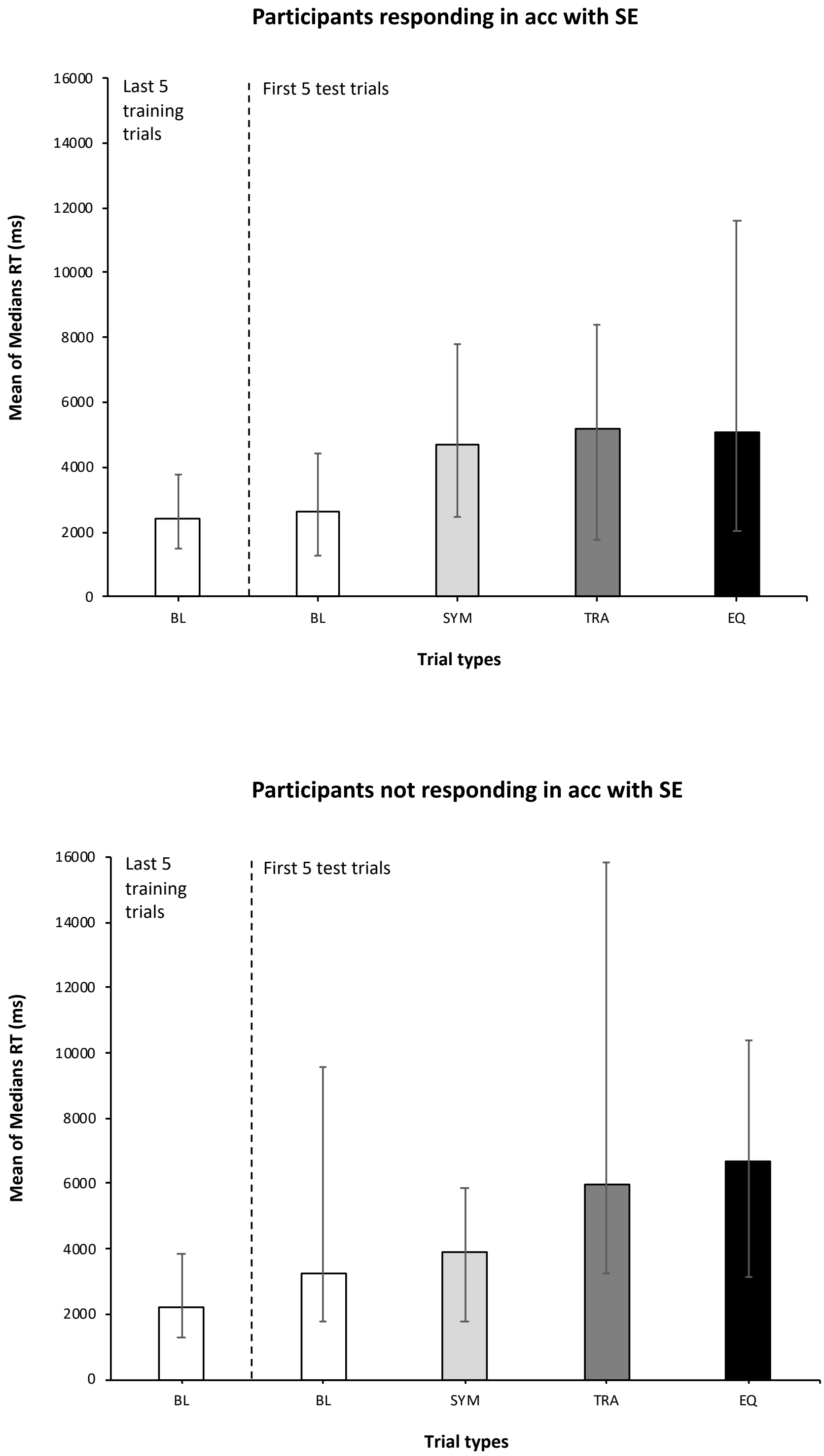

Figure 5. The figure shows mean of median reaction time (RT) in milliseconds (ms) for the last five training trials and five first test trials for participants responding and not responding in accordance with stimulus equivalence. Error bars show highest and lowest RT for each respective relation in both groups. BL = baseline relations; SYM = symmetry; TRA = transitivity; EQ = equivalence.