STIMULUS EQUIVALENCE:

A COMPARISON OF OPERANT AND ASSOCIATIVE PROCEDURES.

A thesis

submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy in Psychology

at the

University of Waikato

by

Jennifer May Kinloch

University of Waikato

2011

ABSTRACT

Previous studies comparing the effectiveness of the stimulus-pairing-observation and matching-to-sample procedures in facilitating the formation of equivalence relations have had conflicting findings. In an attempt to clarify the reasons for this, Experiment 1 replicated one of the experiments from Leader and Barnes-Holmes (2001b) but with the Chinese characters used by Clayton and Hayes (2004) as stimuli. The adult participants completed both the stimulus-pairing-observation and matching-to-sample procedures. Neither procedure was found to be more effective than the other, with few of the participants demonstrating equivalence after either procedure. Due to the failure of most participants to demonstrate equivalence, Experiment 2 replicated Experiment 1 but with the original nonsense syllables used by Leader and Barnes-Holmes (2001b). Equivalence was not demonstrated by any of the participants in Experiment 2. Therefore, the failures in Experiment 1 were not the result of the stimuli used. The use of the same stimuli in conflicting relations was identified as the most likely cause. Experiment 3 addressed this by using different nonsense syllables with each procedure. This resulted in greater accuracy on both the symmetry and equivalence tests compared to the earlier experiments; however, none of the participants demonstrated equivalence, and the procedures did not differ in their effectiveness. In Experiments 4 to 6 participants experienced either the stimulus-pairing observation or matching-to-sample procedures. These three experiments examined the effect of instructional specificity, stimulus arrangement, and the number of training trials on the effectiveness of these two procedures. Experiment 4 found that instructions which outlined the task required more specifically increase the effectiveness of both procedures marginally, and that a larger number of training and testing cycles (compared to e 1-3) did not aid in the development of equivalence. Experiment 5 examined the effectiveness of the many-to-one or one-to-many stimulus arrangements (compared to the linear arrangement used in the earlier experiments). The many-to-one and one-to-many arrangements resulted in more participants demonstrating equivalence than the linear arrangement for both the stimulus-pairing-observation and matching-to-sample procedures. Experiment 6 replicated E 5 but with more training trials prior to each equivalence test. This resulted in more participants demonstrating equivalence across both procedures and all stimulus arrangements. The stimulus-pairing-observation and matching-to-sample procedures were found to be equally effective in terms of accuracy achieved on the equivalence tests; however, the matching-to-sample procedure resulted in the development of equivalence within fewer training trials than the stimulus-pairing-observation procedure. When the stimulus-pairing-observation procedure was used, more participants demonstrated equivalence with the one-to-many arrangement than with the many-to-one or linear arrangements. When the matching-to-sample procedure was used, the one-to-many and many-to-one arrangements resulted in more participants demonstrating equivalence than the linear arrangement. Comparisons across the experiments suggested that the number of training trials completed prior to each equivalence test, but not the total number of training trials completed, affected performance. The effectiveness of the stimulus arrangements differed across the procedures, but one-to-many arrangement was more effective than the linear arrangement for both procedures. Overall, these experiments suggest that there is little difference in the effectiveness of the MTS and SPO procedures in facilitating the formation of equivalence relations, and that the development of equivalence is made more likely for both procedures by the addition of more training trials prior to each test, and the use of a one-to-many stimulus arrangement.

ACKNOWLEDGEMENTS

Many thanks to Prof. Mary Foster and Dr. James McEwan for their continued guidance, and support, for encouraging me to do this PhD, and for providing me with many opportunities during my time at Waikato University. Thank you to Andrew Malcolm for writing the computer programmes for my experiments, and to Allan Eaddy for his assistance with the many computers involved. Thank you to my colleagues and fellow students for their friendship, support, and encouraging words, and for helping to reduce my workload during the final months of writing. Special thanks to Susan Yates for her assistance with proof-reading this thesis. Also special thanks to Clare Browne for her help with data collection and proof-reading. My appreciation goes to my family and friends for all of their support and encouragement over many years. Thanks to the Tertiary Education Commission of New Zealand for their generous scholarship. Finally, thank you to all of the people who gave up their time to participate in this study.

In memory

Kenneth Joseph Keen

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICIES

When a young child is taught the relation between an actual cat and the spoken word ‘cat’, and then subsequently taught the relation between the spoken word ‘cat’ and the written word ‘cat’, they will derive the relation between the actual cat and the written word ‘cat’ without the occurrence of any explicit training or direct association. These types of relations are called arbitrarily derived relations. They are considered to be arbitrary relations as they are not based on the physical properties of the stimuli, but are related arbitrarily. For example, the value of money is arbitrary. It is not based on the physical properties of the notes and coins, but on what it will buy. We learn that five 20c coins have the same value as two 50c coins or one $1 coin even though the five 20c coins involve more physical objects, or that three 50 c coins has less value even though there are more of them than one $2 coin.

Arbitrarily derived relations provide a framework by which the function of words, objects, events, or other stimuli can be evoked by other words, objects, events or other stimuli. In the examples given above, the actual cat and the written word ‘cat’ or the five 20c and two 50c coins come to share the same function. That is, they come to mean the same thing or have the same value. This is known as transfer of function.

Arbitrarily derived relations have been used to explain the development of emotional reactions to words, events, or objects that have not previously been encountered or that have never been explicitly paired with an aversive event (Dougher, Auguston, Markham, & Greenway, 1994). By this argument, many human psychological problems are the result of the inadvertent pairing or associating of objects or stimuli with each other. Blackledge (2003) gives the example that a person who reports a fear of snakes may experience a physiological fear reaction when presented with a stimulus or situation that has been associated with snakes, e.g., a forested area or movement in the undergrowth (Blackledge, 2003), or when presented with stimuli that share some of the same physical properties as a snake, e.g., a long thin stick or a picture of a snake. If these events or stimuli are then associated with other stimuli that are unrelated to snakes, then these may also come to evoke the fear reaction originally elicited in the presence of snakes. However, in this example, the person does not need to have had any direct experience with snakes. As the stimuli involved in derived relations can include words, being told that snakes are dangerous can result in the same physiological reaction and behaviours when presented with objects or situations as a directly experienced aversive experience with a snake.

Therefore, a person can report a fear of snakes (or many other stimuli) without having ever encountered an actual snake. The words themselves can also come to evoke the physiological reactions and behaviours; therefore, simply talking about snakes may evoke the same physiological reaction and behaviours as would the presence of an actual snake.

Arbitrarily derived relations also underpin other human emotional behaviours. Dixon et al. (2006) applies the principle of derived relations to the development of behaviours that are often said to be the result of prejudice or stigma. They give an example they argue accounts for the prejudicial attacks and behaviours that have occurred towards innocent people of Middle Eastern descent in the United States of America since the terrorist attacks in September, 2001. The example they outline is of an American who while watching TV learns that there have been attacks on his country, and responds angrily. The news reader then labels the people responsible for the attack as terrorists and shows the pictures of the people responsible, reporting their ethnicity and religious affiliation. Within that one news report the attacks that evoked the initial anger response have been associated with the label terrorist, and the label terrorist with the pictures of the men and the reported demographics. The initial anger reaction may now come to be evoked by the physical characteristics, ethnicity, and religious affiliation reported in the news report even though they have not been paired directly. However, these characteristics and demographics are shared by a large number of innocent people, some of whom live in the same country. Therefore, attacks or discriminatory behaviours that are attributed to prejudice or stigma may be the result of arbitrarily derived relations and transfer of function.

In summary, arbitrarily derived relations are those relations that emerge between words, events, objects, or other stimuli without explicit training or direct association. These relations involve the transfer of the directly learned function of stimuli to other stimuli but without training or direct association. Arbitrarily derived relations and transfer of function help to explain how people can come to behave in certain ways towards stimuli that they have never encountered.

Applications of stimulus equivalence

The simplest derived relation is stimulus equivalence. Sidman and Tailby (1982) defined these based on mathematical equivalence relations. They suggested that stimulus equivalence was evidenced by relations that had the properties of symmetry, reflexivity and transitivity. These will be addressed in more detail later in this introduction. Simply, stimulus equivalence is where one stimulus evokes the behaviours associated with another stimulus, without training or explicit association. That is, the second stimulus becomes equivalent to the first. Although, as Barnes-Holmes, Barnes-Holmes, Smeets, Cullinan and Leader, (2004) point out, other derived relations have been studied (including greater-/less-than, same/different, coordination, and location and perspective-taking relations such as here/there or me/you), stimulus equivalence has been the focus of much of the research.

Research in stimulus equivalence was pioneered by Sidman (1971; 1994). The earliest studies (Sidman, 1971; Sidman & Cresson, 1973) successfully taught reading comprehension skills to children with intellectual disabilities. More recently, Leader and Barnes-Holmes (2001a) demonstrated the formation of equivalence relations between written fractions, pictorial representations of fractions, and decimal equivalents by 5-year old children, and the generalisation of these relations to similar representations of the fractions. Successful interventions have also been implemented with people with intellectual or developmental disabilities. An example is provided by LeBlanc, Miguel, Cummings, Goldsmith and Carr (2003). This study demonstrated the formation of equivalence relations by two children with autism (aged 6 and 13 years) between the shape, name, and capital city of three sets of three US states. Also, Elias, Goyos, Saunders and Saunders, 2008 used derived relations to teach manual signing to intellectually disabled adults.

Derived relations are the basis of Acceptance and Commitment Therapy (ACT). Since it was developed, ACT has been used to treat a wide range of disorders such as depression (e.g., Forman, Herbert, Moitra, Yeomans & Geller, 2007), anxiety (e.g., Eifert, Forsyth, Arch, Espejo, Keller & Langer, 2009), to help manage psychosis (e.g., Bach & Hayes, 2002) and chronic pain (e.g., Lunde & Nordhus, 2009), and to reduce problematic behaviours such as viewing internet pornography (e.g., Twohig & Crosby, 2010) and disordered eating (e.g., Juarascio, Forman & Herbert, 2010).

Thus derived relations have been and are being applied in many areas. To help understand their application and relevance more fully they have been the subject of many experimental investigations. One question has been how to facilitate, and therefore come to understand, their development.

Procedures in equivalence research

The precursor to any experimental investigation of derived relations is to identify an effective method by which to facilitate their development. The examples given by Blackledge (2004) and Dixon et al. (2006) suggest that much of the 'real-world' development of derived relations is often the result of associative rather than operant learning. It has been suggested that the differences between operant and associative procedures are not distinct (e.g., Rehfeldt & Hayes, 1998). However, the procedures used to study equivalence involve either operant or associative procedures. There are three methods that have been shown to facilitate the emergence of derived relations within an experimental setting. Two of these (matching-to-sample (MTS)) and the precursor to the relational evaluation procedure (pREP)) are based primarily in operant theory, while the other (stimulus-pairing observation) takes the form of associative learning. The research on each is outlined below.

Matching-to-sample. The first and most commonly used method in equivalence research is MTS (Barnes-Holmes, Barnes-Holmes, Smeets, Cullinan & Leader, 2004). MTS is a method used in a number of experimental analyses of behaviour, and was first employed as a method to study equivalence relations by Sidman (1971). In human experimental research, MTS usually takes the form of a computer task. As outlined in Barnes-Holmes et al. (2004), the participants are first presented with a sample stimulus. This is followed by a number of comparison stimuli. Three or four stimuli are commonly used. Two stimuli are considered insufficient as it is not possible to tell if the participant is responding towards one stimulus or away from the other (Sidman, 1994). The participant is then required to make a response, choosing one of the comparison stimuli. Feedback or reinforcement are provided for correct responses, and incorrect responses receive either negative feedback, punishment, or no feedback at all. For example, positive feedback and reinforcement may be provided when a participant chooses Stimulus B1 (comparison stimulus) in the presence of Stimulus A1 (sample stimulus), but not if they choose Stimulus B2.

MTS training can take three forms. The linear method involves training each stimulus to its following alphabetically designated stimulus (e.g., A-B, B-C, and C-D). In one-to-many training the same set of stimuli serve as the sample stimulus in all of the trials (e.g., A-B, A-C, A-D). In many-to-one training this scenario is reversed with the comparison stimuli remaining stable, and the sample stimulus changing (e.g., B-A, C-A, D-A). These outline that basic stimulus arrangement, however, there are many variations of these procedures.

Once these base relations have been established, the participants then undergo tests to demonstrate the formation of equivalence classes. During the testing conditions the participants do not receive reinforcement or feedback of any sort (Barnes-Holmes et al., 2004). Three different types of tests are conducted: reflexivity, symmetry, and transitivity. In reflexivity tests, to answer correctly participants are required to choose the comparison stimulus that is the same as the sample stimulus (e.g., if the participant is presented with sample stimulus A1, a correct response is choosing comparison stimulus A1, and not A2 or A3). To demonstrate symmetry, participants are required to demonstrate the reversal of the trained relationships (e.g., if the relationship A1-B1 has been trained, then the symmetry test involves choosing A1 when B1 is presented as the sample stimulus; Barnes-Holmes et al., 2004). Finally, successful transitivity tests involve the demonstration of equivalence relations between stimuli that have not appeared together in any of the trial tests. For example, if the trained relationships are A1-B1 and B1-C1, then transitivity is demonstrated if a participant chooses C1 when the sample stimulus is A1. Combined symmetry and transitivity is demonstrating the reverse of this relationship (i.e., choosing A1 when C1 is the sample stimulus). The successful completion of all of these tests demonstrates the formation of a stimulus equivalence class, in this case involving the stimuli A1, B1 & C1.

While the general format of MTS is similar across all studies, the procedures used can vary in a number of ways. Procedural differences include the number of stimuli in each class (e.g., Fields & Verhave, 1987), the type of stimuli used (e.g., Holth & Artnzen, 1998), the content and specificity of the instructions given (e.g., Drake & Wilson, 2008; Sigurdardottir, Green, & Saunders, 1990), the number of training trials used, how the stimuli are arranged in the trained (and subsequently tested) relations (e.g., Fields et al., Hobbie-Reeve, Adams, & Reeve, 1999; Saunders & Green, 1999), whether the sample and comparison stimuli are presented together (e.g., Markham, Dougher, & Auguston, 2002) or successively (e.g., Barnes-Holmes, Keane, Barnes-Holmes, & Smeets et al., 2000), and whether the tested relations include symmetry and/or transitivity (e.g., Barnes-Holmes et al., 2000; Hayes, Kohlenberg, & Hayes., 1991) or just equivalence (e.g., Arntzen, 2006).

The research to date on the use of the matching-to-sample procedure in stimulus equivalence is considerable. Therefore, a full review is not included here and relevant studies are outlined as required throughout this thesis. Since being established as a method for producing equivalence relations, MTS has been used as the basis of a wide range of equivalence studies covering topics such as generalisation (Fields, Reeve, Adams, et al., 1997), contextual control (Dibbets, Maes & Voessen, 2002), rule following and instructional control (Green, Sigurdardottir & Saunders, 1991; Hayes, Thompson & Hayes, 1989) and transfer of function (e.g., Markham et al., 2002; Roche & Barnes, 1997). As mentioned previously, the MTS procedure is widely used in the study of equivalence . When it is used some studies have reported that not all of their participants demonstrated equivalence (e.g., Holth & Arntzen, 1998; Rehfeldt, 2003). These findings are presented in more detail later in this thesis.

Stimulus-pairing observation procedure. The second methodology employed in the study of equivalence relations is based on associative learning rather than operant theory. A typical experiment involves a computer task where participants are presented with pairs of stimuli. The participants are not required to choose a comparison stimulus as in MTS, and as such, no feedback is given. This procedure was developed in an effort to assess the formation of equivalence classes using a non-operant theoretical base (Leader, Barnes & Smeets, 1996), and to ensure that equivalence was not a product of the MTS procedure only. Much less research has been done with the SPO than the MTS procedure; therefore, the following section outlines the SPO research in detail. With the exception of one study (Fields, Reeve, Varelas, et al., 1997), the SPO procedure is followed by MTS equivalence tests. The SPO procedure (e.g., Smyth, Barnes-Holmes, & Forsyth et al., 2006) has also been referred to as the respondent-type (ReT) procedure (Leader, Barnes, & Smeets, 1996; Leader, Barnes-Holmes, & Smeets, 2000; Leader & Barnes-Holmes, 2001a, 2001b; Clayton & Hayes, 2004), the stimulus pairing procedure (Fields, Reeve, Valeras et al.,1997), and the stimulus-stimulus pairing procedure (Layng & Chase, 2001).

Leader, Barnes and Smeets (1996) conducted the first study into the development of equivalence classes with an SPO procedure, which they labelled a "respondent-type training procedure" (pg. 685). In this study, six pairs of nonsense syllables (nine syllables in total) were presented to the adult participants. Initially, the syllables in each pair were presented successively, with a 0.5s within-pair delay between the syllables. There was a 3s delay between pairs. Once each pair had been presented 10 times in a quasi-random order, equivalence class formation was tested using a MTS test, with 10 tests for each equivalence relation. A consistency criterion required participants to choose the same, but not necessarily correct response on nine of the 10 trials for each relation. If this criterion was not met, the participant was re-exposed to the training and testing conditions.

They demonstrated an effective associative-based method by which to facilitate the formation of equivalence relations. However, the authors conceded that none of the participants formed the equivalence classes after exposure to the ‘respondent-type’ training alone, but only after a minimum of two exposures to both training and MTS testing (Leader et al., 1996). This study also demonstrated that the likelihood of equivalence was not affected by the specificity of the instructions, or the relative length of the within- and between- pair delay. Somewhat surprisingly, equivalence classes could still be formed when the within- and between-pair delays were the same (resulting in one long chain of stimuli presentations). However, when a fixed, but non-linear (A1-B1, A2-B2, A3-B3, B1-C1, B2-C2, B3-C3), presentation of the stimuli was used, none of the participants demonstrated equivalence with equal within- and between-pair delays. Thus, the authors suggest that the formation of equivalence relations with equal within- and between- pair delays was made possible by the linear (A1-B1, B1-C1, A2-B2, B2-C2, A3-B3, B3-C3) presentation of the stimuli.

The same group of authors completed four further studies (Smeets, Leader & Barnes, 1997; Leader, Barnes-Holmes & Smeets, 2000; Leader & Barnes-Holmes, 2001a; Leader & Barnes-Holmes, 2001b). The second study (Smeets et al., 1997) focused on different training and testing methods within the ‘respondent-type’ training procedure, and employed adult university students and preschool children as participants in two separate experiments. Both experiments assessed the effect of the arrangement in which the stimuli (nonsense syllables for the adult participants, and Greek letters and arbitrary symbols for the preschool children) were presented. Two arrangements were used. With the one-to-many (OTM) arrangement, the first stimulus presented in each stimulus pair always came from the same stimulus set (i.e., A-B, A-C). In the many-to-one (MTO) arrangement, the second stimulus in each pair always came from the same stimulus set (i.e., A-B, C-B). As with the earlier study (Leader et al., 1996) six pairs of stimuli were presented to participants using 0.5s within-pair and 3s between-pair delays. The instructions were the same as the minimal (less-specific) instructions used by Leader et al. (1996). In both conditions, the participants then completed a MTS test for symmetry and equivalence relations.

Overall, this study confirmed that the SPO procedure could facilitate the formation of equivalence classes with adults, and extended these findings to children when a simple-to-complex (i.e., tests for symmetry and equivalence are conducted separately in order of complexity with further training trials in between the tests) protocol was employed. However, the authors note that the pairing procedure may have been aided by the presentation of stimulus pairs on opposite sides of the same card (Smeets et al., 1997) and this was then addressed in a later study (Leader et al., 2000). It is also suggested that the MTO and OTM procedures are not equally effective at facilitating the formation of equivalence classes; however, the experiments in this study provide conflicting results. Therefore Smeets et al. (1997) identified this as an area for further study.

As mentioned above, the next study (Leader et al., 2000) was an extension of Smeets et al. (1997). In this study, 15 five-year old children completed a 'respondent-type' training procedure followed by MTS testing for the formation of equivalence classes with a simple-to-complex protocol and either a linear, OTM or MTO stimulus arrangement. In all experiments the stimuli were Greek letters and arbitrary symbols and were presented on separate cards. The participants first completed a MTS task using stimuli that were not used later in the experiment to ensure familiarity with the MTS procedure. If the participants met the criterion on the equivalence test they were exposed to further respondent training, pairing the most recently trained stimulus in each class to another two stimuli that had not previously been presented. MTS testing was then employed to test for the inclusion of this stimulus in the already formed equivalence class.

The results of this study confirmed the results of the previous studies that the SPO procedure was effective at facilitating the development of stimulus equivalence classes by children when a simple-to-complex protocol was used. Additionally, this study extended these findings to show that the class could be extended without testing for symmetry relations. There was little difference between the effectiveness of the linear, many-to-one, and one-to-many procedures, however; only following the one-to-many condition did the participants demonstrate all necessary equivalence relations with the fourth member of the class. The authors note that none of the participants could provide verbal reports to explain their performance. They argue that this suggests that being able to verbalise the relations is not a prerequisite of equivalence class formation (Leader et al., 2000).

Leader and Barnes-Holmes (2001a) endeavoured to establish equivalence classes involving fractions and their decimal equivalents using a ‘respondent-type’ procedure. Twenty-four 5-year old children were used as participants, eight in each of three experiments. The stimuli were written fractions (A stimuli), visual representations of the fractions (B stimuli), and the decimal proportion represented by the fractions (C stimuli). Prior to the beginning of the respondent training in Experiment 1, the participants completed MTS training and testing using arbitrary symbols as stimuli. Once this was completed the participants were presented with stimulus pairs using a ‘respondent-type’, simple-to-complex training and testing protocol, followed by tests for the combined symmetry and transitivity relations. Some of the participants also completed generalisation tests with modified visual representations of the fractions

The findings of this study demonstrated that young children can both form equivalence classes involving decimal and fraction stimuli, and generalise to other pictorial stimuli that represent the same fractions. Further this demonstrates the utility of the SPO procedure to facilitate the formation of equivalence classes within an applied setting (Leader & Barnes-Holmes, 2001a).

A five-experiment study by Layne and Chase (2001) looked at the effect of using MTS testing on the formation of equivalence classes using a SPO training procedure. In all of the experiments the stimuli were either symbols or Greek letters. During SPO training the participants used the ‘up’ arrow to move through the trials. Across three experiments the arrangement of the SPO training and MTS testing was altered to increase the number of SPO trials prior to the start of MTS testing. SPO training and MTS testing were then alternated until the participants met a criterion during testing. The same participants (university students) completed Experiments 1 and 2, with different participants in Experiment 3. The findings of these experiments suggested that, when participants do not have prior experience of equivalence procedures, the MTS test was important in the development of equivalence relations. Specifically, a large number of trials prior to the first test were less effective than cycles of training and testing involving fewer training trials. In Experiment 4 the participants were exposed to only one presentation of each stimulus pair in SPO training followed by a long testing condition. The authors concluded that testing alone was insufficient to facilitate the formation of equivalence relations. In the fifth experiment, participants learned equivalence relations with one set of stimuli, using the same procedure as Experiment 1. Once those relations were established, the relations between the stimuli in the training trials were rearranged. Most participants then demonstrated reversal of the equivalence relations by demonstrating equivalence with the new relations.

The authors concluded that using a SPO procedure is only effective when it is alternated with MTS testing and that the experimental history of the participants can affect performance on equivalence-class formation tasks. Specifically, a prior history of forming equivalence relations under experimental conditions aids in the acquisition of other classes using different stimuli. They also demonstrated that the reversal of equivalence relations was possible using a SPO procedure.

The findings of Layng and Chase (2001) are, however, contrary to the findings of an earlier study (Fields, Reeve, Varelas, et al., 1997) which demonstrated the formation of equivalence classes using a paired-stimuli procedure, and a yes/no testing procedure. As in much of the previous research, this study used undergraduate university students as participants. The stimuli were nonsense syllables and the participants received an instruction to work out which of the stimuli ‘go together’. The second stimulus in each pair was presented after the first stimulus had been removed from the screen. The subject was required to select either ‘yes’ or ‘no’ and in the training phase, the participants received ‘right’ or ‘wrong’ as feedback. This was faded across the training procedure and no feedback was presented during testing. The experiment employed a simple-to-complex training order. Once a three-member class had been established this was extended to four members, and then a delayed MTS test was used at the end to confirm the formation of the equivalence class.

Ten of the 18 participants met the 97% criterion on the four-member equivalence class using the yes/no procedure, and all of these participants passed the MTS test for equivalence. All of the participants who failed the equivalence test learnt the trained relations but failed to demonstrate the derived relations. The authors argue that the expansion of the equivalence class to four members is evidence of transfer of function, as the fourth stimulus was only paired with one member of the equivalence class and yet came to function as a member of that class. They conclude that the results demonstrate the development of equivalence classes without using the

MTS procedures. However, while MTS is not used, the participants do receive feedback for correct responses, and would therefore, generally be considered an operant procedure. The results are not the result of simple stimulus pairings, and as with the studies outlined previously, the participants received instructions to attend to the task, a further deviation from a simple pairing procedure.

Two studies have examined transfer of function through equivalence classes facilitated with the SPO procedure. Tonneau and Gonzalez (2004) assessed the SPO procedure, and function transfer. Their study employed students between the ages of 11 and 15. The stimuli were geometric shapes and black line figures. In the first experiment, six participants were initially trained to press a different key in the presence of each of three black line figures to a criterion of 100%. The participants then completed a SPO procedure followed by testing for the transfer of the correct response to the geometric figures. Nearly all of the participants demonstrated the transfer of function from the black line figures to the geometric shapes. The authors suggest that the stimulus pairings produced the transfer of function from one stimulus to another. The second experiment involved a second group of participants who received MTS training. The pairings for the participants in the SPO group were yoked to the matching trials completed by the participants in the MTS group. For each pair of participants the SPO procedure provided the same stimuli presentations as for the MTS condition, including corrections. All of the participants showed transfer of function; however those who received SPO training demonstrated transfer of function more quickly than those who received MTS training. The authors also found that attempting to minimise the covert verbal behaviour of the participants by requiring them to vocalise throughout the testing procedures did not affect the formation of equivalence classes. Overall, these findings provide further support for the SPO procedure as an effective method by which to facilitate the formation of equivalence relations. It also extends this to demonstrate transfer of function without an operant training procedure. The authors concluded that stimulus pairings are sufficient for the formation of equivalence classes to occur. However, they do concede that the use of operant measures such as instructions to attend to the screen and to outline the task may aid in the formation of equivalence classes using paired stimuli procedures. Tonneau and Gonzalez (2004) suggested that the SPO procedure resulted in transfer of function more quickly than the MTS procedure.

Smyth et al., (2006) assessed function transfer of a simple conditional discrimination, and the self-reported ratings of fear arousal in undergraduate students who reported that they either were or were not fearful of spiders. In their first Experiment, Smyth et al. (2006) first established the functions of 'correct' or 'wrong' to two stimuli. These stimuli were each used as one of a set of stimuli in the trained relations for an equivalence class. The other stimuli were nonsense syllables. This training used an SPO procedure where each pair of stimuli was presented twice. The pairings had a 0.5 s within-pair delay, and a 3 s between-pair delay. This was followed immediately with a test for the transfer of function. If participants demonstrated transfer of function, they then completed a MTS equivalence test. If they failed to demonstrate transfer of function then they were re-exposed to the SPO procedure. Two of the three participants demonstrated equivalence. Following this, the relations between stimuli in the baseline associations were reversed and the transfer of function, and equivalence tests were repeated. All participants demonstrated transfer of function and two demonstrated the reversal of the equivalence relations.

Experiment 2 looked at transfer of function with undergraduate students who reported being fearful, or not fearful of spiders. The participants in Experiment 2 first completed the same procedure as Experiment 1, but without the reversal task. Eleven of the 16 participants demonstrated transfer of function. All proceeded to the second part of the experiment. Part 2 of Experiment 2 was similar to Part 1, however, two of the stimuli were initially paired with either a video of a spider attack, or a blank screen. Measures of fear and disgust, and ratings of intensity with regard to the nonsense syllable stimuli were also taken initially, and after the SPO training. To demonstrate transfer of function, the participants were presented with the nonsense syllables and asked to choose whether they were related to a spider or blank video. This was followed by an equivalence test. The authors report that the SPO procedure was effective in facilitating the transfer of function, with participants who were fearful of spiders reporting greater levels of fear and disgust and higher intensity ratings for the nonsense syllable stimuli that were in the equivalence class with the nonsense syllable that had been paired with the spider attack video.

Therefore, as with the MTS procedure, research using SPO procedures has demonstrated transfer of function, and Tonneau and Gonzalez (2004) suggest that the

SPO procedure might be more effective at facilitating transfer of function than the traditionally used MTS procedure.

In summary, the SPO procedure has been demonstrated to be effective; however, as with the MTS procedure some failures to demonstrated equivalence have been reported.

Precursor to the relational-evaluation procedure (pREP). The least developed of the three methodologies is the precursor to the relational evaluation procedure (pREP). As with the SPO procedure, there is only a small field of research compared to that done with MTS, and so this is outlined in detail below. pREP was first developed by Cullinan, Barnes and Smeets (1998) and took the form of a go/no-go procedure. On each trial the participants were presented with a sample stimulus, followed by a comparison stimulus. If the comparison stimulus was the correct comparison, then the participants were required to respond by pressing the space bar. If it was a negative comparison, they were required to give no response. Correct and incorrect responses and non-responses resulted in positive or negative feedback. This study (Cullinan et al., 1998) employed both pREP and MTS training and testing procedures. In all of the experiments, the stimuli were two sets of three nonsense syllables. During MTS training the feedback for correct and incorrect responses and increments/decrements of the points tally were the same as during pREP training. Failure to meet the test criterion resulted in retraining followed by a repeat of the test condition. Alternation of training and testing continued until the participant met the testing criterion or exhibited a stable, but incorrect, response pattern over two testing cycles. pREP testing involved test trials for eight symmetry relations and four equivalence relations. MTS testing involved test trials for four symmetry relations and two equivalence relations as two comparison stimuli were presented for each sample stimulus, while one comparison stimulus was presented on each pREP trial.

In conclusion, the findings of the study by Cullinan et al. (1998) suggest that pREP does not facilitate the emergence of equivalence relations when used as a stand-alone procedure. Rather, some MTS procedures must be employed for equivalence relations to emerge. The authors also suggest the pREP procedure may have been affected by other problems associated with go/no-go procedures, the lack of specificity in the instructions, simultaneous versus successive presentation of the sample and comparison stimuli, and the possible failure of the pREP procedure to result in the formation of two separate equivalence classes.

A second study (Cullinan, Barnes-Holmes & Smeets, 2000) was undertaken to examine the possible causes of the superiority of MTS over pREP. The basic training, testing and feedback procedures were the same as in Cullinan et al. (1998). Experiments 1-3 in this study each involved a modification of one of these original procedures including providing the participants with control of the interval length in which they had to respond, providing feedback only on responses (not no-go's), and providing feedback on all trials. Experiments 4 and 5 both examined the effect of a history of MTS training and testing on the formation of equivalence relations.

Overall the results of the study by Cullinan et al. (2000) did not identify any factors that increased the effectiveness of the pREP as a stand-alone procedure for facilitating the formation of equivalence relations but the findings of Experiments 4 and 5 did suggest that a history of MTS training and testing can facilitate the emergence of derived equivalence relations using pREP procedures.

The same set of authors completed a further study (Cullinan, Barnes-Holmes & Smeets, 2001), this time endeavouring to find if different contextual cues would aid in the formation of equivalence relations using a pREP procedure. In Experiment 1, a two response option procedure replaced the go/no-go procedure used in the previous pREP studies. Sixteen participants, four in each of four conditions, were presented with pairs of successive stimuli. This was followed by the presentation of two response options (these were the contextual cues and differed across conditions) of which the participants were required to choose one. The contextual cues used in each condition were as follows: Condition 1: !!!!, ****; Condition 2: *yes, no*; Condition 3: *goes with, does not go with*; Condition 4: *same, different*. The only contextual cues that resulted in the reliable formation of equivalence classes using a pREP procedure were *same* and *different*. Further experiments demonstrated that the facilitation of equivalence relations with a pREP procedure could also be aided by the addition of a pre-training procedure.

Therefore, this study was the first demonstration of the formation of equivalence relations using pREP without MTS. However, the procedure has been modified from the original pREP procedure used in earlier studies (Cullinan et al., 1998; 2000), and closely resembles the yes-no testing procedure used in the study by

Fields et al., (1997). The yes-no procedure was not used in training during that study, but rather, it followed a paired-stimuli procedure (Fields et al., 1997).

Smeets, Wijngaarden, Barnes-Holmes & Cullinan (2004) replicated and modified the study by Cullinan et al. (1998). The rates of equivalence demonstrated in this experiment were the same as in the Cullinan et al. (1998) study with regard to the pREP; however, they are lower in this study than the earlier study for MTS. In contrast to the earlier study, pREP did not appear to facilitate the formation of symmetry relations. As with Cullinan et al. (1998), Smeets et al. (2004) demonstrated that a history of MTS testing facilitates the formation of equivalence relations using the pREP. The extensions in this study demonstrated the use of a pREP training procedure helped to increase the likelihood of symmetry, but not equivalence relations, and that the use of a simple-to-complex procedure aided in the formation of equivalence relations during the pREP test.

Smeets, Barnes-Holmes and Striefel (2006) conducted six experiments. Their first experiment used a procedure modified from Cullinan et al.'s (1998) Conditions 1 and 2. The differences between this study and Cullinan et al. (1998) were that the relations were each trained separately (i.e., the A-B relations were trained to criterion prior to training the B-C relations) and the stability criterion was removed. With a procedure using pREP procedures only, the participants demonstrated symmetry but not equivalence. Equivalence was demonstrated following MTS tests. Experiment 2 was a replication of Experiment 1, however, the instructions included a directive to respond on half of the trials (Smeets et al., 2006). The results of Experiment 2 were similar to Experiment 1. Experiment 3 employed a simple-to-complex procedure, testing for the symmetry relations first, followed by the equivalence relations. The findings of this study showed the simple-to-complex procedure to be slightly more effective than the procedure of Experiment 2. Experiment 4 compared MTO and linear stimulus arrangements, and also required the participants to read the instructions out aloud. Nearly all of the participants demonstrated equivalence with these modifications, and similar results were achieved with each of the stimulus arrangements. In their fifth experiment, the feedback for no-go responses was removed. Therefore, participants only received feedback when they pressed the key. All of the participants demonstrated equivalence, which is different to the findings reported by Cullinan et al. (2000) who reported that nearly all of their participants failed to demonstrate equivalence under a similar procedure. In Experiment 6, once participants demonstrated equivalence (using the same procedure as Experiment 5) the baseline relations were reversed (e.g., where previously B1-C1 was correct, B1-C2 was now correct). Nearly all participants demonstrated equivalence both initially, and following the reversal of the baseline relations. Smeets et al. (2006) conclude that the pREP procedure is effective at both establishing and reversing equivalence relations when a simple-to-complex procedure is used, feedback is only delivered for 'go' responses, and when the instructions ask participants to respond on half of the trials.

Thus far, only one study (Smeets et al., 2006) has demonstrated the successful establishment of equivalence relations using a pREP procedure. Thus, it appears that the facilitation of equivalence using a pREP procedure is possible, as is reversal and subsequent demonstration of new equivalence relations using this procedure. One other study (Cullinan et al., 1988) found the pREP procedure to be less effective than MTS.

Comparison of procedures. As mentioned previously, most equivalence research has used a MTS procedure. However, the formation of equivalence relations in applied settings (as outlined previously) may also result from associative pairings (e.g., Dixon et al., 2006). So far there is little evidence for the effectiveness of the pREP procedure. The MTS and SPO procedures have both been shown to facilitate the formation of equivalence relations, but the question remains as to which of these procedures is most effective.

Two studies (Leader & Barnes-Holmes, 2001b; Clayton & Hayes, 2004) have compared the effectiveness of the MTS and paired-stimuli procedures at facilitating the formation of equivalence relations. In the study by Leader and Barnes-Holmes (2001b), undergraduate university students completed both MTS and SPO training in a series of within-subjects experiments involving the same nonsense syllables in each condition. All of the experiments involved a linear stimulus arrangement during training, a MTS testing procedure, and a criterion of 9/10 correct responses for each tested relation. This study reported that the SPO procedure was more effective than the MTS procedure even when a criterion was introduced during MTS training. The procedures were found to be equally effective when the number of comparison stimuli in MTS training was reduced to the one correct option. It could be argued that this procedural change reduced the MTS procedure to an SPO procedure with a required response. The authors note that many of the incorrect responses made during the MTS procedure were the correct responses for the SPO procedure. This suggests that the use of the same stimuli in both procedures may have interfered with the formation of equivalence relations when the MTS procedure was used.

The more recent study (Clayton & Hayes, 2004) found contradictory results and concluded that MTS was more effective than the SPO procedure. As in the study by Leader and Barnes-Holmes (2001b), Clayton and Hayes (2004) employed university students as participants, and these participants completed both MTS and SPO training; however, they used Chinese characters or arbitrary symbols as stimuli. Prior to the beginning of the first training session, and at the completion of the experiment the participants completed a scaling exercise where they were exposed to each of the stimuli and asked to choose the 6 most related and 6 least related objects from the remaining stimuli. This was to account for the effect of formal similarities on equivalence-class formation. One of the stimulus sets was used in both training conditions; the remainder of the stimuli used in each condition were different. The demonstration of equivalence was followed by an extended MTS test for equivalence combining the two sets of stimuli.

Overall, Clayton and Hayes (2004) reported that their participants performed better on the symmetry and equivalence tests involving the stimuli used in the MTS training than the tests involving the stimuli used in the SPO training procedure. There was also some evidence for extended equivalence. The participants performed better with both procedures when the arbitrary symbols were used as stimuli suggesting that it may be easier to form equivalence relations with arbitrary stimuli than with Chinese characters, however, the SPO procedure still resulted in greater rates of equivalence than the MTS procedure.

Clayton and Hayes (2004) suggest that the differences between their results and those of Leader and Barnes-Holmes (2001b) were due to the stimuli used in each study. The stimuli in the study by Leader and Barnes-Holmes (2001b) were nonsense syllables, and as such were easily named by the participants. In contrast, Clayton and Hayes (2004) used Chinese characters (and in the final experiment, arbitrary symbols), which their subjects could not read and, due to their complexity, could not name easily.

Conclusion

In conclusion, three methods for studying the formation of derived relations have been developed. Of these, two (MTS and SPO) have been shown to result in the formation of derived relations across a variety of settings, participants, and stimuli; however, due to conflicting results, the comparable effectiveness of these two methods has not been determined. The two studies (Clayton & Hayes, 2004; Leader & Barnes-Holmes, 2001b) that have compared the effectiveness MTS and SPO procedures found conflicting results. One factor identified by Clayton and Hayes (2004) that may account for the different results was the stimuli used. Thus the first study of this research replicated the procedure of Leader and Barnes-Holmes (2001b) but used the same stimuli as Clayton and Hayes (2004).

EXPERIMENT 1

Method

Participants

Participants were recruited through an information sheet on the notice boards in the University of Waikato Psychology Department, and through a posting on the University of Waikato e-learning forum site for two first-year psychology papers. The information sheet provided information about the research and the contact details of the researcher. There were 12 participants in this experiment (P1.1-P1.12), three in each of four groups.

Ethics

Ethical approval was granted by the Psychology Department Ethics Committee. At the beginning of the study, participants were provided with an information sheet. They were provided with the opportunity to ask questions both before and after their participation. All of the participants gave informed consent. Those participants who were enrolled in either of the first-year psychology papers received 1% course credit for each session they attended, up to a maximum of 4% (four sessions).

Apparatus and Setting

Participants were seated at a computer in one of the university computer rooms. The room was quiet and free from distraction, and the participants were alone in the room during the experiment. The experimental instructions and experimental task were presented in black, on a white background, on a 19” monitor. Three keys on the keyboard (Z, V & M) were marked with white paper dots. Each experimental session was approximately one hour in length.

Stimuli

The stimuli used were nine of the 18 Chinese characters used by Clayton and Hayes (2004). They were divided into three groups of three stimuli and each stimulus was given an alphanumeric designation (i.e., A1, A2, A3, B1, B2, B3, C1, C2, and C3). The participants were never shown the alphanumeric designations of the stimuli. The stimuli and their alphanumeric designation are shown in Table 1.1.

Table 1.1

Chinese characters and their alphanumeric designation

错	A1	枝	A2	死	A3
博	B1	歌	B2	新	B3
舞	C1	姜	C2	帝	C3

Pre-experimental procedure

Prior to the beginning of the first session, the participants were presented with all of the stimuli and asked "Are you able to name these characters?". If a participant answered 'No' to this question, they were quasi-randomly assigned to either the E1.N.SPO.MTS or E1.N.MTS.SPO group. If a participant answered 'Yes' to this question they were quasi-randomly assigned to either the E1.Y.SPO.MTS or E1.Y.MTS.SPO group. Group assignment and condition order are shown in Table 1.2.

Table 1.2

Group assignment and condition order for all participants.

Participant	Group	Reported that they could (yes) or could not (no) read the characters	Condition order
P1.1	E1N.SPO.MTS	No	SPO/MTS
P1.2	E1N.SPO.MTS	No	SPO/MTS
P1.3	E1N.SPO.MTS	No	SPO/MTS
P1.4	E1N.MTS.SPO	No	MTS/SPO
P1.5	E1N.MTS.SPO	No	MTS/SPO
P1.6	E1N.MTS.SPO	No	MTS/SPO
P1.7	E1Y.SPO.MTS	Yes	SPO/MTS
P1.8	E1Y.SPO.MTS	Yes	SPO/MTS
P1.9	E1Y.SPO.MTS	Yes	SPO/MTS
P1.10	E1Y.MTS.SPO	Yes	MTS/SPO
P1.11	E1Y.MTS.SPO	Yes	MTS/SPO
P1.12	E1Y.MTS.SPO	Yes	MTS/SPO

Procedure

Procedurally, the experiment was very similar to Leader and Barnes-Holmes's (2001b) Experiment 1. Each session included two training conditions: SPO training, and MTS training. Each of the training conditions was followed by MTS testing. The current experiment used the same pairs of stimuli for all participants. The participants in Groups E1.Y.SPO.MTS and E1.N.SPO.MTS completed SPO training followed by MTS training in each session. The participants in Groups E1.N.MTS.SPO completed MTS training prior to SPO training.

SPO training. At the beginning of SPO training the participants were presented with the following instructions. The instructions remained on the screen for 5 s.

In this stage of the experiment your task is to simply watch the screen

During SPO training, the participants were presented with pairs of Chinese characters. The pairs (trained relations) were as follows: A1-B1, A2-B2, A3-B3, B1-C1, B2-C2, and B3-C3. The characters were centred (vertically and horizontally) on the monitor, with only one character presented at a time. Each stimulus remained on the screen for 1 s. There was a within-pair delay of 0.5 s, and a between-pair delay of 3 s. Each pair was presented 10 times quasi-randomly (60 trials in 10 blocks, with each pair presented once in each block). This was followed by a 5 s end-of-stage delay, after which the instructions for the MTS test appeared on the screen. During the within-pair, between-pair, and end of stage delays the screen was blank (white). This procedure is outlined below:

In this stage of this experiment your task is to simply watch the screen (5 s)

⇓

Stimulus 1 (1 s)

⇓

0.5 s within-pair delay

⇓

Stimulus 2 (1 s)

⇓

3 s between-pair delay

5 s end of stage delay (following the presentation of all 60 trials)

MTS test following SPO training. The MTS test that followed the SPO training tested for the emergence of derived relations between the stimuli. The relations tested were the six symmetry (B1-A1, B2-A2, B3-A3, C1-B1, C2-B2, and C3-B3), and three equivalence (C1-A1, C2-A2, and C3-A3) relations that could be formed from the trained relations. The first character in each tested relation above was presented as the sample stimulus. The second character in each tested relation was the correct answer and the selection of that stimulus demonstrated the formation of a derived symmetry or equivalence relation. This stimulus, along with the two others in its group were presented as comparison stimuli (for example, if B1 was the sample stimulus, A1, A2, and A3 were the comparison stimuli, and choosing A1 was the correct response). The location of the comparison stimuli were varied on a quasi-random basis.

Immediately following the end-of-stage delay for SPO training, the following instructions appeared on the screen:

In this stage of the experiment you must look at the character at the top, and then choose one of the three characters at the bottom, by pressing one of the marked keys on the keyboard. To choose the left character, press the marked key on the left. To choose the middle character, press the marked key in the middle. To choose the right character, press the marked key on the right. Press the space-bar twice to continue.

Once the participants had pressed the space bar twice, the first sample stimulus appeared on the screen. The sample stimulus was centred in the top half of the screen and remained on the screen by itself for 1.5 s, after which three comparison stimuli appeared in the bottom half of the screen, with equal distances between each. The participants were then required to make their choice, and press the key corresponding to their chosen character (Z, V, or M, which were marked by white paper dots). The sample and comparison stimuli remained on the screen until a response was made. Each response was followed by a 3 s between-trial delay. No feedback was given following any of the trials. Each of the nine tested relations was presented 10 times in a quasi-random order (a total of 90 trials, in blocks of 10, with each trial presented once within each block). The procedure for MTS testing is outlined below:

In this stage of the experiment you must look at the character at the top, and then choose one of the three characters at the bottom, by pressing one of the marked keys on the keyboard. To choose the left character, press the marked key on the left. To choose the middle character, press the marked key in the middle. To choose the right character, press the marked key to the right. Press the space-bar twice to continue.

⇓

Sample stimulus

remains on screen for 1.5 s

⇓

Sample stimulus

Comp 1 Comp 2 Comp 3

remains on screen until response made

⇓

3 s between-trial delay

MTS training. At the beginning of MTS training the participants were presented with the following instructions:

> *In this stage of the experiment you must look at the character at the top, and then choose one of the three characters at the bottom, by pressing one of the marked keys on the keyboard. To choose the left character, press the marked key on the left. To choose the middle character, press the marked key in the middle. To choose the right character, press the marked key to the right. Press the space-bar twice to continue.*

The procedure used in MTS training was similar to that used in MTS testing, differing in two respects. First, feedback was provided following each response. Correct responses were followed by a tone (1000 Hz) and the presentation on the screen of the word 'CORRECT'. Incorrect responses were followed by the presentation on the screen of the word 'INCORRECT'. No tone accompanied the feedback for incorrect responses. Feedback remained on the screen for 1.5 s. Second, the trained relations involved different arrangements of the stimuli. The trained relations in MTS training were A1-B2, A2-B3, A3-B1, B1-C2, B2-C3, and B3-C1. As with MTS testing, the first stimulus in each pair was the sample stimulus, and the second was the correct comparison stimulus (it was presented along with the other stimuli in the same group). For example, if the sample stimulus was A1, then B1, B2, and B3 were presented as comparison stimuli. If the participant chose B2 then their response was followed by the feedback for a correct response. Choosing either of the other comparison stimuli resulted in feedback for an incorrect response. The procedure for MTS training is outlined below:

In this stage of the experiment you must look at the character at the top, and then choose one of the three characters at the bottom, by pressing one of the marked keys on the keyboard. To choose the left character, press the marked key on the left. To choose the middle character, press the marked key in the middle. To choose the right character, press the marked key to the right. Press the space-bar twice to continue.

⇓

Sample
stimulus

remained on screen for 1.5 s

⇓

Sample
stimulus

Comp 1 Comp 2 Comp 3

remained on screen until response made

⇓

"Correct" and high pitched tone, or "Wrong" appear on screen for 1.5 s

3 s between-trial delay

MTS testing following MTS training. This MTS testing condition was procedurally identical to MTS testing following SPO training; however the tested relations relate to the new arrangements that were trained in the MTS testing. Therefore, the tested relations were six symmetry relations (B1-A3, B2-A1, B3-A2, C1-B3, C2-B1, and C3-B2) and three equivalence relations (C1-A2, C2-A3, and C3-A1).

Condition Criterion and Session Criterion

Once each participant had completed both training and testing conditions, the number of correct responses on each tested relation and the overall percentage of correct responses during each testing condition were calculated. A participant met the condition criterion if they responded correctly on nine of the 10 trials for every tested relation during testing following either training procedure of a session. The session criterion was met if the participant met the condition criterion during testing following both training procedures in one session. If a participant did not meet the session criterion then they were re-exposed to the entire experimental procedure. This continued until they met the session criterion or until they had completed the maximum of four experimental sessions.

At the end of each experimental session, the participant was presented with the following instructions:

Thank you for your participation. Please contact the researcher.

At the end of each session, the participants were advised as to whether they should return for a further session. At the end of their final session, participants were thanked for their participation and given the opportunity to ask questions.

Results

Number of sessions

None of the participants met the session criterion in any of the experimental sessions. Therefore, all participants in this experiment completed the maximum of four experimental sessions.

Testing

Tables 1.3 – 1.6 show the number of correct responses (with a maximum of 10) for each tested relation; the percent of responses that were correct on the symmetry and equivalence trials, and over all trials, during each condition of each experimental session, and whether the criterion for each condition, and the overall session criterion were met, for each of the participants in Groups E1.N.SPO.MTS (P1.1-P1.3), E1.N.MTS.SPO (P1.4-1.6), E1.Y.SPO.MTS (P1.7-P1.9), and E1.Y.MTS.SPO (P1.10-P1.12) respectively. In these tables, numbers that appear in bold indicate that nine or more of the ten trials for that tested relation resulted in a correct response.

Number correct on each tested relation and overall percent correct. Overall, as shown in Tables 1.3 and 1.4, half of the participants who could not read the characters (Groups E1.N.SPO.MTS and E1.N.MTS.SPO) achieved greater percentages of correct responses following SPO training than following MTS training during all sessions but did not meet the condition criterion was in any session. The remaining participants in these groups achieved a greater percentage of correct responses following SPO training in early sessions, but this switched to a greater percentage correct following MTS training in the later sessions. Two of these participants, one in each group (P1.3 and P1.5) either met the condition criterion, or achieved a result that was very close to the condition criterion, following MTS training in the final session.

Table 1.3

Number of correct responses for each trial type during testing of each training and testing cycle in each experimental session for the participants (in Group E1.N.SPO.MTS) who reported that they could not name the stimuli and completed SPO followed by MTS training.

Participant	Session	Condition	Tested relations - sample and correct comparison stimuli											Overall % Correct	Condition criterion met	Session criterion met
			Symmetry relations						% Correct Symmetry	Equivalence relations			% Correct Equivalence			
			B1-A1 (SPO) B1-A3 (MTS)	B2-A2 (SPO) B2-A1 (MTS)	B3-A3 (SPO) B3-A2 (MTS)	C1-B1 (SPO) C1-B3 (MTS)	C2-B2 (SPO) C2-B1 (MTS)	C3-B3 (SPO) C3-B2 (MTS)		C1-A1 (SPO) C1-A2 (MTS)	C2-A2 (SPO) C2-A3 (MTS)	C2-A2 (SPO) C3-A1 (MTS)				
P1.1	Session 1	SPO	8	2	0	**10**	0	8	46.67	**10**	3	2	50.00	47.78	N	N
		MTS	0	**10**	7	0	0	0	28.33	0	2	2	13.33	23.33	N	
	Session 2	SPO	**9**	1	8	**9**	0	**10**	61.67	**10**	0	**9**	63.33	62.22	N	N
		MTS	0	**9**	1	0	0	0	16.67	0	**10**	0	33.33	22.22	N	
	Session 3	SPO	8	1	**10**	**10**	0	**10**	65.00	**10**	2	**9**	70.00	66.67	N	N
		MTS	4	5	8	1	0	0	30.00	0	3	0	10.00	23.33	N	
	Session 4	SPO	4	7	**10**	2	**9**	8	66.67	1	0	4	16.67	50.00	N	N
		MTS	0	2	3	**9**	6	0	33.33	2	**9**	1	40.00	35.56	N	
P1.2	Session 1	SPO	**9**	**10**	**9**	5	4	1	63.33	**10**	0	8	60.00	62.22	N	N
		MTS	3	**9**	2	1	**9**	2	43.33	4	**0**	6	33.33	40.00	N	
	Session 2	SPO	**10**	**9**	**10**	1	4	8	70.00	2	0	**9**	36.67	58.89	N	N
		MTS	2	2	2	6	1	3	26.67	2	5	4	36.67	30.00	N	
	Session 3	SPO	8	8	7	8	8	8	78.33	1	4	2	23.33	60.00	N	N
		MTS	4	4	2	4	7	6	45.00	2	4	6	40.00	43.33	N	
	Session 4	SPO	**9**	**10**	**9**	8	**9**	8	88.33	7	0	1	26.67	67.78	N	N
		MTS	2	2	7	4	3	2	33.33	3	3	6	40.00	35.56	N	
P1.3	Session 1	SPO	7	0	2	**10**	0	**10**	48.33	**9**	3	1	43.33	46.67	N	N
		MTS	**10**	2	**10**	7	0	1	50.00	**9**	0	0	30.00	43.33	N	
	Session 2	SPO	**10**	7	2	7	**10**	**10**	76.67	**10**	0	0	33.33	62.22	N	N
		MTS	5	7	2	5	3	1	38.33	2	2	6	33.33	36.67	N	
	Session 3	SPO	0	**10**	0	**10**	**10**	**10**	66.67	0	**10**	0	33.33	55.56	N	N
		MTS	0	0	**10**	1	1	0	20.00	**10**	0	0	33.33	24.44	N	
	Session 4	SPO	4	0	7	3	5	**9**	46.67	3	1	8	40.00	44.44	N	N
		MTS	**10**	8	**10**	**10**	**10**	**10**	96.67	**10**	**10**	**10**	**100.00**	97.78	N	

Table 1.4

Number of correct responses for each trial type during testing of each training and testing cycle in each experimental session for the participants (in Group E1.N.MTS.SPO) who reported that they could not name the stimuli and completed MTS followed by SPO training.

Participant	Session	Condition	Tested relations - sample and correct comparison stimuli											Overall % Correct	Condition criterion met	Session criterion met
			Symmetry relations						% Correct Symmetry	Equivalence relations			% Correct Equivalence			
			B1-A1 (SPO) B1-A3 (MTS)	B2-A2 (SPO) B2-A1 (MTS)	B3-A3 (SPO) B3-A2 (MTS)	C1-B1 (SPO) C1-B3 (MTS)	C2-B2 (SPO) C2-B1 (MTS)	C3-B3 (SPO) C3-B2 (MTS)		C1-A1 (SPO) C1-A2 (MTS)	C2-A2 (SPO) C2-A3 (MTS)	C2-A2 (SPO) C3-A1 (MTS)				
P1.4	Session 1	SPO	7	2	2	2	1	2	26.67	**10**	2	5	56.67	36.67	N	
		MTS	1	4	1	6	3	0	25.00	3	2	1	20.00	23.33	N	N
	Session 2	SPO	8	7	6	3	1	1	43.33	**10**	0	2	40.00	42.22	N	
		MTS	1	3	1	7	0	0	20.00	3	0	5	26.67	22.22	N	N
	Session 3	SPO	7	5	6	6	6	3	55.00	5	7	1	43.33	51.11	N	
		MTS	4	4	5	4	2	4	38.33	0	2	2	13.33	30.00	N	N
	Session 4	SPO	7	8	8	5	1	6	58.33	2	4	4	33.33	50.00	N	
		MTS	4	6	0	7	3	6	43.33	3	0	5	26.67	37.78	N	N
P1.5	Session 1	SPO	8	3	5	8	4	8	60.00	**9**	4	3	53.33	57.78	N	
		MTS	1	3	3	1	2	2	20.00	3	3	4	33.33	24.44	N	N
	Session 2	SPO	6	6	0	4	1	8	41.67	4	0	7	36.67	40.00	N	
		MTS	1	5	1	2	4	0	21.67	1	4	7	40.00	27.78	N	N
	Session 3	SPO	0	0	0	0	**10**	0	16.67	**10**	0	0	33.33	22.22	N	
		MTS	**9**	**9**	**10**	0	0	9	61.67	7	0	10	56.67	60.00	N	N
	Session 4	SPO	0	0	0	0	0	0	0.00	0	0	0	0.00	0.00	N	
		MTS	**10**	**10**	**10**	**10**	**10**	**9**	98.33	**10**	**10**	**10**	**100.00**	**98.89**	**Y**	N
P1.6	Session 1	SPO	8	**9**	**10**	**10**	**10**	0	78.33	**9**	8	0	56.67	71.11	N	
		MTS	0	4	1	0	2	1	13.33	0	0	2	6.67	11.11	N	N
	Session 2	SPO	9	7	**9**	**10**	8	4	78.33	**10**	**9**	1	66.67	74.44	N	
		MTS	0	6	0	0	5	1	20.00	0	0	7	23.33	21.11	N	N
	Session 3	SPO	8	7	5	7	3	2	53.33	**9**	3	3	50.00	52.22	N	
		MTS	3	5	8	2	4	5	45.00	0	4	5	30.00	40.00	N	N
	Session 4	SPO	0	1	0	0	8	0	15.00	8	2	0	33.33	21.11	N	
		MTS	**9**	7	6	**10**	1	8	68.33	0	0	**9**	30.00	55.56	N	N

Table 1.5

Number of correct responses for each trial type during testing of each training and testing cycle in each experimental session for the participants (in Group E1.Y.SPO.MTS) who reported that they could name the stimuli and completed SPO followed by MTS training.

Participant	Session	Condition	Tested relations - sample and correct comparison stimuli											Overall % Correct	Condition criterion met	Session criterion met
			Symmetry relations						% Correct Symmetry	Equivalence relations			% Correct Equivalence			
			B1-A1 (SPO) B1-A3 (MTS)	B2-A2 (SPO) B2-A1 (MTS)	B3-A3 (SPO) B3-A2 (MTS)	C1-B1 (SPO) C1-B3 (MTS)	C2-B2 (SPO) C2-B1 (MTS)	C3-B3 (SPO) C3-B2 (MTS)		C1-A1 (SPO) C1-A2 (MTS)	C2-A2 (SPO) C2-A3 (MTS)	C2-A2 (SPO) C3-A1 (MTS)				
P1.7	Session 1	SPO	2	**9**	**10**	2	4	2	48.33	5	4	1	33.33	43.33	N	
		MTS	1	1	1	1	6	4	23.33	4	4	3	36.67	27.78	N	N
	Session 2	SPO	5	7	3	8	3	1	45.00	**9**	5	1	50.00	46.67	N	
		MTS	7	1	5	0	2	**9**	40.00	4	2	2	26.67	35.56	N	N
	Session 3	SPO	3	**9**	0	0	0	2	23.33	2	0	**10**	40.00	28.89	N	
		MTS	0	3	**10**	0	5	0	30.00	5	0	0	16.67	25.56	N	N
	Session 4	SPO	0	2	5	0	5	1	21.67	5	5	7	56.67	33.33	N	
		MTS	8	2	8	1	5	3	45.00	1	5	0	20.00	36.67	N	N
P1.8	Session 1	SPO	1	**10**	0	0	3	1	25.00	0	**9**	1	33.33	27.78	N	
		MTS	0	0	**10**	1	**10**	**9**	50.00	8	0	1	30.00	43.33	N	N
	Session 2	SPO	**10**	**10**	7	**10**	**10**	3	83.33	8	8	1	56.67	74.44	N	
		MTS	0	1	10	1	0	**9**	35.00	1	1	0	6.67	25.56	N	N
	Session 3	SPO	**10**	**10**	**10**	**10**	**10**	**10**	**100.00**	**10**	7	1	60.00	86.67	N	
		MTS	1	2	3	0	0	1	11.67	4	0	2	20.00	14.44	N	N
	Session 4	SPO	**10**	**9**	4	**10**	**10**	**10**	88.33	**9**	**9**	0	60.00	78.89	N	
		MTS	7	6	**10**	7	2	4	60.00	4	0	1	16.67	45.56	N	N
P1.9	Session 1	SPO	**10**	7	**9**	2	**10**	**10**	80.00	0	4	**10**	46.67	68.89	N	
		MTS	0	0	0	0	0	5	8.33	7	4	0	36.67	17.78	N	N
	Session 2	SPO	**10**	9	**10**	1	**10**	**10**	83.33	0	9	**10**	63.33	76.67	N	
		MTS	0	0	1	0	0	1	3.33	**10**	0	0	33.33	13.33	N	N
	Session 3	SPO	**10**	**9**	**10**	**9**	**10**	**9**	**95.00**	**10**	**9**	**10**	**96.67**	95.56	**Y**	
		MTS	1	0	0	0	0	0	1.67	0	0	1	3.33	2.22	N	N
	Session 4	SPO	**9**	**10**	**9**	**10**	**10**	**10**	**96.67**	**10**	**10**	**10**	**100.00**	97.78	**Y**	
		MTS	1	0	0	0	0	1	3.33	0	0	0	0.00	2.22	N	N

Table 1.6

Number of correct responses for each trial type during testing of each training and testing cycle in each experimental session for the participants (in Group E1.Y.MTS.SPO) who reported that they could name the stimuli and completed MTS followed by SPO training.

Participant	Session	Condition	Tested relations - sample and correct comparison stimuli											Overall % Correct	Condition criterion met	Session criterion met
			Symmetry relations						% Correct Symmetry	Equivalence relations			% Correct Equivalence			
			B1-A1 (SPO) B1-A3 (MTS)	B2-A2 (SPO) B2-A1 (MTS)	B3-A3 (SPO) B3-A2 (MTS)	C1-B1 (SPO) C1-B3 (MTS)	C2-B2 (SPO) C2-B1 (MTS)	C3-B3 (SPO) C3-B2 (MTS)		C1-A1 (SPO) C1-A2 (MTS)	C2-A2 (SPO) C2-A3 (MTS)	C2-A2 (SPO) C3-A1 (MTS)				
P1.10	Session 1	SPO	0	**10**	0	**10**	2	**10**	53.33	0	0	**9**	30.00	45.56	N	
		MTS	**10**	0	**10**	4	3	1	46.67	**10**	**9**	0	63.33	52.22	N	N
	Session 2	SPO	0	**10**	0	1	0	0	18.33	0	1	**10**	36.67	24.44	N	
		MTS	**10**	0	**10**	**10**	5	**10**	75.00	**10**	**10**	0	66.67	72.22	N	N
	Session 3	SPO	0	**10**	0	0	4	5	31.67	0	1	**10**	36.67	33.33	N	
		MTS	**10**	0	**10**	**10**	2	8	66.67	**10**	8	0	60.00	64.44	N	N
	Session 4	SPO	0	**10**	0	3	1	3	28.33	0	0	**10**	33.33	30.00	N	
		MTS	**10**	0	**10**	7	0	5	53.33	**10**	8	0	60.00	55.56	N	N
P1.11	Session 1	SPO	5	7	**9**	6	8	5	66.67	1	2	2	16.67	50.00	N	
		MTS	5	6	5	6	7	7	60.00	8	4	2	46.67	55.56	N	N
	Session 2	SPO	0	0	0	0	0	0	0.00	0	0	0	0.00	0.00	N	
		MTS	**10**	**10**	**10**	**10**	**9**	**9**	**96.67**	**10**	**10**	**10**	**100.00**	97.78	**Y**	N
	Session 3	SPO	0	0	0	0	0	0	0.00	0	1	0	3.33	1.11	N	
		MTS	**10**	**10**	**10**	**10**	**10**	**10**	**100.00**	**9**	**9**	**10**	**93.33**	97.78	**Y**	N
	Session 4	SPO	0	0	0	1	0	0	1.67	0	0	0	0.00	1.11	N	
		MTS	**10**	**10**	**10**	**10**	**10**	**10**	**100.00**	**9**	**9**	**10**	**93.33**	97.78	**Y**	N
P1.12	Session 1	SPO	4	**10**	7	**10**	0	2	55.00	6	0	3	30.00	46.67	N	
		MTS	2	0	**9**	3	**9**	2	41.67	0	2	6	26.67	36.67	N	N
	Session 2	SPO	7	8	**10**	**10**	1	6	70.00	6	8	**9**	76.67	72.22	N	
		MTS	2	5	1	0	6	**9**	38.33	3	1	2	20.00	32.22	N	N
	Session 3	SPO	5	10	4	**9**	7	4	65.00	2	8	5	50.00	60.00	N	
		MTS	2	4	1	**9**	3	6	41.67	5	3	1	30.00	37.78	N	N
	Session 4	SPO	6	**10**	**10**	**10**	8	**10**	90.00	**10**	**10**	**9**	**96.67**	92.22	N	
		MTS	**10**	1	4	**9**	7	**10**	68.33	3	6	1	33.33	56.67	N	N

The results for the participants who reported that they could read the characters (Groups E1.Y.SPO.MTS and E1.Y.MTS.SPO) were mixed (as shown in Tables 1.5 and 1.6). Overall, in most sessions, most of these participants achieved greater percentages of correct responses following the training condition that they completed first in each session (the exception being P1.12 in Group E1.Y.MTS.SPO). Two participants (P1.9 and P1.11) in these groups met the condition criterion following the training condition that they experienced first. No effect of the order of the training procedures was evident for Groups E1.N.SPO.MTS and E1.N.MTS.SPO.

When accuracy on the symmetry and equivalence relations was assessed separately, the data for one of the groups who reported that they could not read the characters (Group E1.N.MTS.SPO) and for both of the groups who reported that they could read the characters (Groups E1.Y.SPO.MTS and E1.Y.MTS.SPO) tended to acheive greater percentages of correct responses on both the symmetry and equivalence relations following the procedure on which the participants had performed best overall. For the group who could not read the characters and completed SPO training prior to MTS training (Group E1.N.SPO.MTS), this pattern was seen for the symmetry, but not necessarily, the equivalence relations.

Figure 1.1 shows percent correct on the symmetry and equivalence trials following both SPO and MTS training for all participants in this experiment. The data show no consistent trends across the groups. Five of the participants, one in each of Groups E1.N.SPO.MTS, E1.N.MTS.SPO, and E1.Y.SPO.MTS (P1.3, P1.5, and P1.9) and two in Group E1.Y.MTS.SPO, (P1.11 and P1.12) were achieving a very high percent correct on both the symmetry and equivalence relations following one training condition but not the other by the end of the final session. For three of these participants (P1.5, P1.9, and P1.11), this was paired with a very low percent correct following the other training condition.

Data from the final session were used to conduct a factorial repeated-measures ANOVA to compare percent correct on the symmetry and equivalence relations across all groups. The results in Table 1.7 show there was a significant difference in the percent correct achieved on the symmetry and equivalence relations (type of relation) in the final session, and this showed an effect size of 0.406, which would be termed moderate by Fergusson (2009). There were no other significant within-subject effects or interactions. However, in most cases the effect size was above the recommended minimum practical effect (RMPE) size of 0.14 suggested by Ferguson

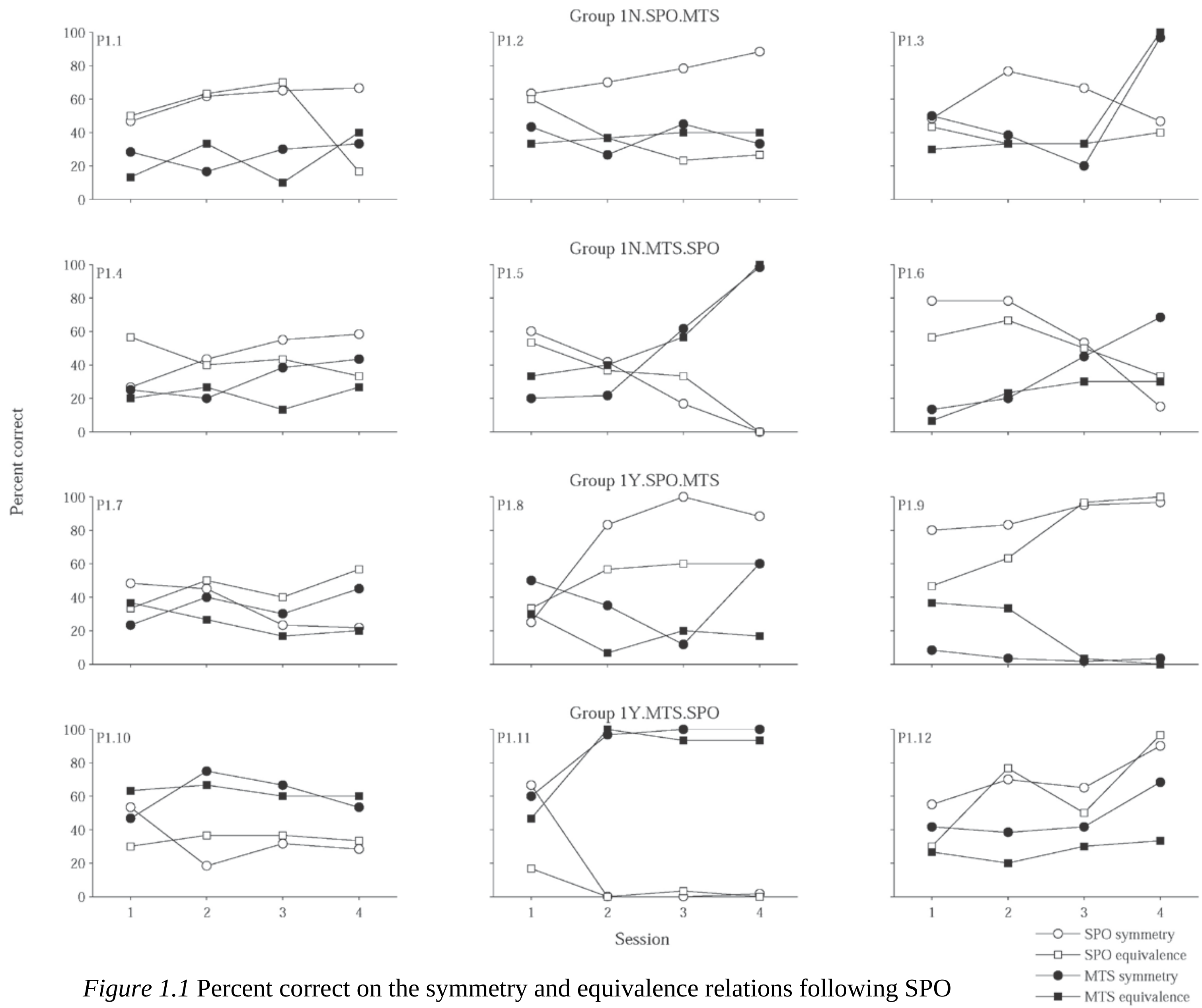

Figure 1.1 Percent correct on the symmetry and equivalence relations following SPO and MTS training across all sessions for all participants in Experiment 1.

(2009). Moderate effect sizes (>0.25 and <0.64) were seen for three of the non-significant results. No large effect sizes (>0.64) were observed. There were no significant between-subjects main effects, and the effect sizes were less than Ferguson's (2009) RMPE. The between-subjects interaction was not significant, and its effect size was small, but greater than Ferguson's (2009) RMPE.

Table 1.7

Results of the ANOVA to compare percent correct on the symmetry and equivalence relations of the final session for both procedures for all groups in Experiment 1.

Source	*df*	*F*	$\eta^2_{partial}$
Within-subjects effects			
Procedure	1,8	0.186	0.023
Procedure x procedural order	1,8	2.630	0.247
Procedure x can/can't read characters	1,8	1.185	0.129
Procedure x procedural order x can/can't read characters	1,8	0.524	0.062
Type of relation	1,8	5.467*	0.406
Type of relation x procedural order	1,8	0.512	0.060
Type of relation x can/can't read characters	1,8	0.559	0.065
Type of relation x procedural order x can/can't read characters	1,8	0.006	0.001
Procedure x Type of relation	1,8	0.111	0.014
Procedure x Type of relation x procedural order	1,8	1.997	0.200
Procedure x Type of relation x can/can't read characters	1,8	4.068	0.337
Procedure x Type of relation x procedural order * can/can't read characters	1,8	4.196	0.344
Between-subjects effects			
Procedural order	1,8	0.026	0.003
Can/can't read characters	1,8	0.249	0.030
Procedural order x can/can't read characters	1,8	1.477	0.156

*=significant at $p<0.05$

As Tables 1.3 - 1.6 show, all of the participants responded correctly on nine or more of the ten trials for some tested relations during the experiment. Overall, for the participants who could not read the Chinese characters, those who completed SPO training first in each session (Group E1.N.SPO.MTS) generally achieved nine or more correct responses on a greater number of the tested relations than the participants who completed MTS training first in each session

(Group E1.N.MTS.SPO). This difference was not evident between the groups who could read the Chinese characters (Groups E1.Y.SPO.MTS and E1.Y.MTS.SPO). However, overall, these groups responded correctly on nine or more trials for more tested relations than the groups who could not read the characters. Two participants (P1.5 and P1.11 in Groups E1.N.MTS.SPO and E1.Y.MTS.SPO) had achieved nine or more responses correct on all individual tested relations following MTS training by the final session, thus, meeting the condition criterion for that session. This was also achieved following SPO training by one participant (P1.9 in Group E1.Y.SPO.MTS). All three of these participants, achieved few, or no, correct responses on the tested relations following the other training procedure.

Correct vs. incorrect (compared to chance) during testing. Due to the failure of most participants to meet the condition criterion during testing following either SPO or MTS training in any experimental session, and the failure of all participants to meet the overall session criterion, χ^2 tests for goodness of fit were conducted to compare the performance of each participant to that predicted by chance. As three comparison stimuli were presented on each trial, a performance that was indistinguishable from chance would have been evidenced by an even distribution of responses across the response options. Therefore, if a participant's performance during one test condition was not different from chance, it was expected that approximately one third (30) of the responses would be correct, and two thirds (60) of the responses would be incorrect. χ^2 tests for goodness of fit were conducted to assess if the number of correct and incorrect responses made during each testing session by each participant differed significantly from a response distribution predicted by chance. In reports of these, and all other χ^2 tests in this section, terms regarding the 'significance' of results refer to statistical significance.

Table 1.8 shows the number of correct and incorrect responses for each condition of each experimental session, made by all of the participants in this experiment, and the χ^2 value for each χ^2 test. χ^2 values that gave significant results at $p<.05$ are indicated by an asterisk (*). As shown in Table 1.8, overall, a significantly greater number of correct responses than chance were recorded during testing following SPO training for all participants in Group E1.N.SPO.MTS, and following MTS training for two of the three participants in Group E1.Y.MTS.SPO (P1.10 and

P1.11). The results for Groups E1.N.MTS.SPO and E1.Y.SPO.MTS showed no consistent pattern; however, results that were significantly different from chance were observed during the majority of the experimental sessions for both groups. Across all of the groups, where results showed a significantly greater number of errors than predicted by chance, these were, in most cases, paired with results that showed a significantly greater number of correct responses following the other training condition of that experimental session.

Consistent vs. inconsistent errors during testing. χ^2 tests for goodness of fit were calculated to assess the distribution of errors during each condition, of each experimental session for all of the participants in this experiment. Of the two incorrect comparison stimuli presented on each trial, one was the stimulus that was the correct stimulus during testing following the other training condition, and one was an incorrect stimulus following both conditions. For example, in testing following SPO training, if A1 was presented as the sample stimulus, then B1, B2, and B3 were provided as comparison stimuli. Of these, B1 was the correct stimulus, B2 was an incorrect stimulus that served as a correct stimulus following MTS training, and B3 was incorrect in both conditions. Therefore, errors could be divided into those that were incorrect, but were consistent with the correct stimulus following the other training condition, and those that were incorrect following both conditions. These are referred to, henceforth, as inconsistent errors. Tables showing the distribution of responses across each alternative during each experimental session by each participant in this experiment are given in Appendix A.

Table 1.9 shows the number of consistent and inconsistent errors made during testing following SPO and MTS training of each experimental session, for all of the participants in this experiment, and the obtained χ^2 value for each test. An asterisk (*) is indicative of a significant difference in the number of consistent and inconsistent errors at $p<.05$. Table 1.9 shows that few participants produced a systematic pattern of errors across the sessions. Three participants, P1.1 (Group1 E1.N.SPO.MTS), P1.8 and, P1.9 (both in Group E1.Y.SPO.MTS) made significantly more consistent than inconsistent errors following MTS training in all, or most, sessions. This was also true following SPO training for P1.11 (Group E1.Y.SPO.MTS). One participant (P1.10 in Group E1.Y.SPO.MTS) made a significantly greater number of consistent than inconsistent

Table 1.8

Chi-square tests for the distribution of correct (Corr.) and incorrect responses (Incor.) compared to the distribution of responses predicted by chance (correct (30/90), incorrect (60/90) for all participants in Experiment 1.

Group	Participant	Session 1						Session 2						Session 3						Session 4					
		SPO			MTS			SPO			MTS			SPO			MTS			SPO			MTS		
		Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2
E1.N.SPO.MTS	P1.1	43	47	8.450*	21	69	4.050*	56	34	33.800*	20	70	5.000*	60	30	45.000*	21	69	4.050*	45	45	11.250*	32	58	0.200
E1.N.SPO.MTS	P1.2	56	34	33.800*	36	54	1.800	53	37	26.450*	27	63	0.450	54	36	28.800*	39	51	4.050*	61	29	48.050*	32	58	0.200
E1.N.SPO.MTS	P1.3	42	48	7.200*	39	51	4.050*	56	34	33.800*	33	57	0.450	50	40	20.000*	22	68	3.200	40	50	5.000*	88	2	168.200*
E1.N.MTS.SPO	P1.4	33	57	0.450	21	69	4.050*	38	52	3.200*	20	70	5.000*	46	44	12.800*	27	63	0.450	45	45	11.250*	34	56	0.800
E1.N.MTS.SPO	P1.5	52	38	24.200*	22	68	3.200	36	54	1.800	25	65	1.250	20	70	5.000*	54	36	28.800*	0	90	45.000*	89	1	174.050*
E1.N.MTS.SPO	P1.6	64	26	57.800*	10	80	20.000*	67	23	68.450*	19	71	6.050*	47	43	14.450*	36	54	1.800	19	71	6.050*	50	40	20.000*
E1.Y.SPO.MTS	P1.7	39	51	4.050*	25	65	1.250	42	48	7.200*	32	58	0.200	26	64	0.800	23	67	2.450	30	60	0.000	33	57	0.450
E1.Y.SPO.MTS	P1.8	25	65	1.250	39	51	4.050*	67	23	68.450*	23	67	2.450	78	12	115.200*	13	77	14.450*	71	19	84.050*	41	49	6.050*
E1.Y.SPO.MTS	P1.9	62	28	51.200*	16	74	9.800*	69	21	76.050*	12	78	16.200*	86	4	156.800*	2	88	39.200*	88	2	168.200*	2	88	39.200*
E1.Y.MTS.SPO	P1.10	41	49	6.050*	47	43	14.450*	22	68	3.200	65	25	61.250*	30	60	0.000	58	32	39.200*	27	63	0.450	50	40	20.000*
E1.Y.MTS.SPO	P1.11	45	45	11.250*	50	40	20.000*	0	90	45.000*	88	2	168.200*	1	89	42.050*	88	2	168.200*	1	89	42.050*	88	2	168.200*
E1.Y.MTS.SPO	P1.12	42	48	7.200*	33	57	0.450	65	25	61.250*	29	61	0.050	54	36	28.800*	34	56	0.800	83	7	140.050*	51	39	22.050*

*=significant at $p < .05$

Table 1.9

Chi-square tests on the distribution of errors (incorrect and consistent with the other condition (Con.), or incorrect and inconsistent with the other condition (Incon.)) for all participants in Experiment 1.

Group	Participant	Session 1						Session 2						Session 3						Session 4					
		SPO			MTS			SPO			MTS			SPO			MTS			SPO			MTS		
		Con.	Incon.	χ^2	Con.	Incon.	χ^2	Con.	Incon.	χ^2	Con.	Incon.	χ^2	Con.	Incon.	χ^2	Con.	Incon.	χ^2	Con.	Incon.	χ^2	Con.	Incon.	χ^2
E1.N.SPO.MTS	P1.1	28	19	1.723	47	22	9.058*	20	14	1.059	58	12	30.229*	17	13	0.533	45	24	6.391*	26	19	1.089	39	19	6.897*
E1.N.SPO.MTS	P1.2	22	12	2.941*	26	28	0.074	20	17	0.243	35	28	0.778	18	18	0.000	28	23	0.490	13	16	0.310	25	33	1.103
E1.N.SPO.MTS	P1.3	25	23	0.083	29	22	0.961	12	22	2.941	27	30	0.158	20	20	0.000	48	20	11.529*	37	13	11.520*	2	0	2.000
E1.N.MTS.SPO	P1.4	17	40	9.281*	33	36	0.130	21	31	1.923	23	47	8.229*	25	19	0.818	36	27	1.286*	34	11	11.756*	26	30	0.286
E1.N.MTS.SPO	P1.5	15	23	1.684	38	30	0.941	26	28	0.074	40	25	3.462	50	20	12.857*	16	20	0.444	89	1	86.044*	0	1	1.000
E1.N.MTS.SPO	P1.6	13	13	0.000	58	22	16.200*	9	14	1.087	59	12	31.113*	19	24	0.581	34	20	3.630	60	11	33.817*	21	19	0.100
E1.Y.SPO.MTS	P1.7	25	26	0.020	42	23	5.554*	18	30	3.000	45	13	17.655*	22	42	6.250*	29	38	1.209	31	29	0.067	29	28	0.018
E1.Y.SPO.MTS	P1.8	27	38	1.862	21	30	1.923	12	11	0.043	51	16	18.284*	2	10	5.333	69	8	48.325	8	11	0.474	36	13	10.796
E1.Y.SPO.MTS	P1.9	13	15	0.143	61	14	29.453*	12	9	0.429	68	10	43.128*	1	3	1.000	88	0	88.000*	0	2	2.000	86	2	80.182*
E1.Y.MTS.SPO	P1.10	48	1	45.082*	27	16	2.814	58	10	33.882*	20	5	9.000*	53	7	35.267*	28	4	18.000*	51	12	24.143*	30	10	10.000*
E1.Y.MTS.SPO	P1.11	14	31	6.422*	17	23	0.900	90	0	90.000*	1	1	0.000	89	0	89.000*	2	0	2.000*	89	0	89.000*	1	1	0.000
E1.Y.MTS.SPO	P1.12	26	22	0.333	33	24	1.421	12	13	0.040	37	24	2.77	21	15	0.317	36	20	4.571*	6	1	3.571	18	21	0.231

*=significant at p <.05

errors following both training procedures in most sessions. For the remaining participants, there was no consistent pattern of error type, however, where significant differences did exist they were indicative of a greater number of consistent than inconsistent errors, with few exceptions.

MTS Training

Number correct on each trained relation and overall percent correct. Table 1.10 and Figure 1.2 show the number of correct responses made on each trial for each relation during MTS training across the sessions, for all participants. Table 1.10 also shows the total percent correct achieved during MTS training of each session for each participant. While there was no criterion on accuracy during training, Table 1.10 and Figure 1.2 show no consistent pattern across any of the groups in the number of correct responses on each trained relation across the sessions. The number of correct responses on each relation for four participants, one in each of Groups 1.N.SPO.MTS and E1.Y.MTS.SPO, and two in Group E1.N.MS.SPO, trended upwards across the sessions for all trained relations. These participants achieved nine or more correct responses on all trained relations during MTS training by the final session. The number of correct responses made by all remaining participants (P1.1 and P1.2 in Group E1.N.SPO.MTS, all participants in Group E1.N.MTS.SPO, P1.8 in Group E1.Y.SPO.MTS, and P1.10 and P1.12 in Group E1.Y.MTS.PS) increased across the sessions for some trained relations, but not others. None of these participants achieved nine or more correct responses on all trained relations of any session. However, most of them (P1.5, P1.6, P1.8, P1.10, and P1.12) did respond correctly on nine or more trials for some trained relations. The percentage of correct responses on some trained relations decreased across the sessions for a small number of participants (P1.2, P1.8, and P1.12).

Correct vs. incorrect (compared to chance) during MTS training. χ^2 tests for goodness of fit were conducted to compare the distribution of correct and incorrect responses to that predicted by chance. As with testing, three response options were provided to the participants on each trial. Therefore, a response distribution indistinguishable from the distribution predicted by chance would result in one third of responses correct (20/60) and two-thirds of responses incorrect (40/60). Table 1.11 shows the number of correct and incorrect responses during MTS

Table 1.10

Number of correct responses for each trained relation during MTS training of each experimental session for all participants in Experiment 1.

Group	Participant	Session	Sample and correct comparison stimuli						% Correct
			A1-B2	A2-B3	A3-B1	B1C2	B2-C3	B3-C1	
E1.N.SPO.MTS	P1.1	Session 1	0	2	0	0	0	0	3.33
		Session 2	6	3	0	0	0	0	15.00
		Session 3	6	2	4	1	5	0	30.00
		Session 4	7	8	2	4	1	1	38.33
E1.N.SPO.MTS	P1.2	Session 1	5	4	8	4	1	3	41.67
		Session 2	1	2	2	1	0	3	15.00
		Session 3	2	2	4	5	4	2	31.67
		Session 4	0	1	8	3	3	4	31.67
E1.N.SPO.MTS	P1.3	Session 1	6	9	5	7	1	2	50.00
		Session 2	**10**	8	8	**9**	8	8	85.00
		Session 3	**10**	**10**	**10**	**10**	**10**	**10**	**100.00**
		Session 4	**10**	**10**	**10**	**10**	**9**	**10**	**98.33**
E1.N.MTS.SPO	P1.4	Session 1	3	4	0	2	2	5	26.67
		Session 2	0	2	3	4	1	3	21.67
		Session 3	2	2	4	1	2	2	21.67
		Session 4	4	6	2	6	5	5	46.67
E1.N.MTS.SPO	P1.5	Session 1	2	4	1	4	4	4	31.67
		Session 2	3	3	2	3	1	5	28.33
		Session 3	1	3	**10**	3	5	1	38.33
		Session 4	**10**	**10**	**10**	4	**9**	6	81.67
E1.N.MTS.SPO	P1.6	Session 1	3	4	4	4	1	4	33.33
		Session 2	1	2	2	4	2	2	21.67
		Session 3	5	6	8	6	4	2	51.67
		Session 4	2	**9**	**9**	3	5	8	60.00

Table 1.10 continued.

Group	Participant	Session	Sample and correct comparison stimuli						% Correct
			A1-B2	A2-B3	A3-B1	B1C2	B2-C3	B3-C1	
E1.Y.SPO.MTS	P1.7	Session 1	3	4	**8**	1	5	2	38.33
		Session 2	5	4	**9**	5	**9**	7	65.00
		Session 3	**10**	**9**	**10**	**9**	**9**	**9**	**93.33**
		Session 4	**10**	**10**	**10**	**10**	**10**	**9**	**98.33**
E1.Y.SPO.MTS	P1.8	Session 1	1	**9**	8	**9**	**9**	1	61.67
		Session 2	**9**	**9**	7	6	5	3	65.00
		Session 3	**9**	4	7	7	5	3	58.33
		Session 4	**10**	**10**	8	3	7	6	73.33
E1.Y.SPO.MTS	P1.9	Session 1	4	5	8	4	5	5	51.67
		Session 2	**10**	**9**	**9**	**10**	**10**	**10**	**96.67**
		Session 3	8	**10**	**10**	**10**	**10**	**10**	96.67
		Session 4	**9**	**9**	**10**	**10**	**10**	**10**	**96.67**
E1.Y.MTS.SPO	P1.10	Session 1	8	**10**	**10**	1	7	7	71.67
		Session 2	5	7	**10**	3	6	8	65.00
		Session 3	**9**	8	**10**	3	5	8	71.67
		Session 4	**10**	**9**	**10**	6	7	**9**	85.00
E1.Y.MTS.SPO	P1.11	Session 1	2	**10**	5	4	8	4	55.00
		Session 2	**9**	**10**	**10**	**10**	**10**	**9**	**96.67**
		Session 3	**10**	**10**	**10**	**10**	**10**	**10**	**100.00**
		Session 4	**10**	**10**	**10**	**9**	**10**	**10**	**98.33**
E1.Y.MTS.SPO	P1.12	Session 1	7	6	3	**10**	8	**10**	73.33
		Session 2	3	1	5	4	**10**	0	38.33
		Session 3	4	**9**	4	8	**10**	1	60.00
		Session 4	1	1	5	8	**10**	0	41.67

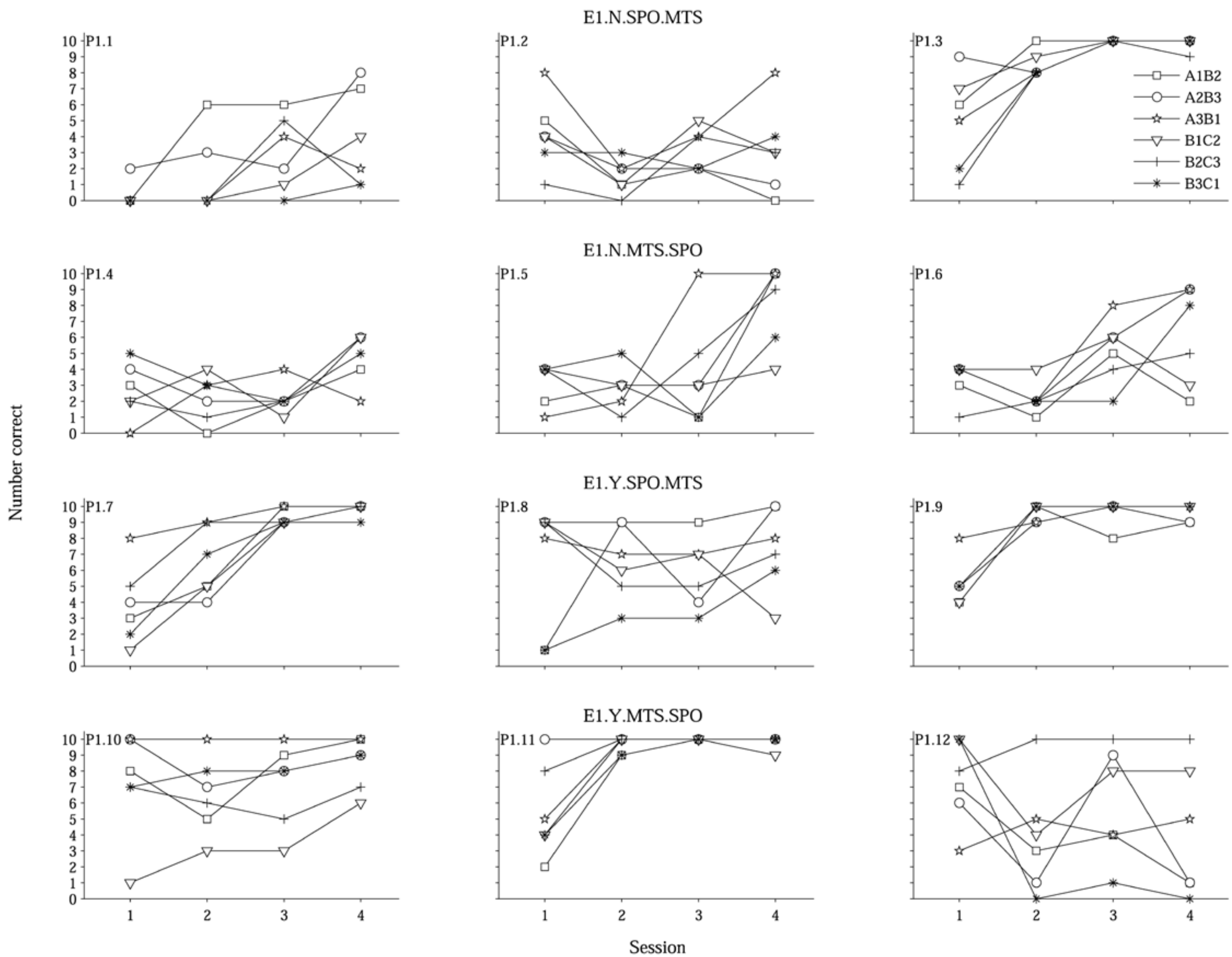

Figure 1.2 Number of correct responses made on each trained relation during MTS training across all sessions for all participants in Experiment 1.

training of each session for all participants in this experiment, and the χ^2 test statistics. χ^2 values that show a response distribution that is significantly different from that predicted by chance (at p<.0.05) are indicated by an asterisk (*). Table 1.11 shows that there was no consistent pattern for Group E1.N.SPO.MTS, with one participant making a greater number of correct responses than predicted by chance during MTS training of all experimental sessions, and two participants failing to do so during any session. The participants in Group E1.N.MTS.SPO were all producing a greater number of correct responses than predicted by chance by the fourth session of MTS training. In contrast, the participants in Groups E1.Y.SPO.MTS and E1.Y.MTS.SPO responded correctly on a greater number of training trials than predicted by chance during all or most of the experimental sessions.

Consistent vs. inconsistent errors during MTS training. χ^2 tests for goodness of fit were conducted to compare the number of errors that were consistent with the correct pairing for the SPO condition, and those that were incorrect in both conditions (inconsistent errors). Table 1.12 shows the number of consistent and inconsistent errors during MTS training of each session for all participants in this experiment, and the statistic for each χ^2 test. χ^2 values that were significant at $p<0.05$ are indicated by an asterisk (*). No clear pattern of consistent and inconsistent errors during MTS training was observed within participants or across groups of this experiment. In few cases were the numbers of each type of error significantly different. Where these did occur, in all but one case, they were indicative of a significantly greater number of errors that were consistent with the correct pairing during SPO training than errors that were incorrect in both conditions.

Table 1.11

Chi-square tests for the distribution of correct and incorrect responses during MTS training (compared to responses expected by chance (correct (20/60), incorrect (40/60)) by all participants in Experiment 1.

Group	Participant	Session 1			Session 2			Session 3			Session 4		
		Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2
E1.N.SPO.MTS	P1.1	2	58	24.300*	9	51	9.075*	18	42	0.300	23	37	0.675
E1.N.SPO.MTS	P1.2	25	35	1.875	9	51	9.075*	19	41	0.075	19	41	0.075
E1.N.SPO.MTS	P1.3	30	30	7.500*	51	9	72.075*	60	0	120.000*	59	1	114.075*
E1.N.MTS.SPO	P1.4	16	44	1.200	13	47	3.675	13	47	3.675	28	32	4.800*
E1.N.MTS.SPO	P1.5	19	41	0.075	17	43	0.675	23	37	0.675	49	11	63.075*
E1.N.MTS.SPO	P1.6	20	40	0.000	13	47	3.675	31	29	9.075*	36	24	19.200*
E1.Y.SPO.MTS	P1.7	23	37	0.675	39	21	27.075*	56	4	97.200*	59	1	114.075*
E1.Y.SPO.MTS	P1.8	37	23	21.675*	39	21	27.075*	35	25	16.875*	44	16	43.200*
E1.Y.SPO.MTS	P1.9	31	29	9.075*	58	2	108.300*	58	2	108.300*	58	2	108.300*
E1.Y.MTS.SPO	P1.10	43	17	39.675*	39	21	27.075*	43	17	39.675*	51	9	72.075*
E1.Y.MTS.SPO	P1.11	33	27	12.675*	58	2	108.300*	60	0	120.000*	59	1	114.075*
E1.Y.MTS.SPO	P1.12	44	16	43.200*	23	37	0.675	36	24	19.200*	25	35	1.875

*=significant at $p < .05$

Table 1.12

Chi-square tests on the distribution of errors (incorrect and consistent with the other condition (Con.), or incorrect and inconsistent with the other condition (Incon.)) during MTS training for all participants in Experiment 1

Group	Participant	Session 1			Session 2			Session 3			Session 4		
		Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2
E1.N.SPO.MTS	P1.1	39	19	6.897*	37	14	10.373*	27	15	3.429	23	14	2.189
E1.N.SPO.MTS	P1.2	21	14	1.400	26	25	0.020	27	14	4.122*	20	21	0.024
E1.N.SPO.MTS	P1.3	17	13	0.533	6	3	1.000	0	0	-	1	0	1.000
E1.N.MTS.SPO	P1.4	25	19	0.818	16	31	4.787*	29	18	2.574	20	12	2.000
E1.N.MTS.SPO	P1.5	24	17	1.195	27	16	2.814	16	21	0.676	5	6	0.091
E1.N.MTS.SPO	P1.6	20	20	0.000	27	20	1.043	13	16	0.310	10	14	0.667
E1.Y.SPO.MTS	P1.7	26	11	6.081*	13	8	1.190	3	1	1.000	0	1	1.000
E1.Y.SPO.MTS	P1.8	11	12	0.043	16	5	5.762*	16	9	1.960	12	4	4.000*
E1.Y.SPO.MTS	P1.9	14	15	0.034	1	1	0.000	2	0	2.000	2	0	2.000
E1.Y.MTS.SPO	P1.10	10	7	0.529	16	5	5.762*	15	2	9.941*	7	2	2.778
E1.Y.MTS.SPO	P1.11	17	10	1.815	2	0	2.000	0	0	2.000	1	0	1.000
E1.Y.MTS.SPO	P1.12	9	7	0.250	20	17	0.243	12	12	0.000	24	11	4.829*

*=significant at p <.05

Discussion

Experiment 1 aimed to provide a procedural replication of the first experiment by Leader and Barnes-Holmes (2001b), but used a selection of the Chinese characters used by Clayton and Hayes (2004). Additionally, two groups were added to the present experiment to include participants who reported that they could read the Chinese characters. There was no difference in overall performance between the SPO and MTS training procedures for any group. This finding is not consistent with those of either Experiment 1 by Leader and Barnes-Holmes (2001b) or by Clayton and Hayes (2004). In the present study, the accuracy achieved on the symmetry relations was significantly better than the accuracy achieved on the equivalence relations regardless of the training procedure. Leader and Barnes-Holmes (2001b) do not present the results for the symmetry and equivalence relations separately so no comparison is possible; however, most participants met the criterion during testing following both training procedures. In order to achieve this result, accuracy would have been above 90% on both the symmetry and equivalence relations. Clayton and Hayes (2004) found no clear difference in accuracy on the symmetry and equivalence trials during testing following either training procedure.

Leader and Barnes-Holmes (2001b) reported that the SPO procedure was more effective in facilitating the formation of equivalence relations than the MTS procedure in their first experiment. Overall, their results showed both procedures to be more effective at facilitating the formation of equivalence relations than was shown here.

The only difference between the present study and Experiment 1 by Leader and Barnes-Holmes (2001b) was the use of the Chinese characters as stimuli. Therefore it is possible that the Chinese characters made the task more difficult, resulting in a lesser likelihood of equivalence here. A study by Holth and Arntzen (1998) showed that the type of stimuli used can affect outcomes on equivalence tasks. In that study, participants were less likely to demonstrate equivalence when the stimuli were arbitrary Greek letters than when they were familiar, nameable, pictures. While the nonsense syllables used by Leader and Barnes-Holmes (2001b) can be considered arbitrary, it is likely that they would have been more ‘nameable’ than the Chinese characters used in the present study for the participants who could not read the Chinese characters.

As Leader and Barnes-Holmes's (2001b) participants could read the stimuli (nonsense syllables), they were most like the participants in the present study who reported that they could read the Chinese characters (Groups E1.Y.SPO.MTS and E1.Y.MTS.SPO). Leader and Barnes-Holmes (2001b) found in favour of the SPO training condition regardless of the condition order experienced by the participants. Their participants were more likely to meet the condition criterion following both procedures if they completed SPO training prior to MTS training in each session, but they still met the criterion following SPO training sooner than meeting the criterion following MTS training.

Leader and Barnes-Holmes's (2001b) participants completed a maximum of six repetitions. In the present study, the number of sessions, and therefore, the number of repetitions, was limited to four. This limitation was due to constraints on the maximum course credit that could be provided for participation. The participants here completed a maximum of 240 training trials across the maximum of four sessions. With six repetitions, Leader and Barnes-Holmes's (2001b) participants could have completed 360 training trials during the experiment. It could be argued, then, that the fewer training trials completed by the participants in the present experiment accounts for the poorer performance compared to that reported by Leader and Barnes-Holmes (2001b). However, all of the participants in Experiment 1 of that study who achieved the criterion following each training procedure did so within the first four repetitions. As there was no clear trend in percent correct achieved on the symmetry and equivalence relations in the present experiment it is unlikely that extra sessions would have been beneficial for most participants here. The training data did show an upwards trend across the sessions for a few participants. For these participants, extra sessions may have resulted in better performance on the training relations, which could have then resulted in better performance on the symmetry and equivalence relations. However, this is true for a few participants only, and does not explain the different findings of the present study and Leader and Barnes-Holmes (2001b).

In the present study, the order in which the training procedures were completed in a session appeared to determine the effectiveness of the training procedure for most participants who could read the Chinese characters. However, there was no significant interaction between the procedural order and the ability of the participants to read the characters during the final session. It is unclear why the results differ from those reported by Leader and Barnes-Holmes (2001b). However, nonsense syllables, as used by Leader and Barnes-Holmes (2001b), are generally considered to be without meaning. In contrast, for the participants who could read the characters, these stimuli carried their own associations and meanings prior to the experiment and this may have contributed to the differences in the findings.

Some research has studied the effect of pre-existing associations or meanings of non-arbitrary stimuli on the formation of equivalence relations. For example, some (Eikeseth & Baer, 1997; Stewart, Barnes-Holmes, Roche & Smeets, 2002; Ybarra Sagarduy, Soriano & Gomez Martin, 2002) have examined the effect of other associations with- or meanings of stimuli that existed pre-experimentally on the formation of new equivalence relations. These studies reported that pre-existing relations involving alphabetic order (Eikeseth & Baer, 1997; Ybarra Sagarduy et al., 2002), numerical order (Ybarra Sagarduy et al., 2002), colours (Stewart et al., 2002) or experimentally induced order relations involving verses from a poem (Ybarra Sagarduy et al., 2002) interfered with the formation of new relations when the existing relations were incompatible with those trained and tested in the experiment. Other studies have extended this research to examine the effect of pre-existing associations on the formation of equivalence classes by specific populations, including people who are highly anxious (Leslie, Tierney, Robinson & Keenan, 1993; Merwin & Wilson, 2005), children with intellectual disabilities (Barnes, Lawlor, Smeets & Roche, 1996) and low-achieving students (Adcock, Merwin, Wilson, Drake, Tucker & Elliot, 2010). As with the previously mentioned studies, these studies reported that the pre-experimentally formed associations interfered with the formation of new equivalence relations when the stimuli had conflicting meanings.

Other studies (e.g., Carr & Blackman, 2001; Dickins, Bentall & Smith, 1993; Peoples, Tierney, Bracken & McKay, 1998; Roche, Barnes & Smeets, 1997) have looked at the effect of conflicting associations that were induced as part of the experimental process. These studies have examined the effect of conflicting relations based on participant-generated stimuli names (Dickins et al., 1993), associations with the onset of a sexual or non-sexual film (Roche et al., 1997), or positive or negative adjectives (Peoples et al., 1998), and associations with novel, but conflicting, stimuli (Carl & Blackman, 2001). In all of these studies, the experimentally trained relations interfered with the formation of equivalence classes involving stimuli that had conflicting meanings or associations.

Several studies (Dixon et al., 2006; Moxon, Keenan & Hine, 1993; Watt, Keenan, Barnes & Cairns, 1991) have examined the effect of previously developed social categorisation on the formation of equivalence relations. In these studies, socially derived associations were found to interfere with the formation of equivalence classes that involved socially conflicting relations between religious (Watt et al., 1991), gender-based (Moxon et al., 1993), or terrorism-based and patriotic (Dixon et al., 2006) stimuli.

All of the studies outlined above suggest that pre-existing associations involving the experimental stimuli (or the meanings of the experimental stimuli) can interfere with the formation of equivalence classes when the stimuli in a class are involved in conflicting associations (or have conflicting meanings). For most participants in the present experiment, the Chinese characters and nonsense syllables were arbitrary. That is, they were not part of pre-experimentally developed associations. However, for the participants who could read the Chinese characters, the characters had meanings that existed outside of the experimental setting. Table 1.13 shows the English translation of the most common definition for each of the Chinese characters used in Experiment 1. The pre-existing meanings of these characters may have made some relations easier, or more difficult, to learn than others and may have interfered with the formation of the correct relations. For example, during symmetry tasks following both SPO or MTS training in the present experiment, when C1 (dance / to dance) was presented as a sample stimulus, the comparison stimuli were B1 (abundant), B2 (song/ to sing), and B3 (new). The correct responses were choosing B1 (abundant) or B3 (new) during testing following SPO and MTS training, respectively. However, the sample stimulus (B1, dance) may have been associated with B2 (song) based on their common meanings. On inspection of the data for these stimuli (see Appendix A for the response distribution tables), three of the participants (P1.7, P1.8 and P1.9) who could read the Chinese characters chose the character meaning ‘song’ (B1) more often than they chose either of the other available characters in the presence of the character meaning ‘dance’ (C1) during some sessions. These sessions followed both procedures equally, and this pattern was more likely to occur in early sessions. By the final session, two of these participants were choosing either of the other two options (the correct character or the character that was the incorrect response but was consistent with the other procedure) most often. Therefore, it seems that a pre-existing relation between these stimuli interfered with the formation of equivalence classes by these participants, at least initially.

Table 1.13

The Chinese characters used in Experiment 1, with their alphanumeric designation and common definition in English.

Chinese characters	Alphanumeric designation	Common definition
错	A1	a mistake / mistaken
枝	A2	branches of a plant
死	A3	death / to die
博	B1	abundant / plentiful / broadly knowledgeable
歌	B2	a song / to sing
新	B3	new
舞	C1	a dance / to dance
姜	C2	ginger / a Chinese family name
帝	C3	a ruler / a monarch

The interaction between procedural order and the ability of the participants to read the characters (discussed above) was not significant. In many studies, where statistical methods have been used to examine effects they have relied on null-hypothesis significance testing (NHST). Though helpful, these tests only inform us about one of the possible dimensions of the observed difference, in this case the chance of observing a difference of this size when in truth no difference can be reliably observed. Another equally important dimension is the size of the observed difference or ‘effect size’. Therefore, the usefulness of NHST has been debated and various problems have been identified (Cohen, 1994; Balluerka, Gomez, & Hidalgo, 2005; Ferguson, 2009; Rosnow & Rosenthal, 2003; Wilkinson & Task Force on Statistical Inference, 1999;). Criticisms of NHST focus on three factors. These are: the sensitivity of statistical significance to sample size (Ferguson, 2009); that NHST does not denote practical or clinical significance; and that there are nearly always, at least small, differences between the sample means (Kirk, 1996). The first two of these criticisms are particularly relevant to the present study. The obtained test statistic in NHST is affected by the size of the sample. For the same magnitude of effect, the larger the sample, the more likely it is the obtained test statistic will be statistically significant (Rosnow & Rosenthal, 2003). This means that, for large samples, the same magnitude of effect that results in statistical significance may not do so in a small sample.

The second problem with NHST is that a statistically significant result does not determine the clinical, or practical, significance of the observed effect. Practical significance is defined by Kirk (1996) as “whether the result is useful in the real world” (p.746). As statistical significance is affected by sample size, it is possible for a small sample with a non-significant result to demonstrate greater practical significance than a significant result obtained with a large sample.

An outcome of this debate is the recommendation that all analyses that employ NHST should also report effect sizes (American Psychological Association, 2010; Wilkinson & Task Force on Statistical Inference, 1999). Cohen (1994) and, more recently, Ferguson (2009) have also argued that effect size is a more useful measure of the importance of the outcome. Cohen (1988) defines effect size as “the degree to which the phenomenon is present in the population” (p.9). Effect size measures are not affected by differences in sample size (Ferguson, 2009), and therefore, allow the comparison of results across studies with differing sample sizes (Ferguson, 2009; Nakagawa & Foster, 2004). However, there are no set rules on the magnitude of effect that demonstrates practical significance. The interpretation of the effect size depends on the context of the research (Rosnow & Rosenthal, 2003). For example, in biomedical research the effect sizes that are important are often very small as the “real-life implications” (Rosnow & Rosenthal, 2003, p.226) of a small effect may still be very serious (for example, if the outcome is death). Ferguson (2009) discusses the use of effect size within the social sciences. He recommends the use of strength of association measures for data that are continuous. One such measure is partial eta-squared ($\eta^2_{partial}$). Ferguson (2009) suggests the following convention for interpreting effect size measured by partial eta-squared. An effect size of .04 is his “recommended minimum effect size representing a “practically” significant effect” (p.533) (abbreviated to RMPE). He defines moderate and strong effects as .25 and .64 respectively. Like Cohen (1994), Ferguson (2009) argues that effect size is a more relevant measure of outcome than whether or not the result is statistically significant. In this study, the effect size for the interaction between procedural order and ability to read the characters is 0.156, and so is above Ferguson's (2009) RMPE. Thus, it represents a practically significant effect. If an effect size over the RMPE is taken to be important, then this finding should not be ignored.

The Chinese characters used in the present study were some of the characters used in the study by Clayton and Hayes (2004). As nearly all of the participants in that study were unable to read them, the results for the participants in the present study who could not read the Chinese characters can be compared with their findings. Clayton and Hayes (2004) reported that, overall, the development of equivalence relations was more likely following MTS training than SPO training. The findings of the present study, for those participants who could not read the Chinese characters did not show either procedure to be the most effective. Half of these participants achieved greater percent corrects following MTS training, and half performed best following SPO training. However, during the final session, this difference was not significant irrespective of whether or not the participants could read the characters, and the effect size was smaller than Ferguson's (2009) RMPE.

The first language for two of Clayton and Hayes's participants (across the three experiments) was Japanese. For these two participants, SPO training proved more effective than MTS training. Clayton and Hayes (2004) argue that the greater familiarity of these types of characters for these participants may account for these differing results.

Clayton and Hayes (2004) counterbalanced the order of the training procedures, but they do not report which order was experienced by individual participants. As such, it is not possible to compare the performance of their participants who were familiar with the stimuli to those participants in the present study who could read the characters. Therefore, the present findings do not clearly support or refute the findings by Clayton and Hayes (2004). As the procedure used by Clayton and Hayes differs in many respects from the one used in the present study, direct comparisons between the findings of the studies should be taken with caution, however, the lack of similarity in the findings suggests that the Chinese characters were not, at least solely, responsible for the different results of the studies by Leader and Barnes-Holmes (2001b) and

Clayton and Hayes (2004). If this had been the case, it would be expected that the use of the Chinese characters in a procedure similar to that used by Leader and Barnes-Holmes (2001b) would result in findings in favour of the MTS procedure.

All but one of the participants in the experiment by Leader and Barnes-Holmes (2001b) met the condition criterion following SPO training. Half of their participants met the condition criterion following MTS training, and met the session criterion (meeting the condition criterion following both training procedures in one session) by the end of the experiment. Only one of the participants in the study by Leader and Barnes-Holmes (2001b) failed to meet the condition criterion following either training procedure in all experimental sessions. In contrast, none of the participants in the present study met the session condition in any session, and few participants met the condition criterion following either training condition of any session. However, chi square tests showed that in most cases the participants were making more correct responses than predicted by chance during testing following one or both of the training conditions in each session. This suggests that both methods were, at least partially, effective at facilitating the formation of equivalence relations, but did not result in the accuracy required by the criterion.

On closer inspection, at the level of the individual tested relation, the number of correct responses was not consistent across all of the tested relations in a session. Nearly all participants responded correctly on nine or more trials for some tested relations during a session, and performed poorly on others. Therefore, using the total numbers of correct and incorrect responses within a session to determine if a performance differed significantly from that predicted by chance did not give a clear picture of how the correct and incorrect responses were distributed across the nine tested relations. A response distribution that was not significantly different from chance did not necessarily indicate that the participant was performing equally poorly across all tested relations. Rather, in most cases, the participants were performing well on some tested relations, and poorly on others. Thus, the low overall percentages of correct responses obtained were often the result of very good performance on some tested relations and very poor performance on other relations. This could suggest that some relations were easier to learn than others. However, as there was no clear pattern in the relations that were or were not learned across participants, this is unlikely. Additionally, as the pairs of stimuli used in the present experiment were the same for all participants (e.g., A1-B1 involved the same two

Chinese characters for all participants) it would be possible to identify if relations involving certain stimuli were easier to learn than others. There were no clear patterns in the relations learned to suggest that the relations involving any individual stimulus were easier to learn than others. The use of the same stimulus relations across subjects is different from the procedure of Experiment 1 by Leader and Barnes-Holmes (2001b). They balanced the presentation of stimuli across the participants. Stimuli are generally balanced to avoid stimulus-specific response patterns. However, it can be argued that this does not remove these patterns, but obscures them so that they cannot be identified easily. If the presence of a particular stimulus in a stimuli pair was to affect the performance of participants on trials in which that stimulus appeared, then this effect would be identified easily when the same stimuli are used in the same relations across all participants. Balancing the presentation of stimuli across the stimulus relations does not remove any effect caused by individual stimuli but distributes evidence of this effect across the stimulus relations making it difficult to identify. Underwood (1949) explains the same idea in terms of sequence effects when the order of treatments in a study is counterbalanced across participants stating that “counterbalancing does not eliminate practice effects; counterbalancing only distributes these practice effects equally over all conditions when the effects are considered for all Ss [subjects] combined” (p.325). While this relates to the practice effects associated with different treatments or conditions, it could be argued that the same principle would apply when balancing the arrangement of individual stimuli in stimulus relations across participants. That is, balancing the arrangement of individual stimuli would not remove any differing effects of these stimuli on performance, and when performance is considered across all participants, any pattern in responding would be difficult to identify. Therefore, the procedure used in the present study allowed for the identification of stimuli that may have consistently affected performance across participants. As pointed out previously, no consistent patterns were seen. It is highly unlikely that using the same pairs of stimuli removed the effect (greater effectiveness of the SPO procedure) seen by Leader and Barnes-Holmes (2001b). If the different results between that, and the present study were due solely to the differences in the stimulus arrangement, then the superiority of the SPO procedure demonstrated by Leader and Barnes-Holmes (2001b) could not have been attributed to the procedures themselves.

Within-subject patterns in accuracy on the individual tested relations did emerge. The participants who reported that they could read the characters (Groups E1.Y.SPO.MTS and E1.Y.MTS.SPO) were more likely to achieve nine out of ten correct on the same individual tested relations across the sessions consistently. This consistency was not seen as frequently for the participants who reported that they could not read the characters (Groups E1.N.SPO.MTS and E1.N.MTS.SPO). In contrast, the tested relations on which these participants achieved greater numbers of correct responses tended to differ across the sessions. Additionally, in some cases, good performances on the tested relations following one training condition were accompanied by poor performance on the conflicting tested relations in the other condition. For example, if a participant responded correctly on nine of the ten trials for the B1-A1 tested relation following SPO training then, in some cases, they were likely to have performed very poorly on the B1-A3 tested relation following MTS training in the same session. The B1 stimulus was present as the sample stimulus in the tests for both relations, but a different response was correct in each test. Therefore, these relations could be deemed conflicting. Some evidence of this pattern was found in the groups who reported that they could not read the characters (Groups E1.N.SPO.MTS and E1.N.MTS.SPO) but this response pattern was particularly evident for two of the three participants in each of the groups that reported they could read the characters (Groups E1.Y.SPO.MTS and E1.Y.MTS.SPO).

The use of the same stimuli in conflicting relations is also reported to have affected the performance of the participants in Experiments 1, 2, and 3 by Leader and Barnes-Holmes (2001b). Leader and Barnes-Holmes (2001b) do not report the results for the individual tested relations. However, overall performance in testing following SPO training was better than following MTS training in three of the four experiments and they state that “upon inspection of the raw data it appears that students “adopted” respondent [SPO] training during the MTS test” (Leader & Barnes-Holmes, 2001b, p.442). Accuracy during MTS testing increased only upon the removal of the incorrect comparison stimuli from the MTS training condition (Experiment 4). However, it could be argued that the removal of these negative stimuli resulted simply in an SPO based training that required a response. This suggests a potential confound in their experiment. For participants to perform equally well on both procedures they would have to learn the stimulus arrangements as two separate equivalence classes. It is likely that the use of the same stimuli in both conditions made this more difficult, and in some cases led to very poor accuracy as the participants responded in accord with relations that were correct in the other procedure. While both studies provide evidence that the use of the same stimuli in both procedures affected performance, the procedure on which this effect was seen in Leader and Barnes-Holmes (2001b) and the present study differ. Leader and Barnes-Holmes (2001b) participants performed better following SPO training but performed poorly following MTS training. In the present study, the procedure that resulted in the best performance differed across participants. It is not clear why these findings differ.

In the present study, the evidence for the effect of conflicting relations is also supported by the finding that a greater number of correct responses than predicted by chance during testing following one training procedure was often accompanied by a greater number of errors than predicted by chance during testing following the other training condition for these participants. It appears that while the ability to read the characters may have made it easier for these participants to learn the relations it may have interfered with the learning of the conflicting relations following the other training procedure more than for those who reported that they could not read the stimuli.

To look at the effect of the conflicting relations further, chi square tests were conducted to assess the distribution of errors. On each trial the participants were presented with three possible response options. One of these options was the incorrect response for testing following the training procedure just completed but was the correct response when presented with the same sample stimulus following the other training procedure. Of the other two responses, one was the correct response and the third was an incorrect response following both training conditions. It could be that the use of the same stimuli in conflicting relations resulted in poor performance following one training condition compared to the other training condition. Evidence for this would be seen if significantly more errors were consistent with the correct response in the other condition than were incorrect in both conditions.

The results of these tests were inconclusive. There were significant differences between the types of errors in some sessions across participants. Where there were significant differences in the type of errors made, these nearly always showed significantly greater numbers of responses consistent with the correct response in the opposing condition. For a few participants this pattern occurred consistently across the sessions. This suggests that the use of the same characters in conflicting relations did interfere, for these participants, with deriving relations following the other training condition. This pattern was most evident for the participants who reported that they could read the characters. The finding that a very good performance on the individual tested relations following one training condition was often followed by poor performance on the conflicting relation following for the other tested condition supports this also. It appears that where the pattern of good performance mirrored by poor performance in the conflicting relation was evident, the poor performance was often due to the participant choosing the response that was correct during testing following the other training condition. In very few cases was the number of inconsistent errors significantly greater than the number of consistent errors. These findings suggest that the use of the same stimuli in conflicting relations interfered with learning other relations involving one of those stimuli.

Due to the overall poor performance of most participants during testing, performance during MTS training was assessed. It is not considered possible for participants to be able to form derived symmetry and equivalence relations during testing if they have not learnt the initial training relations (Sidman, 1994). For this reason, most studies that use MTS employ a training criterion. Participants must reach this criterion prior to beginning the test for derived relations. In these conditions the number of training trials or sessions completed by each participant can vary across participants. It is not possible to have a training criterion during SPO training as no response is required of the participant. A criterion was not used to terminate the MTS training in the present study as the number of training trials was kept the same across conditions, participants and sessions to allow a direct comparison with the SPO procedure. Had a training criterion similar to that used during testing (nine or more correct responses on each of the trained relations) been used, only four of the participants would have met the criterion by the end of their final experimental session. Two of these participants met the condition criterion in testing following MTS training. The remaining two participants were both in the groups who completed SPO training prior to MTS training. One reported that they could read the characters (Group E1.Y.SPO.MTS) and one reported that they could not (Group E1.N.SPO.MTS). The failure of some participants to meet the condition criterion after achieving an accuracy that would have met a criterion during MTS training is not uncommon. For example, Barnes-Holmes et al. (2000)

reported that nine of their 36 participants failed to meet the criterion during equivalence tests following MTS training. Similarly, Rehfeldt (2003) reported that only 7 of their 12 participants demonstrated equivalence following training to a criterion with an MTS procedure. Surprisingly, one of the participants in the present experiment met the condition criterion following MTS training despite having failed to respond correctly (at what would be considered a criterion) during training. This participant reported that they could not read the characters and completed MTS training prior to SPO training (Group E1.N.MTS.SPO). This finding contradicts the idea that criterion-level accuracy during training is necessary for the participant to respond correctly on tests of derived symmetry and equivalence relations. It is not possible to compare these findings to those of Leader and Barnes-Holmes (2001) for their participants who failed to meet the condition criterion following MTS training of any session as they did not present these data.

In summary, the results of the present study failed to replicate the findings of Leader and Barnes-Holmes (2001b) conclusively. It is possible that the Chinese characters in this study are responsible, in part, for this result. However, as the results also failed to agree with the findings of the study by Clayton and Hayes (2004), which did use the Chinese characters, it is unclear whether the findings were the result of the different stimuli. Perhaps the most likely factor responsible for the findings of the present study is the use of the same stimuli in both training conditions. This is supported by the instances of significantly greater numbers of consistent than inconsistent errors and also by the pattern of responding where participants performed well in one condition, but poorly in the other condition of a session. However, there is no clear explanation as to why this affected the participants in this study but did not affect the ability of the participants in the study by Leader and Barnes-Holmes's (2001b) study to form derived relations following both training conditions.

The effect of using the same stimuli in conflicting relations warrants further investigation. However, due to the inconsistent results between- and within-groups the first factor to be explored further was the one procedural difference between the current study and the study by Leader and Barnes-Holmes (2001b) that may have affected the findings. This is the use of the Chinese characters as stimuli. Therefore, the next experiment was a procedural replication of the present experiment with the Chinese characters replaced by the original nonsense syllables used by Leader and Barnes-Holmes (2001b).

EXPERIMENT 2

Experiment 1 failed to replicate the findings of Leader and Barnes-Holmes (2001b) and did not demonstrate a clear finding in favour of either procedure. Additionally, the participants performed poorly on tests for derived symmetry and equivalence relations overall. It was argued that one possible reason for the failure was the use of Chinese characters. Therefore, the present experiment was a procedural replication of Experiment 1 with the Chinese characters replaced with the nonsense syllables used in the original study by Leader and Barnes-Holmes (2001b).

Method

Participants, Ethics, Apparatus and Setting

Participant recruitment, ethics procedures, apparatus and setting were identical to those in Experiment 1. There were six participants, three in each of two groups. The participants were assigned to a group quasi-randomly in the order they were recruited. Table 2.1 shows the participant number, the group they were in, and the order in which the procedures were experienced in each session.

Table 2.1

Group assignment and condition order for all participants in Experiment 2.

Participant	Group	Condition order
P2.1	E2.SPO.MTS	SPO/MTS
P2.2	E2.SPO.MTS	SPO/MTS
P2.3	E2.SPO.MTS	SPO/MTS
P2.4	E2.MTS.SPO	MTS/SPO
P2.5	E2.MTS.SPO	MTS/SPO
P2.6	E2.MTS.SPO	MTS/SPO

Stimuli

The stimuli used were nine nonsense syllables and are presented in Table 2.2. As in Experiment 1, each stimulus was given an alphanumeric designation (i.e., A1, A2, A3, B1, B2, B3, C1, C2, and C3). The participants were never shown these alphabetic designations.

Table 2.2

Nonsense syllables and their alphanumeric designation

CUG	A1	ZID	A2	VEK	A3
YIM	B1	DAX	B2	PAF	B3
ROG	C1	MAU	C2	JOM	C3

Procedure

The procedure was a replication of that used in Experiment 1. The only difference was the use of nonsense syllables as stimuli. The participants in Group E2.SPO.MTS experienced SPO training prior to MTS training. These were reversed for the participants in Group E2.MTS.SPO. The instructions were modified slightly from those in Experiment 1. The word ‘character’ was replaced with ‘syllable’ in the instructions presented to the participants in line with the change in the stimuli.

Condition Criterion and Session Criterion

As in Experiment 1, participants were required to respond correctly on nine of the 10 trials for each tested relation in the MTS test following either training procedure to meet the condition criterion. The session criterion required that the participants meet the condition criterion on testing following both training conditions of an experimental session.

Results

Number of sessions

None of the participants met the condition criterion following either training condition of any session. Thus, all participants completed four experimental sessions.

Testing

Number correct on each tested relation and overall percent correct. Tables 2.3 and 2.4 show the number of correct responses for each tested relation; percent correct on the symmetry and equivalence relations; and total percent correct during testing following each training condition of each session for all participants in both groups of the present experiment. On both tables, numbers that appear in bold indicate correct responses for nine or more of the 10 trials for that tested relation.

As shown in Tables 2.3 and 2.4, none of the participants in either group met the condition criterion following either SPO or MTS training of any session. Overall, four of the participants achieved their greatest total percent corrects following the training procedure they experienced first in each session by the end of the final session. Figure 2.1 showed that this pattern was also evident in the percent correct on the symmetry and equivalence relations when assessed separately, for two participants in Group E2.SPO.MTS. These participants achieved a greater percentage of correct responses on the symmetry and equivalence relations following the training procedure they experienced first (SPO training). This pattern was not seen for the participants in Group E2.MTS.SPO, although one was performing better overall after MTS training in the last session.

A repeated-measures ANOVA was conducted to compare percent correct on the symmetry and equivalence relations of the final session for both procedures across the order of procedures (SPO/MTS and MTS/SPO). The results of the ANOVA are shown in Table 2.5. This table shows that there were no significant within-subject main effects of the procedure (SPO or MTS) nor was there a significant interaction between the procedure and the type of relation (symmetry or equivalence). For these findings, the effect sizes were less than the RMPE of 0.14 suggested by Ferguson (2009) as denoting an effect of practical significance. The other interaction effects were also not significant; however, the effect sizes of most comparisons were above 0.14. Moderate effect sizes (>0.25) were observed for the main effect of the type of relation and for the interactions between type of relation and procedural order, and the procedure and procedural order. There was no significant between-subjects effect of procedural order, and the effect size was less than the RMPE.

Most of the participants achieved nine or more correct responses on some individual relations, and these were more likely to be symmetry than equivalence relations. This pattern was strongest for the participants who experienced MTS training followed by SPO training.

Table 2.3

Number of correct responses for each tested relation of each training and testing cycle in each experimental session for the participants (in Group E2.SPO.MTS) who completed SPO followed by MTS training.

Participant	Session	Condition	Tested relations - sample and correct comparison stimuli											Overall % Correct	Condition criterion met	Session criterion met
			Symmetry relations						% Correct Symmetry	Equivalence relations			% Correct Equivalence			
			B1-A1 (SPO) B1-A3 (MTS)	B2-A2 (SPO) B2-A1 (MTS)	B3-A3 (SPO) B3-A2 (MTS)	C1-B1 (SPO) C1-B3 (MTS)	C2-B2 (SPO) C2-B1 (MTS)	C3-B3 (SPO) C3-B2 (MTS)		C1-A1 (SPO) C1-A2 (MTS)	C2-A2 (SPO) C2-A3 (MTS)	C2-A2 (SPO) C3-A1 (MTS)				
P2.1	Session 1	SPO	1	5	7	5	8	0	43.33	**9**	2	1	40.00	42.22	N	
		MTS	**10**	1	0	**10**	**9**	0	50.00	0	**9**	**10**	63.33	54.44	N	N
	Session 2	SPO	1	**9**	5	1	**10**	2	46.67	4	6	5	50.00	47.78	N	
		MTS	**9**	5	0	**10**	0	**10**	56.67	0	0	7	23.33	45.56	N	N
	Session 3	SPO	1	6	1	1	**10**	3	36.67	1	6	3	33.33	35.56	N	
		MTS	**10**	**10**	0	**10**	0	**10**	66.67	0	5	2	23.33	52.22	N	N
	Session 4	SPO	0	3	0	0	**10**	0	21.67	8	0	**10**	60.00	34.44	N	
		MTS	**10**	8	0	**10**	0	**10**	63.33	0	**10**	0	33.33	53.33	N	N
P2.2	Session 1	SPO	2	0	8	4	0	1	25.00	8	2	1	36.67	28.89	N	
		MTS	1	0	2	0	1	2	10.00	6	0	1	23.33	14.44	N	N
	Session 2	SPO	0	7	**10**	**9**	**10**	6	70.00	7	**10**	2	63.33	67.78	N	
		MTS	0	1	1	**9**	0	2	21.67	1	0	1	6.67	16.67	N	N
	Session 3	SPO	0	**10**	**10**	1	**10**	**10**	68.33	6	**10**	1	56.67	64.44	N	
		MTS	2	0	1	**9**	1	0	21.67	1	1	5	23.33	22.22	N	N
	Session 4	SPO	1	**10**	8	3	7	7	60.00	1	**10**	0	36.67	52.22	N	
		MTS	0	0	1	2	0	1	6.67	6	0	0	20.00	11.11	N	N
P2.2	Session 1	SPO	0	0	8	1	0	**9**	30.00	1	1	**9**	36.67	32.22	N	
		MTS	0	0	0	1	0	0	1.67	0	**10**	0	33.33	12.22	N	N
	Session 2	SPO	4	8	6	7	**9**	8	70.00	2	**9**	6	56.67	65.56	N	
		MTS	0	0	0	0	0	0	0.00	0	**10**	0	33.33	11.11	N	N
	Session 3	SPO	0	0	**10**	**10**	7	8	58.33	0	1	**10**	36.67	51.11	N	
		MTS	2	0	0	0	1	0	5.00	0	**10**	0	33.33	14.44	N	N
	Session 4	SPO	0	0	**10**	**9**	0	**9**	46.67	0	0	**10**	33.33	42.22	N	
		MTS	1	0	0	0	1	0	3.33	0	**10**	0	33.33	13.33	N	N

Table 2.4

Number of correct responses for each tested relation of each training and testing cycle in each experimental session for the participants (in Group 2.MTS.SPO) who completed MTS followed by SPO training.

Participant	Session	Condition	Tested relations - sample and correct comparison stimuli											Overall % Correct	Condition criterion met	Session criterion met
			Symmetry relations						% Correct Symmetry	Equivalence relations			% Correct Equivalence			
			B1-A1 (SPO) B1-A3 (MTS)	B2-A2 (SPO) B2-A1 (MTS)	B3-A3 (SPO) B3-A2 (MTS)	C1-B1 (SPO) C1-B3 (MTS)	C2-B2 (SPO) C2-B1 (MTS)	C3-B3 (SPO) C3-B2 (MTS)		C1-A1 (SPO) C1-A2 (MTS)	C2-A2 (SPO) C2-A3 (MTS)	C2-A2 (SPO) C3-A1 (MTS)				
P2.4	Session 1	SPO	2	7	7	6	**10**	6	63.33	0	5	3	26.67	51.11	N	
		MTS	8	4	8	3	5	3	51.67	4	7	1	40.00	47.78	N	N
	Session 2	SPO	1	**9**	**9**	**9**	**10**	**9**	78.33	6	2	6	46.67	67.78	N	
		MTS	**9**	4	8	6	**9**	2	63.33	6	3	5	46.67	57.78	N	N
	Session 3	SPO	0	0	0	0	1	0	1.67	5	0	0	16.67	6.67	N	
		MTS	**10**	**10**	**10**	**10**	**10**	**9**	98.33	5	7	3	50.00	82.22	N	N
	Session 4	SPO	0	0	1	0	0	0	1.67	0	0	0	0.00	1.11	N	
		MTS	**10**	**9**	**10**	**10**	**10**	**10**	98.33	1	0	1	6.67	67.78	N	N
P2.5	Session 1	SPO	3	4	6	3	2	3	35.00	4	1	3	26.67	32.22	N	
		MTS	4	4	4	3	4	5	40.00	5	4	3	40.00	40.00	N	N
	Session 2	SPO	3	4	7	2	0	2	30.00	3	2	6	36.67	32.22	N	
		MTS	3	1	3	1	4	6	30.00	3	4	2	30.00	30.00	N	N
	Session 3	SPO	4	3	2	4	2	2	28.33	4	6	2	40.00	32.22	N	
		MTS	2	4	1	4	2	4	28.33	3	2	2	23.33	26.67	N	N
	Session 4	SPO	0	4	2	6	5	3	33.33	0	4	5	30.00	32.22	N	
		MTS	4	3	2	4	3	2	30.00	2	4	1	23.33	27.78	N	N
P2.6	Session 1	SPO	7	0	1	3	**10**	**10**	51.67	**10**	1	0	36.67	46.67	N	
		MTS	**10**	5	0	4	6	2	45.00	0	8	5	43.33	44.44	N	N
	Session 2	SPO	8	**10**	7	0	**10**	**10**	75.00	7	0	2	30.00	60.00	N	
		MTS	**10**	**10**	4	**10**	0	0	56.67	0	0	6	20.00	44.44	N	N
	Session 3	SPO	**10**	8	2	**9**	**10**	**10**	81.67	**10**	1	0	36.67	66.67	N	
		MTS	0	**10**	4	**10**	5	4	55.00	1	0	5	20.00	43.33	N	N
	Session 4	SPO	2	0	8	8	**10**	**10**	63.33	**10**	0	0	33.33	53.33	N	
		MTS	9	**10**	2	**10**	**10**	2	71.67	1	0	**9**	33.33	58.89	N	N

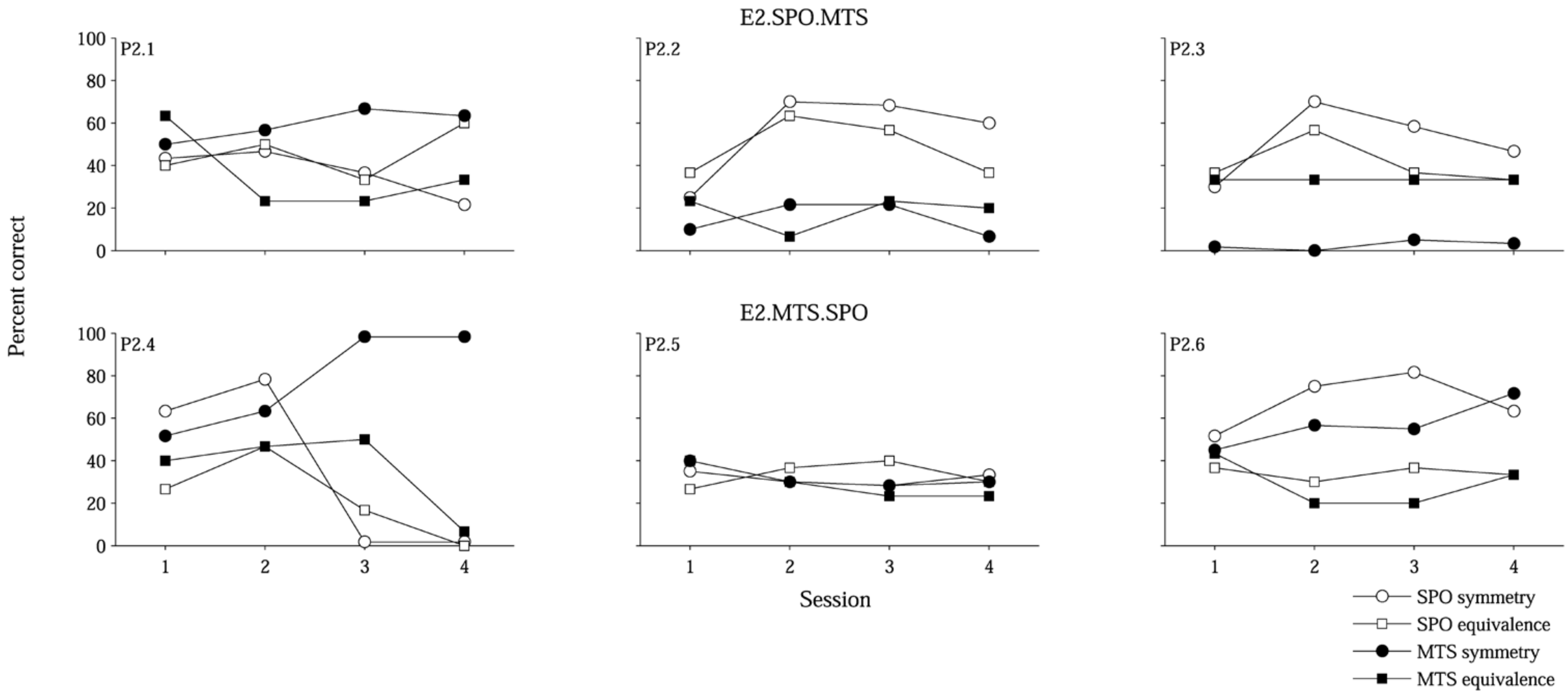

Figure 2.1 Percent correct on the symmetry and equivalence relations following SPO and MTS training across all sessions for all participants in Experiment 2.

Table 2.5

Results of an ANOVA to compare percent correct on the symmetry and equivalence relations of the final session for both procedures for both groups in

Source	*df*	*F*	$\eta^2_{partial}$
Within-subjects effects			
Procedure	1,4	0.001	0.000
Procedure x Procedural Order	1,4	2.386	0.374
Type of relation	1,4	4.062	0.504
Type of relation x Procedural Order	1,4	5.765	0.590
Procedure x Type of relation	1,4	0.429	0.097
Procedure x Type of relation x Procedural Order	1,4	0.681	0.145
Between-subjects effects			
Procedural order	1,4	0.004	0.001

*=significant at $p<0.05$

Correct vs. Incorrect (compared to chance) during testing. As in Experiment 1, χ^2 tests of goodness of fit were conducted to assess the performance of each participant compared to chance. Table 2.6 shows the number of correct and incorrect responses made following both training conditions in each experimental session by each participant in this experiment, and the obtained chi square statistic for each χ^2 test. As shown in Table 2.6, the number of correct responses achieved was significantly different from chance in most experimental sessions, for most participants. Two participants who completed SPO training prior to MTS training (P2.2 and P2.3 in Group E2.SPO.MTS) made more errors than predicted by chance following MTS training in all sessions. In most cases, these were paired with more correct responses than predicted by chance following SPO training of that session. The results for the other participant in that group (P2.1) showed significantly more correct responses than predicted by chance following MTS training in three sessions. For this participant, the distribution of responses following SPO training did not differ significantly from chance in most sessions. Two participants who completed MTS training followed by SPO training (P2.4 and P2.6 in Group E2.MTS.SPO) produced response distributions that were significantly different from chance following both training conditions of all sessions, and these indicated, with two exceptions, a greater number of correct responses than predicted by chance.

Consistent vs. inconsistent errors during testing. χ^2 tests for goodness of fit were calculated to assess the distribution of errors during testing following each training condition. Table 2.7 shows the number of consistent and inconsistent errors made during testing following both training conditions in each experimental session by all participants. Overall (as shown in Table 2.7), where there were significant differences in the number of consistent and inconsistent errors, these were nearly all indicative of a greater number of consistent than inconsistent errors. Two participants in each group (P2.1 and P2.2 in Group E2.SPO.MTS, and P2.4 and P2.6 in Group E2.MTS.SPO) made significantly more consistent than inconsistent errors in some sessions. The remaining participant in each group (P2.3 and P2.5) made similar numbers of consistent and inconsistent errors in all (P2.5), or nearly all (P2.3), sessions.

MTS training

Number correct on each trained relation and overall percent correct. Table 2.8 shows the number of correct responses for each trained relation and the total percent correct during MTS training for all participants. Numbers for each trained relation that appear in bold indicate that the participant responded correctly on nine or more trials for that relation. Where a total percent correct appears in bold, this participant responded correctly on nine or more trials for every trained relation in MTS training of that session. As shown in Table 2.8, two participants in each group (P2.1 and P2.3 in Group E2.SPO.MTS, and P2.4 and P2.6 in Group E2.MTS.SPO) responded correctly on nine or more trials for some trained relations during most sessions. Figure 2.2 shows the number of correct responses made on each trained relation across the four sessions. This figure shows that the number of correct responses on all relations increased across the sessions for those two participants in Group 2.MTS.PS (P2.4 and P2.6). One participant (P2.4) achieved nine or more correct responses on all trained relations in the final session. Two participants (P2.2 in Group E2.SPO.MTS, and P2.5 in Group E2.MTS.SPO) failed to respond correctly on nine or more trials of any trained relation following MTS training of any session. The data for these participants showed no clear trend in the percent of correct responses made across the sessions.

Table 2.6

Chi-square tests for the distribution of correct (Corr.) and incorrect (Incor.) responses (compared to responses expected by chance (correct (30/90), incorrect (60/90)) for all participants in Experiment 2.

Group	Participant	Session 1						Session 2						Session 3						Session 4					
		SPO			MTS			SPO			MTS			SPO			MTS			SPO			MTS		
		Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2
E2.SPO.MTS	P2.1	38	52	3.200	49	41	18.050*	43	47	8.450*	41	49	6.050*	32	58	0.200	47	43	14.450*	31	59	0.050	48	42	16.200
E2.SPO.MTS	P2.2	26	64	0.800	13	77	14.450*	61	29	48.050*	15	75	11.250*	58	32	39.200*	20	70	5.000*	47	43	14.450*	10	80	20.000*
E2.SPO.MTS	P2.3	29	61	0.050	11	79	18.050*	59	31	42.050*	10	80	20.000*	46	44	12.800*	13	77	14.450*	38	52	3.200	12	78	16.200*
E2.MTS.SPO	P2.4	46	44	12.800*	43	47	8.450*	61	29	48.050*	52	38	24.200*	6	84	28.800*	74	16	96.800*	1	89	42.050*	61	29	48.050*
E2.MTS.SPO	P2.5	29	61	0.050	36	54	1.800	29	61	0.050	27	63	0.450	29	61	0.050	24	66	1.800	29	61	0.050	25	65	1.250
E2.MTS.SPO	P2.6	42	48	7.200*	40	50	5.000*	54	36	28.800*	40	50	5.000*	60	30	45.000*	39	51	4.050*	48	42	16.200*	53	37	26.450*

*=significant at $p < .05$

Table 2.7

Chi-square tests on the distribution of errors (incorrect and consistent with the other condition (Con.), or incorrect and inconsistent with the other condition (Incon.)) for all participants in Experiment 2.

Group	Participant	Session 1						Session 2						Session 3						Session 4					
		SPO			MTS			SPO			MTS			SPO			MTS			SPO			MTS		
		Con.	Incon.	χ^2	Con.	Incon.	χ^2	Con.	Incon.	χ^2	Con.	Incon.	χ^2	Con.	Incon.	χ^2	Con.	Incon.	χ^2	Con.	Incon.	χ^2	Con.	Incon.	χ^2
E2.SPO.MTS	P2.1	22	30	1.231	30	11	8.805*	35	12	11.255*	35	14	9.000*	39	19	6.897*	29	14	5.233*	47	12	20.763*	24	18	0.857*
E2.SPO.MTS	P2.2	27	37	1.563	37	40	0.117	8	21	5.828*	52	23	11.213*	14	18	0.500	54	16	20.629*	22	21	0.023	57	23	14.450*
E2.SPO.MTS	P2.3	33	28	0.410	37	42	0.316	13	18	0.806	40	40	0.000	11	33	11.000*	39	38	0.013	19	33	3.769	40	38	0.051
E2.MTS.SPO	P2.4	28	16	3.273	28	19	1.723	21	8	5.828*	30	8	12.737*	79	5	65.190*	8	8	0.000	59	30	9.449*	0	29	29.000*
E2.MTS.SPO	P2.5	26	35	1.328	23	31	1.185	23	38	3.689	26	37	1.921	29	32	0.148	34	32	0.061	32	29	0.148	31	34	0.138
E2.MTS.SPO	P2.6	38	10	16.333*	19	31	2.880	25	11	5.444*	30	20	2.000	13	17	0.533	31	20	2.373	30	12	7.714*	26	11	6.081*

*=significant at $p < .05$

Table 2.8

Number of correct responses for each trained relation during MTS training of each experimental session for all participants in Experiment 2.

Group	Participant	Session	Sample and correct comparison stimuli						% Correct
			A1-B2	A2-B3	A3-B1	B1C2	B2-C3	B3-C1	
E2.SPO.MTS	P2.1	Session 1	2	1	**9**	8	2	8	50.00
		Session 2	6	1	**10**	5	4	7	55.00
		Session 3	8	7	**10**	8	**9**	**10**	86.67
		Session 4	**10**	6	**10**	**10**	**10**	6	86.67
E2.SPO.MTS	P2.2	Session 1	2	2	2	7	3	4	33.33
		Session 2	3	2	4	3	6	3	35.00
		Session 3	3	2	5	2	6	4	36.67
		Session 4	0	0	6	0	0	1	11.67
E2.SPO.MTS	P2.3	Session 1	**10**	7	**10**	6	**9**	**10**	86.67
		Session 2	**9**	1	**9**	5	**9**	**9**	70.00
		Session 3	**9**	8	**10**	7	8	**10**	86.67
		Session 4	**10**	8	**10**	**9**	**9**	**10**	93.33
E2.MTS.SPO	P2.4	Session 1	3	4	**9**	7	4	4	51.67
		Session 2	8	8	**10**	**10**	7	8	85.00
		Session 3	**10**	**10**	**10**	**10**	**10**	**9**	**98.33**
		Session 4	**10**	**9**	**10**	**10**	**9**	**10**	**96.67**
E2.MTS.SPO	P2.5	Session 1	3	4	6	4	6	4	45.00
		Session 2	2	1	5	2	5	4	31.67
		Session 3	4	1	6	3	6	2	36.67
		Session 4	3	3	4	6	4	5	41.67
E2.MTS.SPO	P2.6	Session 1	3	2	5	5	2	5	36.67
		Session 2	4	5	**10**	4	5	**10**	63.33
		Session 3	5	6	**10**	7	7	**10**	75.00
		Session 4	**10**	8	**9**	7	8	**10**	86.67

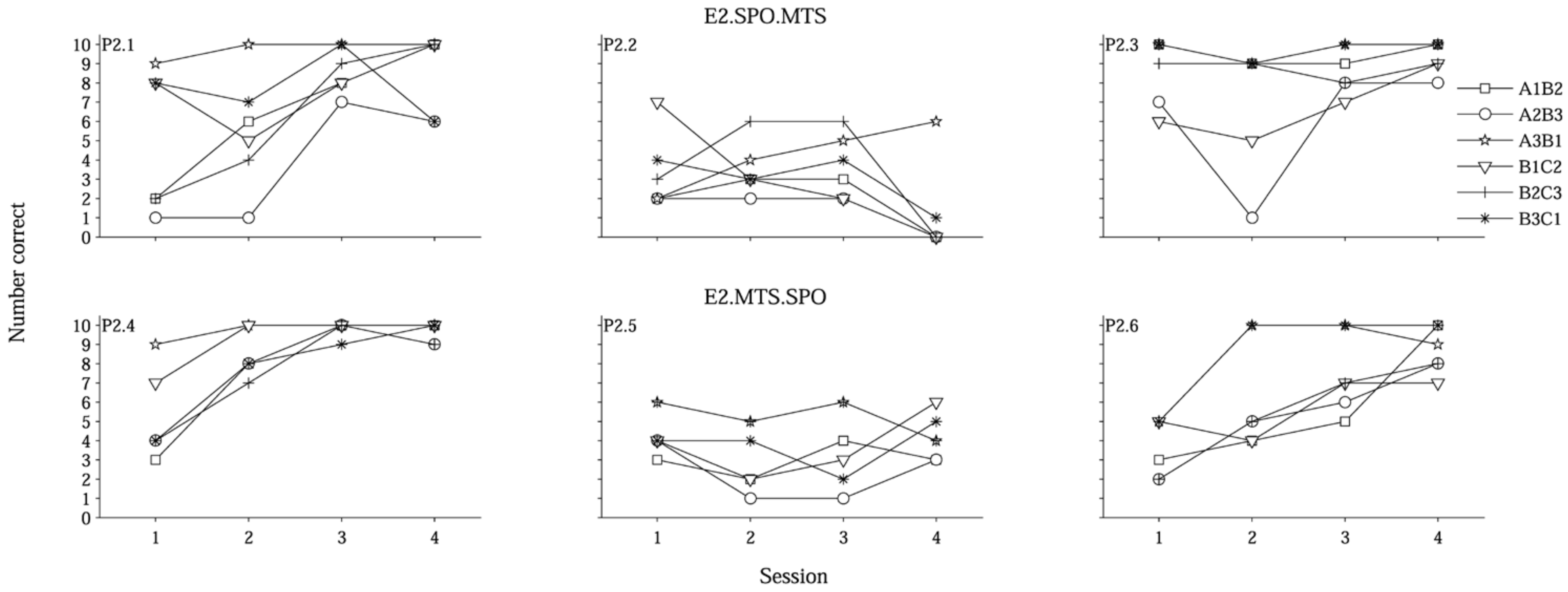

Figure 2.2. Number of correct responses made on each trained relation during MTS training across all sessions for all participants in Experiment 2.

Correct vs. incorrect (compared to chance) during MTS training. As in Experiment 1, χ^2 tests for goodness of fit were conduced to compare the distribution of correct and incorrect responses to the distribution that would be predicted by chance. Table 2.9 shows the number of correct and incorrect responses made during MTS training of each session by each participant in this experiment, and the χ^2 statistic for each test. Two participants in each group (P2.1 and P2.3 in Group E2.SPO.MTS, and P2.4 and P2.6 in Group E2.MTS.SPO) achieved a greater number of correct responses than predicted by chance during MTS training in all, or nearly all, sessions. One participant in each group (P2.2 in Group E2.SPO.MTS and P2.5 in Group E2.MTS.SPO) produced response distributions during MTS training that did not differ significantly from chance in most (P2.2) or all sessions (P2.5). One participant who completed SPO training prior to MTS training (P2.2) made a significantly greater number of incorrect than correct responses during their final session.

Consistent vs. inconsistent errors during MTS training. As in Experiment 1, $\chi 2$ tests with unequal expected frequencies were conducted to compare the number of errors where the participant chose the stimulus paired with that sample stimulus in the SPO condition, and those that were incorrect in both conditions. Table 2.10 shows the number of consistent and inconsistent errors made by each of the participants during MTS training of each session. The χ^2 statistic for each test is also shown. There were no significant differences between the number of consistent and inconsistent errors made by all but one participant during any session. Most of these participants made very few errors during MTS training of the final two sessions. The exception to this was P2.5, who completed MTS training followed by SPO training, and made a large number of both consistent and inconsistent errors in most sessions.

Table 2.9

Chi-square tests for the distribution of correct and incorrect responses during MTS training (compared to responses expected by chance (correct (20/60), incorrect (40/60)) by all participants in Experiment 2.

Group	Participant	Session 1			Session 2			Session 3			Session 4		
		Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2
E2.SPO.MTS	P2.1	30	30	7.500*	33	27	12.675*	52	8	76.800*	52	8	76.800*
E2.SPO.MTS	P2.2	20	40	0.000	21	39	0.075	22	38	0.300	7	53	12.675*
E2.SPO.MTS	P2.3	52	8	76.800*	42	18	36.300*	52	8	76.800*	56	4	97.200*
E2.MTS.SPO	P2.4	31	29	9.075*	51	9	72.075*	59	1	114.075*	58	2	108.300*
E2.MTS.SPO	P2.5	27	33	3.675	19	41	0.075	22	38	0.300	25	35	1.875
E2.MTS.SPO	P2.6	33	27	12.675*	38	22	24.300*	45	15	46.875*	52	8	76.800*

*=significant at p <.05

Table 2.10

Chi-square tests on the distribution of errors (incorrect and consistent with the other condition (Con.), or incorrect and inconsistent with the other condition (Incon.)) during MTS training for all participants in Experiment 2.

Group	Participant	Session 1			Session 2			Session 3			Session 4		
		Con.	Incon.	χ^2	Con.	Incon.	χ^2	Con.	Incon.	χ^2	Con.	Incon.	χ^2
E2.SPO.MTS	P2.1	18	12	1.200	15	12	0.333	5	3	0.500	8	0	8.000
E2.SPO.MTS	P2.2	15	25	2.500	17	22	0.641	22	16	0.947	51	2	45.302*
E2.SPO.MTS	P2.3	3	5	0.500	10	8	0.222	6	2	2.000	4	0	4.000
E2.MTS.SPO	P2.4	12	17	0.862	4	5	0.111	1	0	1.000	1	1	0.000
E2.MTS.SPO	P2.5	14	19	0.758	20	21	0.024	19	19	0.000	22	13	2.314
E2.MTS.SPO	P2.6	13	24	3.271	8	14	1.636	9	6	0.600	6	2	2.000

*=significant at p <.05

Discussion

This experiment was a procedural replication of Experiment 1 but with the nonsense syllables used by Leader and Barnes-Holmes (2001b). As in Experiment 1, the findings show no clear difference the effectiveness of the SPO and MTS procedures. Thus, as with the findings of Experiment 1, this finding differs from the findings of both Leader and Barnes-Holmes (2001b) (Experiment 1) and Clayton and Hayes (2004). Those studies reported conflicting findings in favour of the SPO and MTS procedures, respectively. The session criterion was not met during any session but most participants responded correctly on a greater percentage of responses than would be predicted by chance during some sessions. Therefore, both procedures could be viewed as partially effective. That is, while the participants failed to meet the condition criterion following training in each condition of each session, their performance was not random.

The findings of Experiment 1 showed better performance on the symmetry than on the equivalence relations during the final session regardless of the training procedure completed. The results of the present study do not show a significant difference in performance between the symmetry and equivalence relations, however, the effect size (0.504) is greater than that for this same effect in Experiment 1 (0.406) where a significant result was obtained. However, the smaller number of participants in the present experiment meant that the F statistic was not significant. The arguments for using effect size as a measure of the true magnitude of an effect, particularly with small samples, were outlined in Experiment 1. If this effect size is taken to indicate better performance on the symmetry than equivalence trials in the present experiment, then, as with Experiment 1, these findings do not agree with the findings of Clayton and Hayes (2004) who found no difference in performance on the symmetry and equivalence relations. A greater effect size than that found for the significant result (type of relation) in Experiment 1, was also found here for the interaction between the type of relation (symmetry or equivalence) and the procedural order. Specifically, this interaction shows that the participants who experienced SPO training prior to MTS training performed better on the symmetry than the equivalence relations, and the opposite was true for the participants who experienced MTS training prior to SPO training.

Overall, most of the participants who completed SPO training prior to MTS training (Group E2.SPO.MTS) achieved their greatest percentage of correct responses following SPO training. The results of those participants who experienced MTS training prior to SPO training (Group E2.MTS.SPO) were mixed. However, there was no significant effect of procedural order in the final session. Most of these participants performed best on testing following SPO training in early sessions, but this reversed in later sessions. These findings differ from those of Leader and Barnes-Holmes (2001b). Nearly all of their participants met the condition criterion following SPO training, and half of their participants also met the criterion following MTS training. The findings of the present study are more similar to those of Experiment 1 of this study in that none of the participants met the session criterion. However, three participants in Experiment 1 met the condition criterion following one training procedure, where none did in the present experiment.

Thus, despite the use of the nonsense syllables from the study by Leader and Barnes-Holmes (2001b), the findings of the present study were more similar to those of Experiment 1 where the stimuli were Chinese characters. As noted in the *Introduction*, Clayton and Hayes (2004) suggested that nonsense syllables are more nameable, and that this would make the formation of equivalence relations easier than with Chinese characters. The findings of the present study do not support this, as three of the participants in Experiment 1 (which used Chinese characters) met the condition criterion following one training condition and no participant in the present study met the criterion following either training condition. This makes it unlikely then that the failure of most participants to demonstrate derived equivalence relations in Experiment 1 was due to the use of the Chinese characters as stimuli.

There is limited evidence of an order effect in the present experiment. There was some evidence of an order effect in Experiment 1 for those participants who could read the Chinese characters (Groups 1Y.SPO.MTS and 1Y.MTS.SPO). In that experiment, the participants who could read the Chinese characters performed best following the training procedure that they experienced first in each session. A moderate effect size, but not a significant result, was shown for the effect of procedural order in the final session of Experiment 1. The conditions experienced by the participants in the present experiment were similar to those in Groups 1Y.SPO.MTS and 1Y.MTS.SPO of Experiment 1 as all could read the stimuli used. However, the procedures of the two experiments differ in that the Chinese characters were not arbitrary symbols for those participants who could read them. Therefore, it is possible that the differences observed were the result of the participation of the Chinese characters in equivalence relations within the context of the Chinese language. That is, previously established associations that involve the characters may have affected the findings for the participants in Experiment 1 who could read the Chinese characters.

As in Experiment 1, where participants in the present study made a greater number of errors than would be predicted by chance following one training condition, this was nearly always paired with a greater number of correct responses than predicted by chance following the other training condition in the session. This was evident for two of the participants who experienced SPO training prior to MTS training especially. Thus, it appears that a good performance in one condition precluded a good performance during the other condition of that session for some participants. This finding provides further support for the idea that the use of the same stimuli in conflicting relations was interfering with the formation of derived equivalence relations. However, this finding is not evident in all cases, and was not observed for the other three participants in this experiment.

At the level of the individual tested relations, five of the six participants responded correctly on nine or more trials for some tested relations in most sessions. Therefore, as with Experiment 1, an overall performance that did not differ significantly from chance did not necessarily indicate equal performance across all relations. Across the groups of the present experiment, most of the participants who experienced MTS first in each session responded correctly on nine or more trials on the symmetry relations. Relations on which nine or more correct responses were made included both the symmetry and equivalence relations for the participants who experienced SPO training prior to MTS training, however, the overall percentages of correct responses achieved by these participants were similar across the groups. Across the participants in Experiment 1 there was no consistent pattern in the relations on which nine or more correct responses were achieved. As in Experiment 1, no pattern involving particular stimuli were noted, and as mentioned earlier, this suggests that none of the relations, or relations involving particular stimuli were easier to learn than were others.

A within-subjects pattern of correct responses on nine or more trials for some tested relations following one condition paired with no, or few, correct responses in the conflicting relation of the other condition emerged for four participants. Additionally, for most participants, the errors made were more likely to be consistent with the correct response in the other condition than due to choosing the response option that was incorrect in both conditions. Thus, it appears that even when the number of errors was not significantly different from that predicted by chance, the distribution of these errors still indicated that the use of the same stimuli in conflicting relations was impacting on the ability of these participants to derive the correct equivalence relations during testing.

The performance of all participants during MTS training was investigated. As in Experiment 1, there was no criterion on performance during MTS testing as the number of trials completed was held constant with the number of SPO trials (on which no criterion was possible). Had a criterion of nine or more correct responses on each tested relation been present, only one participant in the present experiment would have met this criterion. In contrast, four of the 12 participants in Experiment 1 would have met such a criterion. Therefore, overall performance on MTS training by the participants in Experiment 2 was worse than that achieved by the participants in Experiment 1. The results of the MTS training in these two studies cannot be compared to those of Leader and Barnes-Holmes (2001b) who do not report the results for their participants during MTS training. The introduction of the nonsense syllables in the present study failed to result in a better performance during both MTS training and testing following both training procedures. This makes it unlikely that the differences observed between the findings of Leader and Barnes-Holmes (2001b), Clayton and Hayes (2004), and either of Experiments 1 and 2 of the present study had anything to do with the different stimuli used.

It is possible that the poor performance during MTS training by some participants was due to the use of the same stimuli in conflicting relations across the training procedures. While it is not possible to look at performance during training of the SPO condition, the better performance achieved during testing following SPO training was paired with poor performance during both training and testing of the MTS condition for two participants. Both of these participants experienced SPO training prior to MTS training. This pattern was reversed for one participant who experienced MTS first in a session. This participant had a good performance after MTS training, but then had a poor following SPO training (in the final two sessions). This participant also achieved a criterion-level performance during MTS training. Therefore, a pattern of good performance following one training condition, paired with poor performance following the other training condition was observed for half of the participants in the present experiment. This, along with the results of Experiment 1, provides some support the idea that the use of the same stimuli in conflicting relations was likely to have affected performance.

Four participants increased their correct responding across the sessions of MTS training (Figure 2.2). Similar increases were seen during MTS training for 10 of the 12 participants in Experiment 1, with four of those participants achieving a criterion-level performance by the final session. Therefore, while not all of these participants achieved a criterion-level performance by the end of training, they were learning some relations, and their performance was improving with repeated exposures to training and testing. This suggests that exposure to a greater number of training trials, or to a greater number of training and testing cycles was contributing to improved performance. Thus, had more training trials been conducted, it is possible that more of these participants may have reached a criterion-level performance, at least during MTS training.

In summary, the findings of the present experiment are similar, in some respects, to those of Experiment 1, and the findings of both these experiments differ from those of Leader and Barnes-Holmes (2001b). Thus, it seems unlikely that the use of the Chinese characters in Experiment 1 had much of an effect on the findings of that experiment. However, the use of non-nonsense stimuli in Experiment 1 may have produced the order effects observed for those participants who could read the Chinese characters.

Additionally, the finding for some participants in both experiments, that a good performance in one condition was often paired with a poor performance in the other condition of a session suggests that the use of the same stimuli in both conditions may have interfered with the development of the equivalence and symmetry relations. This possibility is supported further by the significantly greater proportion of the errors that were consistent with the correct response for the opposing condition than incorrect in both conditions that was seen for these participants. Specifically, the use of the same stimuli in one condition may interfere with the facilitation of associations using those same stimuli in the other condition. Performance during MTS training may have improved with further exposure to the training and testing cycles. It is possible then that performance during testing may improve with repeated exposure also, although it is not clear that this would be expected from the testing results alone. In conclusion, the most likely factor to have affected the findings of the present study was the use of the same stimuli in both training conditions.

EXPERIMENT 3

The findings of Experiments 1 and 2 suggest that the use of the same stimuli in conflicting relations may have contributed to the poor performance seen for most participants. Other studies have also suggested that the use of conflicting relations can interfere with the formation of new equivalence relations.

Both of the studies that have compared the effectiveness of the SPO and MTS procedures used the same stimuli across different relations in the different procedures. Leader and Barnes-Holmes (2001) used the same sample and comparison stimuli for each relation, however, the comparison stimulus that was designated as the correct response differed between the procedures. Therefore, the relations trained in one procedure conflicted with those trained under the other. In Clayton and Hayes (2004) the stimuli that were presented as sample stimuli were the same under both procedures; however, different sets of comparison stimuli were used. It is not clear what the outcomes would have been had conflicting relations not been involved. There is some research examining the effect of conflicting relations in equivalence tasks. These studies have all examined the reversal of equivalence relations (e.g., Michael & Bernstein, 1991; Pilgrim, Chambers & Galizio, 1995; Pilgrim & Galizio, 1990; 1995; Saunders, Drake & Spradlin, 1999; Saunders, Saunders, Kirby and Spradlin, 1988; Smeets, Barnes-Holmes, Akpinar & Barnes-Holmes, 2003; Spradlin, Cotter & Baxley, 1973, cited in Spradlin, Saunders and Saunders, 1992; Spradlin, Saunders & Saunders, 1992; Wirth & Chase, 2002). In reversal studies, participants receive training, and then complete tests for equivalence relations. Thus, once the equivalence classes have been established, the same stimuli are rearranged into new relations, and training (and subsequent testing) of these relations begins. If the participants respond correctly on equivalence tests for the new sets of relations, then the equivalence relations are said to have been reversed.

An early study by Spradlin et al. (1973, cited in Spradlin et al., 1992) demonstrated that equivalence classes involving different arrangements of the same stimuli could be achieved. Following training and testing of two equivalence classes, the baseline relations were altered so that the classes became mixed. The participants in this study then responded in accordance with the new relations during testing (Spradlin et al., 1973, cited in Spradlin et al., 1992). Spradlin et al. (1992) also report the reversal of equivalence relations by two child participants; however, they report that one of these participants required a large number of training trials to learn the new baseline relations. Saunders et al. (1988) demonstrated interference with the reversal of previously established equivalence relations. In Experiment 2 by Saunders et al. (1988), participants who had demonstrated the formation of equivalence relations received training in which the correct response was reversed (the correct response became choosing the other comparison stimulus that was available). However, on equivalence test probes, most responses were consistent with the original equivalence relations. Pilgrim and Galizio (1990) reported that during testing following a reversal procedure, most participants responded correctly on the new symmetry relations. However, responses on the transitivity relations were consistent with the correct response for the original relations (Pilgrim & Galizio, 1990). This suggests that the originally developed equivalence relations interfered with performance on the reversal task. Pilgrim and Galizio (1995) extended this study to a second adult sample. This study reported similar findings to those of Pilgrim and Galizio (1990) where following the reversal training, most participants responded correctly on the symmetry trials, but responses on the equivalence trials were consistent with the original relations. Another study by this research group (Pilgrim et al., 1995) examined the reversal of equivalence relations by child participants. That study reported that trained relations also interfered with the development of equivalence relations by the children. However, the reversal resulted in an unsystematic pattern of responses on tests for the derived relations (Pilgrim et al., 1995). Michael and Bernstein (1991) also reported an unsystematic pattern of responses when reversing some of the trained relations in equivalence tests with children. In another study with children, Saunders, Drake and Spradlin (1999) reported the reversal of equivalence relations by most of their participants. Wirth and Chase (2001) found some evidence of reversal in equivalence tests in their study with adults who demonstrated a greater proportion of responses that were consistent with the reversed contingencies than were consistent with the original relations compared to a set of relations that were not subject to reversal training. Different findings were reported by Smeets et al. (2003). In that study, most participants performed well on a new equivalence test following reversal training, even when Smeets et al. (2003) modified their procedure so that it would match more closely the procedures of Pilgrim and Galizio (1990) and Pilgrim et al. (1995).

Equivalence relations involving different arrangements of the same stimuli have also been used in studies that have examined the role of context in equivalence class formation (e.g., Bush, Sidman & de Rose, 1989; Randell & Remington, 2006). Typically, these studies have used a reversal procedure but the different arrangements have been signalled by different contextual cues. For example, Bush et al. (1990) demonstrated the successful formation of equivalence relations involving the same stimuli but presented in different contexts (signalled by a high or low pitched tone) by adult participants. It could be argued that, in the present study, the differing presentation of the training procedures (SPO or MTS) could be contextual cues. This suggests that the reversal task would be easier in this study than in previous studies where the different conditions appeared identical. The failure of the participants in the present study to form the equivalence relations suggests that this was not the case. Contextual cues in the formation of equivalence relations were also studied by Randell and Remington (2006) who used reversed relations involving pictorial stimuli with rhyming or non-rhyming names in a computer task. In that study, the changing relations were signalled by a change in colour of the background on the monitor (blue or red). Randell and Remington (2006) reported that the participants learnt the relations when the names of the stimuli in a class rhymed, but not when the classes were rearranged so that the names of the stimuli within a class did not rhyme. However, it could be argued that the rhyming nature of the words was a pre-existing relation that interfered with the development of equivalence relations. This finding would then be similar to those of Eikeseth and Baer (1997), Ybarra Sagarduy et al. (2002) and Stewart et al. (2002) as outlined in the *Discussion* of Experiment 1.

All of the studies outlined so far have studied the reversal of equivalence relations using MTS procedures. Three studies (Layng & Chase, 2001; Leader and Barnes-Holmes, 2001; Smeets, Barnes-Holmes & Striefel, 2006) have demonstrated the reversal of equivalence relations using procedures other than MTS.

Most of Layng and Chase’s (2001) participants demonstrated the reversal of equivalence relations using an SPO procedure. Leader and Barnes-Holmes (2001) demonstrated the reversal of equivalence relations by most participants using both MTS and SPO procedures when they removed negative comparisons from the MTS task (so that only the correct comparison could be selected on each trial). The participants in a study by Smeets et al. (2006) initially failed to demonstrate reversal with the pREP procedure. However, the successful reversal of equivalence relations using a pREP procedure was shown when it was combined with instructions to respond on half of the trials and the procedure involved training and testing first the simple (symmetry) and then more complex (transitivity and equivalence) relations procedure was used.

In summary, most of the studies that have examined the reversal of equivalence relations have demonstrated disruption to the original equivalence classes, but failure of the participants to respond correctly on reversed relations. Therefore, while the effect produced by conflicting relations is unclear, there is evidence to suggest that previously developed relations can interfere with the development of new relations involving the same stimuli. As such, it seems likely that the use of the same stimuli in both the SPO and MTS procedures contributed to the poor performance of most participants in Experiments 1 and 2 of the present study. Therefore, Experiment 3 was a replication of Experiment 2, but with different stimuli used in each procedure. The stimuli used in the SPO procedure were the same nonsense syllables as those used in Experiment 2, and by Leader and Barnes-Holmes (2001). The stimuli used in the MTS procedure were nine different, three-letter, nonsense syllables. In all other respects the procedures were identical to those of Experiment 2. To allow comparisons with the two previous experiments and Leader and Barnes-Holmes (2001), there were no other procedural changes.

Method

Participants, Ethics, Apparatus and Setting

Participant recruitment, ethics procedures, apparatus and setting were identical to those in Experiments 1 and 2. Additionally, an ethics application to allow the inclusion of participants who were not enrolled in either of the first year psychology courses was submitted, and approved. Participants who were not eligible for course credit received book or MTA (petrol) vouchers, at $5 per session, to a total of $20 for their participation. The present study involved 6 participants in total, 3 in each of 2 groups that differed only in the order of the conditions completed by the participants (3.SPO.MTS, 3.MTS.SPO). The participants were assigned to an experimental group quasi-randomly in the order in which they were recruited.

Stimuli

The stimuli used in SPO training and testing following SPO training were the nine nonsense syllables used by Leader and Barnes-Holmes (2001), and in Experiment 2 of the present study. Nine different nonsense syllables were used in MTS training, and testing following MTS training. These nonsense syllables were checked against definitions from the Oxford English Dictionary (www.oed.com) to ensure that they were not common words. These stimuli are presented in Table 3.1. As in Experiments 1 and 2, each of the syllables was given an alphanumeric designation that was not shown to the participants.

Table 3.1

Nonsense syllables used in MTS training and their alphanumeric designation.

BUQ	A1	NIF	A2	GOE	A3
LAJ	B1	WOB	B2	SUL	B3
TIW	C1	HAC	C2	KAP	C3

Procedure

The procedure was a replication of Experiment 2, with one alteration. This was that different nonsense syllables were used during MTS training (and subsequent testing), as outlined above. The instructions and condition order for the participants in the two groups of the present study were identical to those in Experiment 2.

Condition and Session Criteria

The condition and session criteria were the same as in Experiments 1 and 2.

Results

Experimental Sessions

The session criterion was met during the third experimental session by one participant who completed MTS training prior to SPO training (P3.5). The remaining participants in Group 3.MTS.SPO, and all of the participants in Group 3.SPO.MTS failed to meet the session criterion during any experimental session, thus, completing the maximum of four experimental sessions.

Testing

Number correct on each tested relation and overall percent correct. Tables 3.2 and 3.4 show the number of correct responses (with a maximum of 10) for each trial type, and the total percentage of responses on symmetry and equivalence trials that were correct during each condition of each experimental session, for each of the participants in Groups

3.SPO.MTS (P3.1-P3.3) and 3.MTS.SPO (P3.4-P3.6) of Experiment 3. On these tables, numbers that appear in bold indicate that nine or more of the 10 trials on this tested relation resulted in a correct response.

Overall, the participants in who completed SPO testing prior to MTS testing (Group 3.SPO.MTS) achieved greater percentages of correct responses following MTS than SPO training (see Table 3.2). This pattern reversed during the final session for one participant (P3.3), and one other participant (P3.1) responded correctly on a greater percentage of equivalence trials following SPO than MTS training when the symmetry and equivalence relations were assessed separately. All of the participants responded correctly on nine or more trials of some tested relations, and these were primarily symmetry relations for two of the three participants in this group (P3.1 and P3.3). No clear pattern emerged for P3.2 who responded correctly on nine or more trials for few tested relations.

As shown in Table 3.3, the results for the participants in Group 3.MTS.SPO were mixed. One participant (P3.4) responded correctly on a greater percentage of responses following MTS than SPO training and this pattern was reversed for P3.6. The remaining participant in this group met the session criterion during their third experimental session, responding equally well following both training conditions. Two participants (P3.1 and P3.4) responded correctly on a greater percentage of responses following MTS than SPO training overall, but achieved a greater percentage correct on the equivalence relations following SPO than MTS training when the symmetry and equivalence relations were assessed separately.

Figure 3.1 shows percent correct on the symmetry and equivalence relations following each training condition across the sessions for all participants in this experiment. There was no clear pattern in percent correct on the symmetry and equivalence relations across the groups. Two participants (P3.4 and P3.6) who completed MTS training followed by SPO training achieved greater percentages of correct responses on the symmetry than the equivalence relations in most sessions. The percentage of correct responses achieved by one participant (P3.5 in Group 3.MTS.SPO) increased across the session for all relations. This participant met the session criterion in their third session. One participant (P3.2) who experienced SPO training prior to MTS training achieved slightly greater percent corrects following MTS training than SPO training in most sessions. No clear pattern was evident for the remaining participants in Group 3.SPO.MTS.

Table 3.4 shows the results of a factorial repeated-measures ANOVA that was calculated to compared the percent of correct responses achieved on the symmetry and

Table 3.2

Number of correct responses for each tested relation of each traning and testing cycle in each experimental session for the participants (in Group E3.SPO.MTS) who completed SPO followed by MTS training.

Participant	Session	Condition	Tested relations - sample and correct comparison stimuli											Overall % Correct	Condition criterion met	Session criterion met
			Symmetry relations						% Correct Symmetry	Equivalence relations			% Correct Equivalence			
			B1-A1	B2-A2	B3-A3	C1-B1	C2-B2	C3-B3		C1-A1	C2-A2	C3-A3				
P3.1	Session 1	SPO	6	4	4	2	6	4	43.33	3	2	2	23.33	36.67	N	
		MTS	3	7	5	3	1	1	33.33	1	3	6	33.33	33.33	N	N
	Session 2	SPO	3	1	6	4	2	2	30.00	3	5	3	36.67	32.22	N	
		MTS	0	8	**10**	**10**	3	1	53.33	0	2	0	6.67	37.78	N	N
	Session 3	SPO	1	1	6	1	9	7	41.67	8	1	4	43.33	42.22	N	
		MTS	1	**10**	**10**	**9**	**10**	**9**	81.67	0	0	1	3.33	55.56	N	N
	Session 4	SPO	0	0	8	0	8	**10**	43.33	**9**	0	0	30.00	38.89	N	
		MTS	2	**10**	**10**	**10**	**9**	**10**	85.00	1	0	1	6.67	58.89	N	N
P3.2	Session 1	SPO	1	4	3	6	4	5	38.33	4	3	6	43.33	40.00	N	
		MTS	5	3	3	6	6	5	46.67	6	3	5	46.67	46.67	N	N
	Session 2	SPO	2	3	2	3	5	4	31.67	4	3	3	33.33	32.22	N	
		MTS	**9**	1	5	7	7	4	55.00	7	7	6	66.67	58.89	N	N
	Session 3	SPO	2	4	4	5	1	3	31.67	2	4	2	26.67	30.00	N	
		MTS	6	2	4	4	5	4	41.67	3	2	3	26.67	36.67	N	N
	Session 4	SPO	1	3	2	5	4	1	26.67	1	5	6	40.00	31.11	N	
		MTS	**9**	6	5	**10**	6	5	68.33	**9**	5	5	63.33	66.67	N	N
P3.2	Session 1	SPO	2	3	4	3	5	1	30.00	0	2	1	10.00	23.33	N	
		MTS	5	1	**9**	7	3	4	48.33	0	**9**	0	30.00	42.22	N	N
	Session 2	SPO	1	2	**10**	2	7	0	36.67	1	1	8	33.33	35.56	N	
		MTS	8	5	**9**	**9**	7	8	76.67	6	5	7	60.00	71.11	N	N
	Session 3	SPO	1	1	**9**	1	**9**	0	35.00	0	0	9	30.00	33.33	N	
		MTS	4	6	3	5	5	4	45.00	3	2	3	26.67	38.89	N	N
	Session 4	SPO	**10**	8	**10**	**10**	**10**	**10**	96.67	2	3	7	40.00	77.78	N	
		MTS	3	2	4	5	3	5	36.67	3	2	4	30.00	34.44	N	N

Table 3.3

Number of correct responses for each tested relation of each training and testing cycle in each experimental session for the participants (in Group E3.MTS.SPO) who completed MTS followed by SPO training.

Participant	Session	Condition	Tested relations - sample and correct comparison stimuli: Symmetry relations						% Correct Symmetry	Equivalence relations			% Correct Equivalence	Overall % Correct	Condition criterion met	Session criterion met
			B1-A1	B2-A2	B3-A3	C1-B1	C2-B2	C3-B3		C1-A1	C2-A2	C3-A3				
P3.4	Session 1	SPO	**9**	8	7	**9**	8	**10**	85.00	1	4	7	40.00	70.00	N	
		MTS	1	2	**10**	6	**9**	5	55.00	3	0	0	10.00	40.00	N	N
	Session 2	SPO	7	5	0	8	3	**10**	55.00	1	0	8	30.00	46.67	N	
		MTS	2	1	10	10	10	10	71.67	0	0	0	0.00	47.78	N	N
	Session 3	SPO	7	**10**	0	4	8	**10**	65.00	1	0	**10**	36.67	55.56	N	
		MTS	8	**9**	**10**	**9**	**10**	**10**	93.33	0	0	**10**	33.33	73.33	N	N
	Session 4	SPO	8	**10**	0	2	4	**10**	56.67	**9**	0	**10**	63.33	58.89	N	
		MTS	**10**	**10**	**10**	**9**	**10**	**9**	**96.67**	0	0	**9**	30.00	74.44	N	N
P3.5	Session 1	SPO	0	5	0	1	6	8	33.33	0	2	0	6.67	24.44	N	
		MTS	**10**	7	7	**10**	8	**9**	85.00	**10**	5	5	66.67	78.89	N	N
	Session 2	SPO	**10**	**10**	**10**	**10**	**9**	**10**	**98.33**	**10**	**10**	**10**	**100.00**	**98.89**	**Y**	
		MTS	**10**	**10**	**10**	**10**	**10**	**10**	**100.00**	**10**	**10**	8	93.33	97.78	N	N
	Session 3	SPO	**10**	**10**	**10**	**10**	**10**	**10**	**100.00**	**10**	**10**	**10**	**100.00**	**100.00**	**Y**	
		MTS	**10**	**10**	**10**	**10**	**10**	**10**	**100.00**	**10**	**10**	**9**	**96.67**	**98.89**	**Y**	**Y**
	Session 4	SPO	–	–	–	–	–	–	–	–	–	–	–	–	–	
		MTS	–	–	–	–	–	–	–	–	–	–	–	–	–	–
P3.6	Session 1	SPO	5	2	2	4	4	4	35.00	3	4	3	33.33	34.44	N	
		MTS	2	3	0	4	1	1	18.33	3	4	3	33.33	23.33	N	N
	Session 2	SPO	3	5	**10**	**9**	**10**	**10**	78.33	5	5	**10**	66.67	74.44	N	
		MTS	6	7	**9**	**9**	**10**	**10**	85.00	5	2	1	26.67	65.56	N	N
	Session 3	SPO	7	**10**	**10**	**9**	**10**	**10**	93.33	3	7	**10**	66.67	84.44	N	
		MTS	**10**	8	**9**	**10**	**9**	**10**	93.33	0	2	4	20.00	68.89	N	N
	Session 4	SPO	8	**9**	**10**	**10**	**10**	**9**	93.33	6	3	**10**	63.33	83.33	N	
		MTS	8	1	**9**	**10**	**10**	**10**	80.00	2	0	7	30.00	63.33	N	N

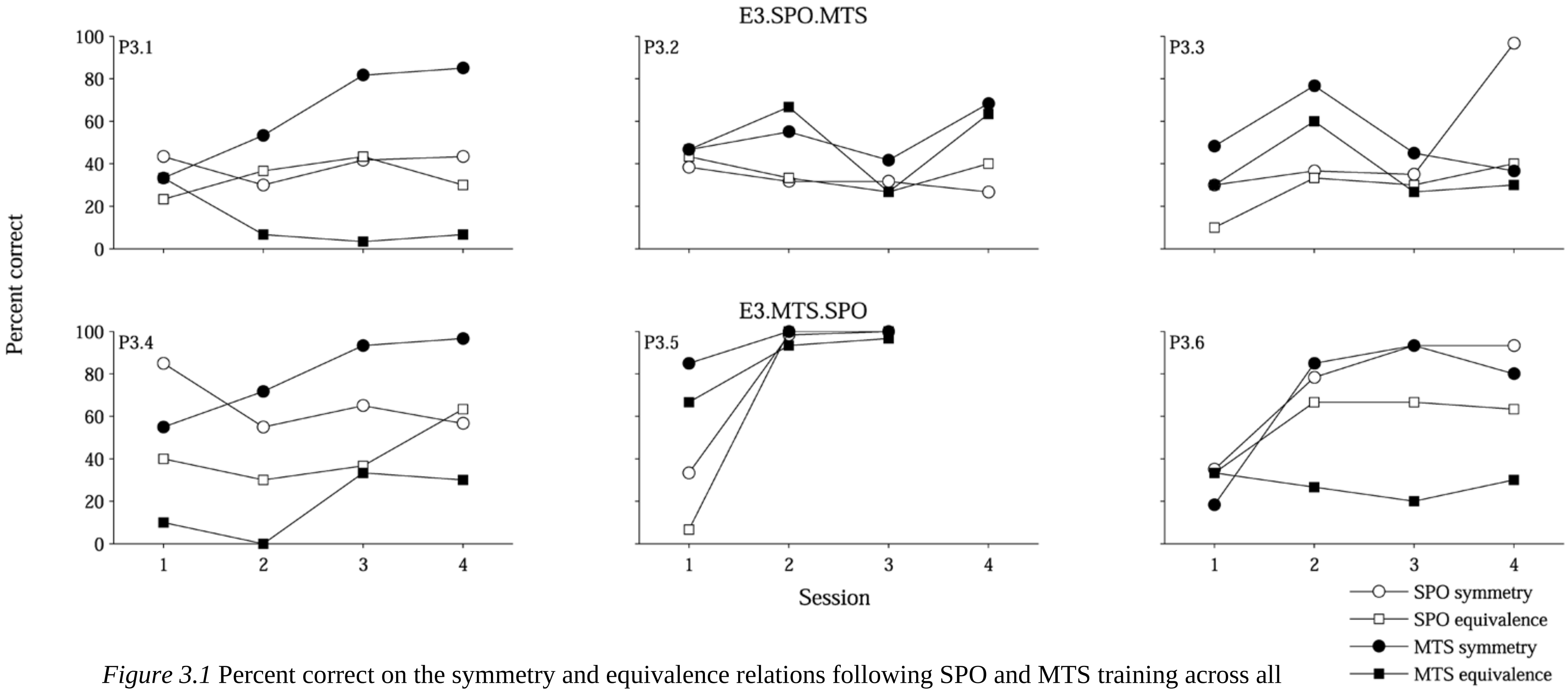

Figure 3.1 Percent correct on the symmetry and equivalence relations following SPO and MTS training across all sessions for all participants in Experiment 3.

equivalence relations in the final session by each participant. As shown in Table 3.5, none of the within-subjects effects were statistically significant. Nor was there a statistically significant effect of procedural order between the groups. Moderate effect sizes were seen for type of relation (symmetry or equivalence) (0.623) and procedural order (0.582). The effect size for the within-subjects interaction between procedure (SPO or MTS) and type of relation (0.231) was greater than the RMPE (0.14) suggested by Ferguson (2009). The effect sizes of the remaining results were below Fergusson's (2009) RMPE.

Table 3.4

Results of an ANOVA to compare percent correct on the symmetry and equivalence relations of the final session for both procedures for both groups in Experiment 3.

Source	*df*	*F*	$\eta^2_{partial}$
Within-subjects effects			
Procedure	1,4	0.054	0.013
Procedure x Procedural Order	1,4	0.194	0.046
Type of relation	1,4	6.615	0.623
Type of relation x Procedural Order	1,4	0.001	0.000
Procedure x Type of relation	1,4	1.203	0.231
Procedure x Type of relation x Procedural Order	1,4	0.286	0.067
Between-subjects effects			
Procedural order	1,4	5.559	0.582

*=significant at $p<0.05$

Correct vs. Incorrect (compared to chance) during testing. The session criterion was not met following training of either SPO or MTS training by five of the six participants in Experiment 3. Therefore, (as with previous experiments) χ^2 tests of goodness of fit were conducted to assess the performance of each participant compared to chance. The expected distribution of responses was the same as expected in Experiments 1 and 2. As with the previous experiments, terms regarding 'significance' denote statistical significance.

Table 3.5 shows the number of correct and incorrect responses for each experimental session for each participant, and the χ^2statistic for each test. Test statistics that are followed by an asterisk (*) are significant at $p<0.05$.

As shown in Table 3.5, the distribution of correct and incorrect errors were significantly different from chance during testing following less than half of all training sessions for the participants in Group 3.SPO.MTS. Where the results for this group did differ significantly from that predicted by chance these were nearly always a result of a greater number of correct responses than predicted by chance following MTS training. For Group 3.MTS.SPO, the number of correct responses was significantly greater than predicted by chance for all participants in most sessions. The participants in Group 3.MTS.SPO were achieved more correct responses than predicted by chance than the participants in Group 3.SPO.MTS.

MTS Training

Number correct on each trained relation and overall percent correct. Table 3.6 shows the number of correct responses for each trained relation, and the overall percentage of correct responses achieved during MTS training for each of the participants. As with Experiments 1 and 2, there was no criterion on performance during training. Figure 3.2 shows the number of trials on which correct responses were made for each of the trained relations in MTS training.

Table 3.6 and Figure 3.2 show that the number of MTS training relations on which nine or more trials were correct increased across the sessions for five of the six participants. Four of these participants responded correctly on nine or more trials for all relations during their final session. No clear pattern was evident across the sessions for one participant (P3.2 in Group SPO.MTS).

Correct vs. Incorrect (compared to chance). χ^2 tests were conducted to assess the distribution of correct and incorrect responses made during MTS training. As with Experiments 1 and 2, the expected distribution at chance would not have differed significantly from 20/60 correct responses and 40/60 incorrect responses. Table 3.7 shows the number of correct and incorrect responses made during MTS training in each session by each participant in Experiment 3, and the χ^2 statistic for each test. A greater number of correct responses than predicted by chance were achieved by three participants (P3.2 and

P3.3 in Group 3.SPO.MTS, and P3.5 in Group 3.5). The remaining three participants (P3.1, P3.4, and P3.6) responded correctly on a greater number of responses than predicted by chance during all but their first session, where their distribution of responses did not differ from chance significantly.

Table 3.5

Chi-square tests for the distribution of correct (Corr.) and incorrect (Incor.) responses (compared to responses expected by chance (correct (30/90), incorrect (60/90)) for all participants in Experiment 3.

Group	Participant	Session 1						Session 2						Session 3						Session 4					
		SPO			MTS			SPO			MTS			SPO			MTS			SPO			MTS		
		Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2
E3.SPO.MTS	P3.1	33	57	0.450	30	60	1.000	29	61	0.050	34	56	0.800	38	52	3.200	50	40	20.000*	35	55	1.254	53	37	26.450*
E3.SPO.MTS	P3.2	36	54	1.800	42	48	7.200*	29	61	0.050	53	37	26.450*	29	61	0.050	33	57	0.450	28	62	0.200	60	30	45.000*
E3.SPO.MTS	P3.3	21	69	4.050*	38	52	3.200	32	58	0.200	64	26	57.800*	30	60	0.000	35	55	1.250	70	20	80.000*	31	59	0.050
E3.MTS.SPO	P3.4	63	27	54.450*	36	54	1.800	42	48	7.200*	43	47	8.450*	50	40	20.000*	66	24	64.800*	53	37	26.450	67	23	68.450*
E3.MTS.SPO	P3.5	22	68	3.200	71	19	84.050*	89	1	174.050*	88	2	168.200*	90	0	180.000*	89	1	174.050*						
E3.MTS.SPO	P3.6	31	59	0.050	21	69	4.050*	67	23	68.450*	59	31	42.050*	76	14	105.800*	62	28	51.200*	75	15	101.250*	87	3	162.450*

*=significant at p <.05

Table 3.6

Number of correct responses for each trained relation during MTS training of each experimental session for all participants in Experiment 3.

Group	Participant	Session	Sample and correct comparison stimuli						Total % correct
			A1-B2	A2-B3	A3-B1	B1C2	B2-C3	B3-C1	
E3.SPO.MTS	P3.1	Session 1	2	2	5	7	3	2	35.00
		Session 2	3	**6**	6	**10**	**9**	**9**	71.67
		Session 3	6	**10**	**9**	**10**	**9**	8	86.67
		Session 4	**10**	**10**	8	**10**	**10**	**10**	96.67
E3.SPO.MTS	P3.2	Session 1	**9**	4	5	**10**	4	2	56.67
		Session 2	**10**	5	8	**9**	6	5	71.67
		Session 3	**10**	6	**9**	**10**	7	8	83.33
		Session 4	**10**	**9**	5	7	7	8	76.67
E3.SPO.MTS	P3.3	Session 1	7	3	**10**	5	5	4	56.67
		Session 2	**9**	**9**	**10**	**10**	**10**	**10**	96.67
		Session 3	**10**	**9**	**10**	**9**	**9**	**10**	95.00
		Session 4	**9**	**10**	**10**	**10**	**10**	**10**	98.33
E3.MTS.SPO	P3.4	Session 1	5	1	3	6	4	3	36.67
		Session 2	6	5	**10**	**10**	**10**	**10**	85.00
		Session 3	**10**	8	**10**	**10**	**10**	**10**	96.67
		Session 4	**10**	**10**	**10**	**10**	**10**	**10**	100.00
E3.MTS.SPO	P3.5	Session 1	9	6	4	10	5	**4**	63.33
		Session 2	**10**	**9**	**10**	**10**	**10**	**10**	98.33
		Session 3	**10**	**10**	**10**	**10**	**10**	**10**	100.00
		Session 4	–	–	–	–	–	–	–
E3.MTS.SPO	P3.6	Session 1	1	4	2	5	4	3	31.67
		Session 2	5	4	7	7	6	6	58.33
		Session 3	8	7	8	**10**	**10**	**9**	86.67
		Session 4	**10**	**10**	**10**	**10**	**10**	**10**	100.00

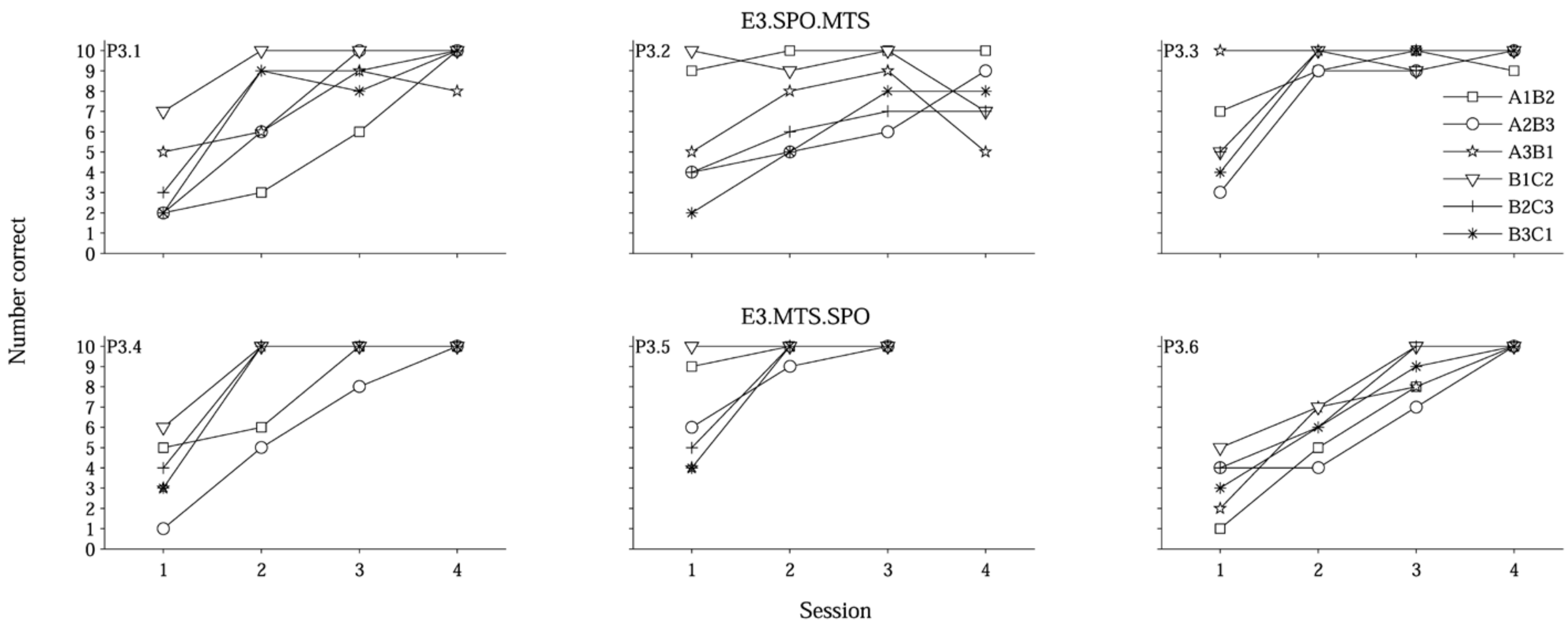

Figure 3.2 Number of correct responses made on each trained relation during MTS training across all sessions for all participants in Experiment 3.

Table 3.7

Chi-square tests for the distribution of correct and incorrect responses during MTS training (compared to responses expected by chance (correct (20/60), incorrect (40/60)) by all participants in Experiment 3.

Group	Participant	Session 1			Session 2			Session 3			Session 4		
		Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2
E3.SPO.MTS	P3.1	21	39	0.750	43	17	39.675*	52	8	76.800*	58	2	108.300*
E3.SPO.MTS	P3.2	34	26	14.700*	43	17	39.675*	50	10	67.500*	46	14	50.700*
E3.SPO.MTS	P3.3	34	26	14.700*	58	2	108.300*	57	3	102.675*	59	1	114.075*
E3.MTS.SPO	P3.4	22	38	0.300	51	9	72.075*	58	2	76.800*	60	0	120.000*
E3.MTS.SPO	P3.5	38	22	24.300*	59	1	114.075*	60	0	120.000*	_	_	_
E3.MTS.SPO	P3.6	19	41	0.750	35	25	16.875*	52	8	76.800*	60	0	120.000*

*=significant at p <.05

Discussion

This experiment examined the effect of using different stimuli in each training procedure on performance during the equivalence tests. Similarly to Experiments 1 and 2, only one participant here met the condition criterion during testing following either training procedure and there was no observed difference in effectiveness between the SPO and MTS procedures. However, the overall percentages of correct responses achieved by the participants here were generally greater than were achieved by the participants in Experiments 1 and 2.

A factorial repeated-measures ANOVA was conducted to compare percent correct on the symmetry and equivalence relations during the final session across procedures, procedural orders, and experiments. The results of this ANOVA are presented in Table 3.8. Findings that were statistically significant (at $p<0.05$) are indicated by an asterisk (*). Table 3.8 shows that there was significant between-subjects difference between the mean percent correct achieved in the final session across experiments, showing a moderate effect (η^2 = 0.415). Scheffé post hoc tests were conducted to compare each pair of experiments. The percentage of correct responses achieved by the participants in Experiments 1 ($\bar{x}$ = 49.13, *SD* = 33.30) and 2 ($\bar{x}$ = 35.14, *SD* = 22.96), and Experiments 1 and 3 ($\bar{x}$ = 61.53, *SD* = 27.994) did not differ significantly. However, the participants in the present experiment achieved a significantly greater mean percentage of correct responses than the participants in Experiment 2. The only procedural difference between Experiments 2 and 3 was the introduction of a new set of stimuli for MTS training and subsequent testing, so that different stimuli were used in each procedure. This suggests that the use of different stimuli in the SPO and MTS conditions improved the participants' accuracy on the symmetry and equivalence relations during testing of their final session, regardless of the order in which the procedures were completed.

As shown in Table 3.11, the mean percent correct on the symmetry relations ($\bar{x}$ = 56.505) was significantly greater than on the equivalence relations ($\bar{x}$ = 40.694). This finding showed a moderate effect size (0.515). This finding is supported by the results for each experiment individually. A statistically significant difference in performance on the symmetry and equivalence was reported for the results of Experiment 1. Although the findings for this effect in the separate ANOVAs for Experiments 2 and 3 were not statistically significant, the effect sizes reported were greater than those associated with the significant results of Experiment 1 and this current analysis. It is likely that the small samples used in Experiments 2 and 3 are responsible for this result.

Table 3.8

Results of an ANOVA to compare percent correct on the symmetry and equivalence relations of the final session for Experiments 1, 2, and 3.

Source	*df*	*F*	$\eta^2_{partial}$
Within-subjects effects			
Procedure	1,18	0.026	0.001
Procedure x Procedural Order	1,18	1.767	0.089
Procedure x Experiment	2,18	0.104	0.011
Procedure x Procedural Order x Experiment	2,18	0.959	0.096
Type of relation	1,18	19.113*	0.515
Type of relation x Procedural Order	1,18	1.255	0.065
Type of relation x Experiment	2,18	1.406	0.135
Type of relation x Procedural Order x Experiment	2,18	2.561	0.222
Procedure x Type of relation	1,18	1.835	0.093
Procedure x Type of relation x Procedural Order	1,18	2.048	0.102
Procedure x Type of relation x Experiment	2,18	0.381	0.041
Prcoedure x Type of relation x Procedural Order x Experiment	2,18	0.059	0.007
Between-subjects effects			
Procedural order	1,18	2.864	0.137
Experiment	2,18	6.380*	0.415
Procedural Order x Experiment	2,18	2.936	0.079

*=significant at $p<0.05$

The asymmetry in performance on the symmetry and equivalence relations in Experiments 1 and 2 is supported by the findings of Pilgrim and Galizio (1990; 1995) and Pilgrim et al. (1995). As outlined previously, those studies reported the reversal of training relations resulted in poorer performance on tests for equivalence involving the new relations. In those studies, participants were more likely to perform well on the symmetry relations, while responses on the equivalence relations were more likely to be consistent with the correct response for the equivalence test prior to reversal training. It could then have been suggested that the poor performance overall, and differences in performance on the symmetry and equivalence relations by the participants in Experiments 1 and 2 was due to the use of the same stimuli in conflicting relations. However, the asymmetry in performance on the symmetry and equivalence relations is also present in the findings of the present experiment. Most of the participants also failed to meet the condition criterion following either procedure of either session. However, as noted previously, the overall performance of the participants in the present experiment was significantly better than that of the participants in Experiment 2. Therefore, although only one participant met the session criterion, the use of different stimuli in each procedure did result in improved performance on the symmetry and equivalence tests. However, there was no difference in the effectiveness of the SPO and MTS procedures, and performance on the equivalence relations was still below the criterion for nearly all participants.

A possible contributor to the poor performance could be the form of the instructions given. In the three experiments reported so far, a few participants mentioned during the debriefing that they were not sure that they understood what they were supposed to be doing during their participation in the experiment. The instructions used in this experiment were the same as used by Leader and Barnes-Holmes (2001) but appear not to have been effective for all participants here. Another possible contributor to the present results was the number of training trials. As mentioned in the *Discussion* of Experiment 1, the maximum possible number of training trials completed by the participants in this study was less than the maximum that could have been completed by Leader and Barnes-Holmes (2001) participants. However, the participants who achieved equivalence in that study did so within fewer trials than completed by the participants here. As the percentage of correct responses achieved by the participants in the present, and previous, experiments increased across the sessions, it is possible that additional training and testing cycles or more training trials may aid in the formation of equivalence classes. The next experiment attempted to address some of these issues.

EXPERIMENT 4

Experiment 3 showed that the use of different stimuli in each training procedure aided in the formation of equivalence relations. However, while accuracy was better than that in the previous experiment, all participants still failed to meet the session criterion. Therefore, the use of the same stimuli in conflicting relations was not solely responsible for the failure of the participants to demonstrate equivalence. For some participants the debriefing revealed that they had found the instructions unclear. It is not known if the same was true for Leader and Barnes-Holmes's (2001b) participants. However, it appeared worth investigating this further.

The instructions provided in stimulus equivalence tasks have varied greatly across studies. Many studies have provided instructions that outline the behaviour required to complete a trial (e.g., Dymond & Barnes, 1997; Holth & Arntzen, 1998; O'Toole, Barnes-Holmes & Smith, 2007; Plaud, Gaither, Franklin, Weller, & Barth et al., 1998). In addition, some studies provide extra information, for example, an outline of what will happen during a trial (e.g., Plaud et al., 1998; Rehfeldt, 2003; Wilson & Hayes, 1996), information about feedback (e.g., de Rose, McIlvane, Dube, Galpin, & Stoddard, 1988; Minster, Jones, Eliffe & Muthukumaraswamy, 2006), a description of the contingencies (e.g., Schenk, 1994), a description of the criteria for progression through the experiment (e.g., Arntzen, 2006; Holth & Arntzen, 1998), or an instruction for the participants to try their best (e.g., Dixon et al., 2006; Minster et al., 2006) or to choose the comparison that matches or goes with the sample (Minster et al., 2006; Peoples et al, 1998). Some studies provide pre-training, or practice, trials (Barnes & Keenan, 1993; de Rose et al., 1988; Smeets & Barnes-Holmes, 2000), or prompting (e.g., de Rose et al., 1988; Dube et al., 1989). Instructions may be given verbally (e.g., Dixon et al., 2006; Harrison & Green, 1990), as part of a computer programme (e.g., Hayes et al., 1991; Minster et al., 2006), or in written or typed form (e.g., Eikeseth et al., 1997; Duarte et al., 1998; Smeets, Dymond & Barnes-Holmes, 2000). Some studies provided participants with all of their instructions at the beginning of the experiment (e.g., Markham et al., 2000; Peoples et al, 1998), while others gave instructions at the start of different stages of the experiment (e.g., Minster et al., 2006; O'Toole et al., 2007; Smeets et al., 1997). Thus, there was no consistent instructional procedure used across these studies.

There is some research that has studied the role of instructions in the formation of equivalence relations specifically. Nearly all of this research has used a MTS procedure (e.g., de Medeiros et al., 2003; Drake & Wilson, 2008; Duarte et al., 1998; Eikeseth et al., 1997;

Green et al., 1991; Rosales-Ruiz et al., 2000; Saunders et al., 1993; Sigurdardottir et al., 1990). Sigurdardottir et al. (1990), and Green et al. (1991) used spoken instructions. Sigurdardottir et al. (1990) found that detailed instructions that outlined the task, the contingencies, changes in the rates of reinforcement, and a directive to “try to get as many correct as you can” (p.63) increased rates of equivalence compared to minimal verbal instructions that outlined the feedback, and how to complete a trial. They suggest that the use of the words “go with” in the instructions prior to testing may have been instrumental in producing this difference (Sigurdardottir et al., 1990). Green et al. (1991) conducted a study that followed Sigurdardottir et al. (1990) and found little difference in the effectiveness of the instructions in producing equivalence responding. However, that study did identify that the more detailed instructions may have affected performance on a subsequent transfer of function task (Green et al., 1991).

Saunders et al. (1993) concluded that instructions during the initial training trials that outlined the correct baseline relations were necessary to facilitate the formation of equivalence relations by adolescents with mild intellectual disabilities. The abstract of de Medeiros et al. (2003) reports that providing instructions “clarifying the participant’s tasks” (p.165) increased the likelihood that the participants would demonstrate equivalence. A more recent study (Drake & Wilson, 2008) reported that undergraduate students performed better on equivalence tasks when the consequences for correct participation (finishing the experiment early and course credit) were outlined in the instructions. This suggests that these factors, and not just the feedback provided during training, were important.

Other studies have examined the use of written instructions (Duarte et al., 1998; Eikeseth et al., 1997; Rosales-Ruiz et al., 2000). Eikeseth et al. (1997) demonstrated that it is possible to train the baseline relations using written instructions by establishing rules about these relations. Using a similar format, Duarte et al. (1998) used written instructions that incorporated examples to train the baseline relations. That study found that restrictive instructions that required participants to adhere to the rules provided in the written instructions reduced the likelihood of equivalence compared to less restrictive instructions. Rosalez-Ruiz et al. (2000) looked at the role of different verbs in the instructions used in a written-format equivalence task. This involved instructions containing verbs that facilitated equivalence, “equals, is, is parallel to, goes with, and matches” (Rosalez-Ruiz et al., 2000, p.180) or did not facilitate equivalence, “eats, owes, pays, likes, and teaches” (p.180). They report that most participants responded differently relative to the verb used.

The instructions provided in SPO procedures have also varied across the studies that have used this procedure. For example, Smeets et al. (1997) used a simple directive to their adult participants to “look at the screen” (p.288) prior to beginning SPO training. The instructions provided prior to MTS testing in that study were the same as those used in the present study prior to MTS testing. Another experiment, with children, in that study used the instruction to “watch these pictures carefully” (p.294). An instructive demonstration was used to teach the children how to complete a trial on a MTS task prior to MTS testing (Smeets et al., 1997). In an associative-pairing based procedure, Fields et al. (1997) included an outline of the trial, how to respond on a trial and an instruction to “select the bottom word that goes with the top word” (p.671). The instructions used by Layng and Chase (2001) outlined how to complete a trial, that there would be a second part to the experiment, and the contingencies for correct responses. Most studies involving SPO procedures also involve MTS testing. The instructions used during testing in these studies (e.g., Leader et al, 1996; Leader and Barnes-Holmes, 2001b; Smeets et al., 1997) are similar to those used in other studies involving MTS. Therefore, as with the studies that have used an MTS task, there is little consistency in the instructions used across the studies that have used an associative based task such as SPO.

Only one study (Leader et al., 1996) has compared the use of different instructions in an SPO procedure. Experiment 2 of that study compared the effectiveness of minimal instructions “Look at the screen” (Leader et al., 1996, p.692) and more detailed instructions “During the first stage of this experiment you will be presented with nonsense syllables on the computer screen. You should pay close attention to this first stage because it is relevant to the second stage of the experiment” (Leader et al., p.688). Nearly all of the participants in that study demonstrated the formation of equivalence relations under both sets of instructions and the authors concluded that the minimal instructions were as effective as the detailed instructions (Leader et al., 1996). The instructions used in the MTS test following SPO training of that study were very similar to those used in the present study.

The two studies that compared the effectiveness of SPO and MTS procedures in the formation of equivalence relations used different instructions. The instructions used by Leader and Barnes-Holmes (2001b) are the same as used in Experiments 1to 3 of the present study. The instructions presented prior to SPO training by Clayton and Hayes (2004) differ from these in that they provide a more detailed outline of how each trial is presented and an instruction prior to SPO training to “please pay close attention because what you encounter here will be relevant during future stages of the experiment” (p.586). The instructions presented prior to matching-to-sample training also outline how to complete a trial in more detail than do the instructions used here and by Leader and Barnes-Holmes (2001b). Therefore, the present experiment assessed the effect of instructional specificity on the formation of equivalence relations in SPO and MTS procedures.

In summary, instructions provided prior to an equivalence task often outline how to complete a trial. There appears to be little consistency in the information that is provided supplementary to this. Additionally, there is little consistency in the focus of the studies that have looked at the role of instructions in equivalence tasks. While there has been a range of research completed, nearly all of these studies have used an MTS procedure, and they have reported conflicting findings. Only one study has examined the effect of instructions using a SPO procedure, and that study (Leader et al., 1996) found no difference in effect between minimal and detailed instructions.

As mentioned previously, some participants reported being unclear as to the task required. The research is unclear on the effect of the specificity of the instructions but it is possible that more specific instructions would have helped accuracy in that experiment. Thus, this next experiment aimed to compare the effect of the instructions used in the previous experiments with instructions that were more specific to the task. It was decided to use a between-subjects design to reduce the possibility of carryover effects if the same participants received differing instructions.

As pointed out previously, some of the participants in the previous experiments in the present study appeared to be getting more accurate across the training and testing cycles. Thus, it is possible that increasing the number of trials might improve performance. Given the constraints on the time that participants were available in the earlier experiments only four sessions were possible. However, it would be possible to increase the number of trials with a procedure in a session if only one procedure was used in a session. A between-subjects design across procedures allowed the number of training and testing cycles completed during each procedure to be increased. Within the maximum of four sessions and only one procedure, each participant could complete eight training and testing cycles, twice the number of cycles that could be completed by a participant on a single procedure in Experiments 1to 3.

Therefore, this experiment used a between-subjects design where each participant completed either SPO training followed by testing, or MTS training followed by testing. Half of the participants received instructions that were more specific to the task than those provided in Experiments 1to 3. The modified instructions aimed to clarify the task of the participant with an additional clause to each set of instructions. In SPO training the modified instructions made reference to the relevance of the SPO training procedure to the second part (testing) of the experiment. The instructions presented prior to MTS training or testing outlined that the participant's task was to choose the syllable that 'goes with' the one at the top. These participants also answered questions that were aimed to test their comprehension of the instructions prior to beginning the first experimental session. The remaining participants received the same instructions as were used in the previous experiments. Other procedural factors remained the same as in Experiments 1to 3 to allow comparison across the experiments.

Method

Participants, Ethics, Apparatus and Setting

Participant recruitment, ethics procedures, apparatus and setting were as those used in Experiment 3. Experiment 4 involved 12 participants in total, three in each of four groups (4.SPO, 4.MTS, 4.SPO.INST, 4.MTS.INST).

Stimuli

The stimuli used in Experiment 4 were the nine nonsense syllables used by Leader and Barnes-Holmes (2001b), in both conditions of Experiment 2 and in the SPO procedure of Experiment 3. To allow comparison to earlier experiments, the stimuli were organised into the same relations as used in Experiments 1 and 2 for each procedure.

Procedure

Experiment 4 was similar to Experiments 2 and 3, however each participant experienced one training procedure only, and two groups received instructions that were more specific to the task.

One procedure only. The participants in Groups 4.SPO (P4.1 - 4.3) and 4.SPO.INST (P4.7 - P4.9) completed two cycles of SPO training, and subsequent testing, during each experimental session. The participants in Group 4.MTS (P4.4 - P4.6) and 4.MTS.INST (P4.10 – P4.12) experienced MTS training, and subsequent testing, twice during each experimental session. The participants were informed by the experimenter that they would be completing two cycles within each session.

Instructions. The instructions provided at the beginning of training and testing for the participants in Groups 4.SPO and 4.MTS were the same as was provided to the participants in Experiments 1, 2, and 3. The participants in Groups 4.SPO.INST and 4.MTS.INST received instructions that were more specific to the task, and are given below:

SPO training:

In this stage of the experiment your task is to simply watch the screen. PLEASE PAY CLOSE ATTENTION TO WHAT YOU SEE AS THIS WILL BE RELEVANT DURING THE 2ND STAGE OF THE EXPERIMENT.

MTS training; testing following SPO training and MTS training:

In this stage of the experiment you must look at the syllable at the top, and then choose one of the three syllables at the bottom, by pressing one of the marked keys on the keyboard. YOUR TASK IS TO CHOOSE THE SYLLABLE THAT GOES WITH THE ONE AT THE TOP.

To choose the left syllable, press the marked key on the left.
To choose the middle syllable, press the marked key in the middle.
To choose the right syllable, press the marked key on the right.
Press the spacebar twice to continue

The participants in Group 4.MTS .INST were also provided with information regarding the consequences that they would receive during training and testing:

1st Stage:

In the first stage of the experiment you will receive feedback for each response. If you make a correct response, then the word 'correct' will be displayed on the monitor and you will hear a tone. If you make an incorrect response, then the word 'wrong' will be displayed on the monitor. If your response is incorrect you will not hear a tone.

2nd Stage

In the second stage of the experiment you will not receive any feedback on your responses.

No information regarding feedback was provided to the participants in Group 4.SPO.INST as they received no feedback during either training or testing. Once the participants in this group had read the instructions and feedback information, the instruction sheet was removed and they were required to answer a series of questions designed to test their comprehension of the instructions. For the participants in Group 4.SPO.INST, these questions were: 1) What is your task during the first stage of the experiment? 2) What is your task during the second stage of the experiment? 3) During the second stage of the experiment, how do you choose a syllable?

The participants in Group 4.MTS.INST had read the instructions they were asked: 1) What is your task during the experiment? 2) How do you choose a syllable during the experiment? 3) What feedback will you receive during the first stage of the experiment? 4) What feedback will you receive during the second stage of the experiment?

If a participant was unable to answer any question they were given the instruction sheet to read and were then asked the question again. The participants were required to provide correct answers to these questions prior to beginning the experiment. The instruction sheet and comprehension questions are given in Appendix B.

In all other respects, the procedure used in Experiment 4 was identical to Experiments 1 to 3.

Session criterion

The session criterion was met if a participant responded correctly on nine or more trials for each tested relation during one training and testing cycle.

Results

Number of sessions

The session criterion was met during the third training and testing cycle by two participants (P4.8 in Group SPO.INST and P4.12 in Group MTS.INST) in the present experiment. The remaining participants completed the maximum of four sessions (eight cycles).

Responses to comprehension questions

All of the participants who completed the instruction comprehension questions answered them all correctly, demonstrating that they understood the task required. Their answers to these responses are given in Appendix B.

Testing

Number correct on each tested relation and overall percent correct. Tables 4.1-4.4 show the number of correct responses for each trial type and the total percentages of correct responses on the symmetry and equivalence trials for the participants in Groups 4.SPO, 4.MTS, 4.SPO.INST, and 4.MTS.INST respectively. On these tables, numbers that appear in bold indicate that nine or more of the 10 trials on this tested relation resulted in a correct response.

Overall, for the participants who received the less specific instructions (Groups 4.SPO and 4.MTS), those who completed SPO training achieved greater percent corrects on the symmetry and equivalence relations during testing than was achieved by the participants who completed MTS training. This difference was not observed for the participants who received the more specific instructions and comprehension questions (Groups 4.SPO.INST and 4.MTS.INST). The overall percentage of correct responses achieved during testing increased across the cycles for all participants. Six participants (P4.1 and P4.2 in Group 4.SPO, P4.5 in Group 4.MTS, P4.8 in Group 4.SPO.INST, and P.4.11 and P4.12 in Group 4.MTS.INST) responded correctly on nine or more responses for all of the symmetry relations by the final training cycle.

Figure 4.1 shows percent correct on the symmetry and equivalence relations for each training and testing cycle for the participants in this experiment. All participants achieved a greater percentage of correct responses on the symmetry than equivalence relations in most cycles. Overall, the participants who received the more specific instructions achieved greater percentages of correct responses on the symmetry relations than the participants who received the less specific instructions. There appeared to be no effect of the instructions on percent correct achieved on the equivalence relations.

Table 4.1

Number of correct responses for each tested relation of each training and testing cycle in each experimental session for the participants who completed SPO training (Group E4.SPO)

Participant	Session	Condition	Tested relations - sample and correct comparison stimuli											Overall % Correct	Session criterion met
			Symmetry relations						% Correct Symmetry	Equivalence relations			% Correct Equivalence		
			B1-A1	B2-A2	B3-A3	C1-B1	C2-B2	C3-B3		C1-A1	C2-A2	C3-A3			
P4.1	Session 1	Cycle 1	**10**	2	3	6	5	1	45.00	4	4	6	46.67	45.56	N
		Cycle 2	**10**	6	7	**10**	**10**	**10**	88.33	2	6	7	50.00	75.56	N
	Session 2	Cycle 3	**10**	6	**9**	**10**	**10**	7	86.67	0	0	6	20.00	64.44	N
		Cycle 4	**10**	8	**9**	**10**	**10**	**10**	95.00	0	0	8	26.67	72.22	N
	Session 3	Cycle 5	**10**	**10**	**10**	**10**	**10**	**10**	100.00	0	0	7	23.33	74.44	N
		Cycle 6	**10**	**10**	**10**	**9**	**10**	**10**	98.33	0	0	7	23.33	73.33	N
	Session 4	Cycle 7	**10**	**9**	**10**	**9**	**10**	**9**	95.00	0	1	**9**	33.33	74.44	N
		Cycle 8	**10**	**10**	**10**	**10**	**10**	**10**	100.00	0	0	**9**	30.00	76.67	N
P4.2	Session 1	Cycle 1	6	3	4	**9**	6	7	58.33	0	5	5	33.33	50.00	N
		Cycle 2	**10**	8	**10**	**10**	8	**10**	93.33	8	3	7	60.00	82.22	N
	Session 2	Cycle 3	**10**	**10**	**10**	**10**	**10**	**10**	**100.00**	**10**	0	**9**	63.33	87.78	N
		Cycle 4	**10**	**10**	**10**	**10**	**10**	**10**	**100.00**	**10**	0	**10**	66.67	88.89	N
	Session 3	Cycle 5	**10**	**10**	**10**	**10**	**10**	**10**	**100.00**	**10**	0	**9**	63.33	87.78	N
		Cycle 6	**10**	**10**	**10**	**10**	**10**	**9**	**98.33**	**10**	0	**10**	66.67	87.78	N
	Session 4	Cycle 7	**10**	**10**	**10**	**10**	**10**	**10**	**100.00**	**10**	0	**10**	66.67	88.89	N
		Cycle 8	**10**	**10**	**10**	**10**	**10**	**10**	**100.00**	**10**	0	**10**	66.67	88.89	N
P4.3	Session 1	Cycle 1	4	3	2	4	2	7	36.67	2	0	8	33.33	35.56	N
		Cycle 2	3	6	2	2	2	2	28.33	4	3	4	36.67	31.11	N
	Session 2	Cycle 3	7	4	6	2	1	7	45.00	1	5	5	36.67	42.22	N
		Cycle 4	7	5	4	6	2	**9**	55.00	5	1	2	26.67	45.56	N
	Session 3	Cycle 5	**9**	**9**	3	8	2	**10**	68.33	2	1	7	33.33	56.67	N
		Cycle 6	7	**9**	6	5	4	**9**	66.67	2	4	2	26.67	53.33	N
	Session 4	Cycle 7	**9**	**10**	1	**9**	0	**9**	63.33	1	1	4	20.00	48.89	N
		Cycle 8	**10**	**10**	2	8	1	**10**	68.33	5	6	5	53.33	63.33	N

Table 4.2

Number of correct responses for each tested relation of each training and testing cycle in each experimental session for the participants who completed MTS training (E4.MTS)

Participant	Session	Condition	Tested relations - sample and correct comparison stimuli											Overall % Correct	Session criterion met
			Symmetry relations						% Correct Symmetry	Equivalence relations			% Correct Equivalence		
			B1-A3	B2-A1	B3-A2	C1-B3	C2-B1	C3-B2		C1-A2	C2-A3	C3-A1			
P4.4	Session 1	Cycle 1	6	6	4	2	5	5	46.67	1	5	2	26.67	40.00	N
		Cycle 2	3	4	5	2	0	7	35.00	4	1	1	20.00	30.00	N
	Session 2	Cycle 3	5	**10**	5	5	8	**9**	70.00	4	0	0	13.33	51.11	N
		Cycle 4	2	8	4	3	5	4	43.33	4	0	1	16.67	34.44	N
	Session 3	Cycle 5	3	6	1	4	8	5	45.00	1	7	5	43.33	44.44	N
		Cycle 6	4	5	1	3	9	5	45.00	3	3	3	30.00	40.00	N
	Session 4	Cycle 7	2	**9**	3	7	**10**	**9**	66.67	2	1	3	20.00	51.11	N
		Cycle 8	3	**10**	8	4	**10**	7	70.00	3	1	5	30.00	56.67	N
P4.5	Session 1	Cycle 1	7	**10**	4	8	1	**10**	66.67	5	7	7	63.33	65.56	N
		Cycle 2	**9**	**10**	6	**10**	8	8	85.00	0	**9**	6	50.00	73.33	N
	Session 2	Cycle 3	**10**	**10**	2	**10**	6	**10**	80.00	0	7	8	50.00	70.00	N
		Cycle 4	**9**	**10**	8	**10**	**9**	**10**	93.33	0	8	**10**	60.00	82.22	N
	Session 3	Cycle 5	**10**	**10**	**9**	**10**	**10**	**10**	**98.33**	0	**10**	**10**	66.67	87.78	N
		Cycle 6	**10**	**10**	8	**10**	**10**	**10**	96.67	0	**10**	**10**	66.67	86.67	N
	Session 4	Cycle 7	**10**	**9**	7	**10**	**10**	**10**	93.33	0	**10**	**10**	66.67	84.44	N
		Cycle 8	**9**	**10**	**9**	**10**	**10**	**9**	**95.00**	0	**10**	**10**	66.67	85.56	N
P4.6	Session 1	Cycle 1	3	2	1	4	5	3	30.00	5	1	4	33.33	31.11	N
		Cycle 2	4	3	5	1	4	2	31.67	3	4	3	33.33	32.22	N
	Session 2	Cycle 3	2	3	5	3	4	4	35.00	5	5	2	40.00	36.67	N
		Cycle 4	3	2	0	3	7	5	33.33	3	1	2	20.00	28.89	N
	Session 3	Cycle 5	4	3	3	4	7	1	36.67	0	3	4	23.33	32.22	N
		Cycle 6	1	3	4	2	**9**	6	41.67	4	3	1	26.67	36.67	N
	Session 4	Cycle 7	0	1	2	8	**10**	8	48.33	3	1	0	13.33	36.67	N
		Cycle 8	1	4	2	8	**10**	7	53.33	2	0	1	10.00	38.89	N

Table 4.3

Number of correct responses for each tested relation of each training and testing cycle in each experimental session for the participants who completed SPO training and received specific instructions and a comprehension test (E4.SPO.INST)

Participant	Session	Condition	Tested relations - sample and correct comparison stimuli											Overall % Correct	Session criterion met
			Symmetry relations						% Correct Symmetry	Equivalence relations			% Correct Equivalence		
			B1-A1	B2-A2	B3-A3	C1-B1	C2-B2	C3-B3		C1-A1	C2-A2	C3-A3			
P4.7	Session 1	Cycle 1	2	6	4	**10**	2	9	55.00	2	7	7	53.33	54.44	N
		Cycle 2	1	5	6	**10**	**10**	7	65.00	2	3	7	40.00	56.67	N
	Session 2	Cycle 3	6	**10**	8	**10**	**9**	10	88.33	1	5	4	33.33	70.00	N
		Cycle 4	5	**10**	**10**	**10**	**10**	**10**	91.67	0	2	8	33.33	72.22	N
	Session 3	Cycle 5	3	**10**	**9**	**10**	**9**	**10**	85.00	3	4	**9**	53.33	74.44	N
		Cycle 6	7	**9**	**10**	**10**	**10**	**10**	93.33	3	1	3	23.33	70.00	N
	Session 4	Cycle 7	7	**10**	**10**	**10**	**10**	**10**	95.00	1	0	4	16.67	68.89	N
		Cycle 8	**10**	8	**10**	**10**	**9**	**10**	95.00	2	5	3	33.33	74.44	N
P4.8	Session 1	Cycle 1	3	5	2	8	6	1	41.67	2	1	7	33.33	38.89	N
		Cycle 2	**0**	10	1	**10**	0	**10**	51.67	0	8	2	33.33	45.56	N
	Session 2	Cycle 3	**9**	**10**	**10**	**10**	**10**	**10**	**98.33**	**9**	**10**	**10**	**96.67**	**97.78**	**Y**
		Cycle 4	_	–	–	–	–	–	–	–	–	–	–	–	–
	Session 3	Cycle 5	_	–	–	–	–	–	–	–	–	–		–	–
		Cycle 6	_	–	–	–	–	–	–	–	–	–	–	–	–
	Session 4	Cycle 7	_	–	–	–	–	–	–	–	–	–	–	–	–
		Cycle 8	_	–	–	–	–	–	–	–	–	–	–	–	–
P4.9	Session 1	Cycle 1	0	4	6	**10**	**9**	**9**	63.33	0	4	4	26.67	51.11	N
		Cycle 2	**9**	**9**	8	**9**	**10**	**9**	90.00	3	3	3	30.00	70.00	N
	Session 2	Cycle 3	**10**	6	8	**9**	**10**	**10**	88.33	3	2	5	33.33	70.00	N
		Cycle 4	**9**	**9**	8	**10**	8	**10**	90.00	6	4	5	50.00	76.67	N
	Session 3	Cycle 5	**9**	8	7	**10**	**10**	**10**	90.00	7	3	4	46.67	75.56	N
		Cycle 6	**9**	**10**	8	**10**	**10**	**10**	95.00	8	7	5	66.67	85.56	N
	Session 4	Cycle 7	**9**	**9**	**9**	**9**	7	**10**	88.33	7	1	4	40.00	72.22	N
		Cycle 8	**10**	**9**	8	**10**	**10**	**10**	95.00	**9**	5	5	63.33	84.44	N

Table 4.4

Number of correct responses for each tested relation of each training and testing cycle in each experimental session for the participants who completed MTS training and received specific instructions and a comprehension test (E4.MTS.INST)

Participant	Session	Condition	Tested relations - sample and correct comparison stimuli											Overall % Correct	Session criterion met
			Symmetry relations						% Correct Symmetry	Equivalence relations			% Correct Equivalence		
			B1-A1	B2-A2	B3-A3	C1-B1	C2-B2	C3-B3		C1-A1	C2-A2	C3-A3			
P4.10	Session 1	Cycle 1	5	0	8	1	6	4	40.00	4	8	0	40.00	40.00	N
		Cycle 2	3	4	1	3	6	4	35.00	2	3	5	33.33	34.44	N
	Session 2	Cycle 3	0	**9**	8	**10**	**10**	**10**	78.33	8	1	0	30.00	62.22	N
		Cycle 4	0	**10**	**10**	**10**	**10**	**10**	83.33	**10**	0	0	33.33	66.67	N
	Session 3	Cycle 5	**9**	**10**	8	**10**	**10**	**10**	95.00	0	0	0	0.00	63.33	N
		Cycle 6	**9**	**10**	**10**	**10**	**10**	**10**	**98.33**	0	0	1	3.33	66.67	N
	Session 4	Cycle 7	**10**	**10**	**10**	**10**	**10**	**10**	**100.00**	0	0	0	0.00	66.67	N
		Cycle 8	**10**	6	**10**	**10**	**10**	**10**	93.33	0	1	0	3.33	63.33	N
P4.11	Session 1	Cycle 1	3	2	1	**9**	1	4	33.33	0	2	1	10.00	25.56	N
		Cycle 2	2	3	1	**9**	**9**	**9**	55.00	1	7	1	30.00	46.67	N
	Session 2	Cycle 3	5	8	0	**10**	**9**	**10**	70.00	0	**9**	6	50.00	63.33	N
		Cycle 4	8	6	6	**10**	**10**	**9**	81.67	0	**10**	3	43.33	68.89	N
	Session 3	Cycle 5	**9**	**9**	**10**	**9**	**10**	**10**	**95.00**	0	**10**	0	33.33	74.44	N
		Cycle 6	**10**	**9**	**10**	**10**	**10**	**10**	**98.33**	0	**10**	0	33.33	76.67	N
	Session 4	Cycle 7	**10**	**9**	**10**	**10**	**10**	**10**	**98.33**	0	**10**	0	33.33	76.67	N
		Cycle 8	**10**	**10**	**10**	**10**	**10**	**10**	**100.00**	0	**10**	0	33.33	77.78	N
P4.12	Session 1	Cycle 1	2	2	2	2	6	**10**	40.00	0	1	0	3.33	27.78	N
		Cycle 2	**10**	3	7	**10**	**9**	**10**	81.67	0	1	3	13.33	58.89	N
	Session 2	Cycle 3	**10**	**9**	**10**	**10**	**10**	**10**	**98.33**	**10**	**10**	**10**	**100.00**	**98.89**	**Y**
		Cycle 4	–	–	–	–	–	–	–	–	–	–	–	–	–
	Session 3	Cycle 5	–	–	–	–	–	–	–	–	–	–		–	–
		Cycle 6	–	–	–	–	–	–	–	–	–	–	–	–	–
	Session 4	Cycle 7	–	–	–	–	–	–	–	–	–	–	–	–	–
		Cycle 8	–	–	–	–	–	–	–	–	–	–	–	–	–

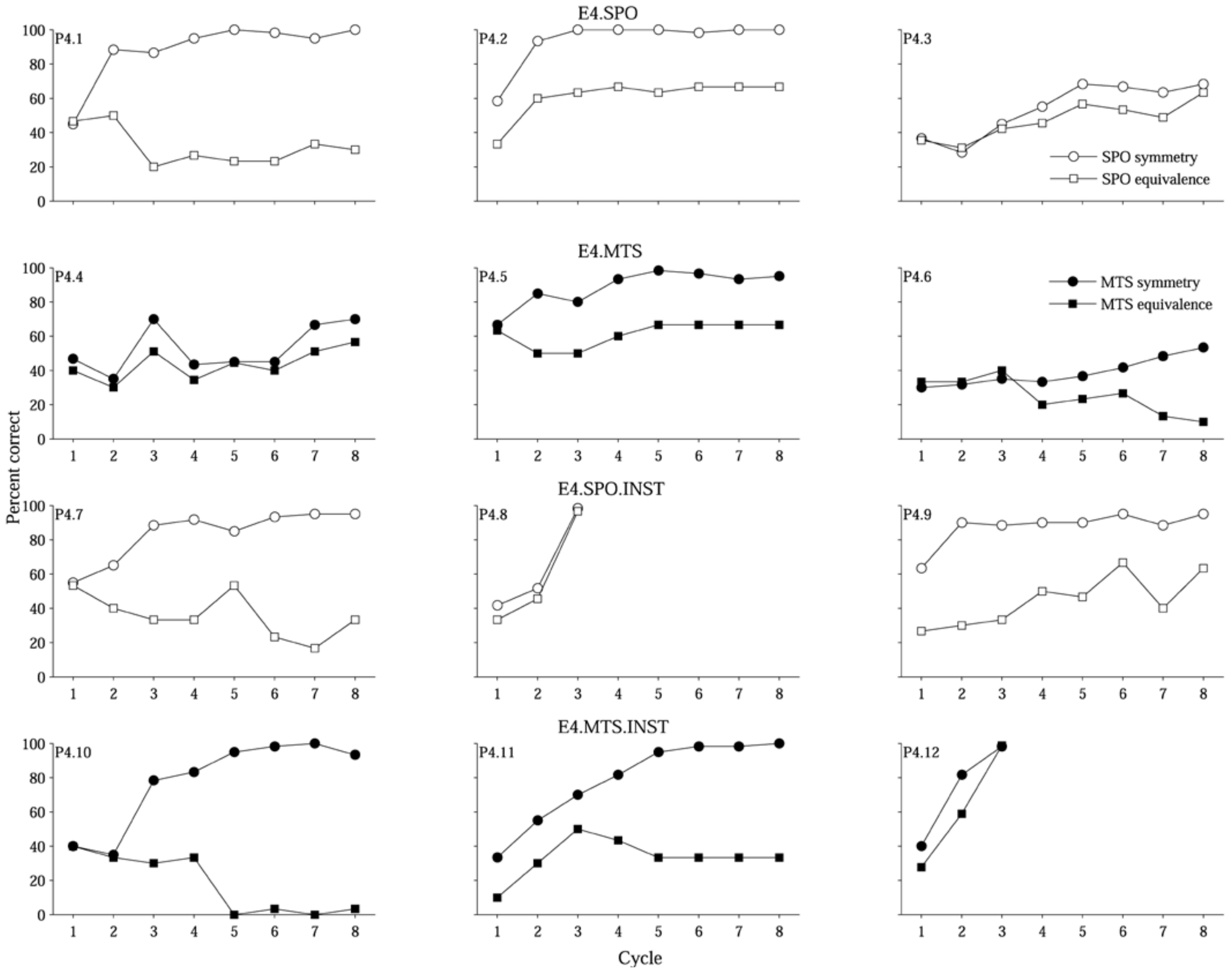

Figure 4.1 Percent correct on the symmetry and equivalence relations following SPO and MTS training across all sessions for all participants in Experiment 4.

The data for five participants (P4.3 in Group 4.SPO, P4.4 in Group 4.MTS, P4.8 and P4.9 in Group 4.SPO.INST, and P4.12 in Group 4.MTS.INST) show an upward trend in percent correct on both the symmetry and equivalence relations across the cycles. Data for only three cycles are presented for two of these participants (P4.8 and P4.12) as these participants met the session criterion during the third cycle.

Table 4.5 shows the results of the factorial repeated-measures ANOVA conducted to compare the percentage of correct responses achieved on the symmetry and equivalence relations of the final training and testing cycle by each participant. Statistically significant results are indicated by an asterisk (*). There was a statistically significant difference in percent correct on the symmetry and equivalence relations, and this difference was associated with a large effect size (0.705). Non-significant results were obtained for all of the within-subjects interactions. There were no significant between-subjects effects. The effect size for the instructions was above Fergusson's (2009) RMPE of 0.14. All other effect sizes were below the RMPE.

Table 4.5

Results of an ANOVA to compare percent correct on the symmetry and equivalence relations of the final session for both procedures for all groups in Experiment 4.

Source	*df*	*F*	$\eta^2_{partial}$
Within-subjects effects			
Type of relation	1,8	19.121*	0.705
Type of relation x Procedure	1,8	0.236	0.029
Type of relation x Instructions	1,8	0.033	0.004
Type of relation x Procedure x Instructions	1,8	0.369	0.044
Between-subjects effects			
Procedure	1,8	1.053	0.116
Instructions	1,8	1.360	0.145
Procedure x Instructions	1,8	0.078	0.010

*=significant at $p<0.05$

Correct vs. Incorrect (compared to chance) during testing. Most participants failed to meet the session criterion during any training and testing cycle. As with Experiments 1-3, χ^2 tests were conducted to compare the number of correct and incorrect responses by each participant compared to the number predicted by chance. The distribution of responses predicted by chance was the same as in previous experiments.

Table 4.6 shows the number of correct and incorrect responses made by each participant during each training and testing cycle, and the χ^2 statistic for each test. Statistically significant χ^2 statistics are followed by an asterisk (*). Most participants made a significantly greater number of correct responses than predicted by chance during most training and testing cycles. The exception to this was P4.6 (Group 4.MTS) who produced a response distribution than did not differ from chance significantly during any training and testing cycle.

MTS training

Number correct on each tested relation and overall percent correct. Table 4.7 and Figure 4.2 show the number of correct responses achieved on each trained relation during each training and testing cycle by the participants who completed MTS training (Groups 4.MTS and 4.MTS.INST). As shown in Table 4.7 and Figure 4.2, five of the six participants who completed MTS training responded correctly on nine or more trials for each of the trained relations by their final training and testing cycle. One participant (P4.6 in Group 4.MTS) did not achieve this during any cycle. Overall, the number of relations on which nine or more correct responses were made increased across the sessions for all participants.

Correct vs. incorrect (compared to chance) during MTS training. As in Experiments 1-3, χ^2 tests for goodness of fit were conducted to compare the distribution of correct and incorrect responses made during MTS training to the distribution predicted by chance. The response distribution predicted by chance was the same as in the previous experiments.

Table 4.8 shows the number of correct and incorrect responses made on each trained relation during each training and testing cycle by the participants who completed MTS training (Groups 4.MTS and 4.MTS.INST), and the χ^2 statistic for each test. As shown in Table 4.8, all participants made a greater number of correct responses than predicted by chance during most sessions of MTS training.

Table 4.6

Chi-square tests for the distribution of correct (Corr.) and incorrect (Incor.) responses (compared to responses expected by chance (correct (30/90), incorrect (60/90)) for all participants in Experiment 4.

Group	Participant	Session 1						Session 2						Session 3						Session 4					
		Cycle 1			Cycle 2			Cycle 3			Cycle 4			Cycle 5			Cycle 6			Cycle 7			Cycle 8		
		Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2
E4.SPO	P4.1	41	49	6.050*	68	22	72.200*	58	32	39.200*	65	25	61.250*	67	23	68.450*	66	24	64.800*	67	23	68.450*	69	21	76.050*
E4.SPO	P4.2	45	45	11.250*	74	16	96.800*	79	11	120.050*	80	10	125.000*	79	11	120.050*	79	11	120.050*	80	10	125.000*	80	10	125.000*
E4.SPO	P4.3	32	58	0.200	28	62	0.200	38	52	3.200	41	49	6.050*	51	39	22.050*	48	42	16.200*	44	46	9.800*	57	33	36.450*
E4.MTS	P4.4	36	54	1.800	27	63	0.450	46	44	12.800*	31	59	0.050	40	50	5.000*	36	54	1.800	46	44	12.800*	51	39	22.050*
E4.MTS	P4.5	59	31	42.050*	66	24	64.800*	63	27	54.450*	74	16	96.800*	79	11	120.050*	78	12	115.200*	76	14	105.800*	77	13	110.450*
E4.MTS	P4.6	28	62	0.200	29	61	0.050	33	57	0.450	26	64	0.800	29	61	0.050	33	57	0.450	33	57	0.450	35	55	1.250
E4.SPO.INST	P4.7	49	41	18.050*	51	39	22.050*	63	27	54.450*	65	25	61.250*	67	23	68.450*	63	27	54.450*	62	28	51.200*	67	23	68.450*
E4.SPO.INST	P4.8	35	55	1.250	41	49	6.050*	88	2	168.200*	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
E4.SPO.INST	P4.9	46	44	12.800*	63	27	54.450*	63	27	54.450*	69	21	76.050*	68	22	72.200*	77	13	72.200*	65	25	61.250*	76	14	105.800
E4.MTS.INST	P4.10	36	54	1.800	31	59	0.050	56	34	33.800*	60	30	45.000*	57	33	36.450*	60	30	45.000*	60	30	45.000*	57	33	36.450*
E4.MTS.INST	P4.11	23	67	2.450	42	48	7.200*	57	33	36.450*	62	28	51.200*	67	23	68.450*	69	21	76.050*	69	21	76.050*	70	20	80.000*
E4.MTS.INST	P4.12	25	65	1.250	53	37	25.219*	89	1	174.050*	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–

*=significant at p <.05

Table 4.7

Number of correct responses for each trained relation during MTS training of each experimental session for all participants in E4.MTS and E4.MTS.INST.

Group	Participant	Session	Cycle	Sample and correct comparison stimuli						Total % correct
				A1-B2	A2-B3	A3-B1	B1C2	B2-C3	B3-C1	
E4.MTS	P4.4	Session 1	Cycle 1	**9**	5	5	3	2	4	46.67
			Cycle 2	**10**	5	0	**10**	7	4	60.00
		Session 2	Cycle 3	**10**	3	5	**10**	**10**	**9**	78.33
			Cycle 4	**10**	**9**	2	**10**	**10**	**10**	85.00
		Session 3	Cycle 5	**10**	8	7	**10**	**9**	**10**	90.00
			Cycle 6	**10**	**10**	7	**10**	**10**	**10**	95.00
		Session 4	Cycle 7	**10**	**10**	**10**	**9**	**9**	**10**	**96.67**
			Cycle 8	**10**	**10**	**10**	**10**	**10**	**10**	**100.00**
E4.MTS	P4.5	Session 1	Cycle 1	5	3	5	5	8	**6**	53.33
			Cycle 2	**10**	6	**10**	**9**	**10**	7	86.67
		Session 2	Cycle 3	**10**	**10**	**10**	8	**10**	**10**	96.67
			Cycle 4	**10**	**10**	**10**	**10**	**10**	**10**	**100.00**
		Session 3	Cycle 5	**10**	**10**	**10**	**10**	**10**	**10**	**100.00**
			Cycle 6	**10**	8	**10**	**10**	**10**	**10**	96.67
		Session 4	Cycle 7	**10**	**10**	**10**	**10**	**10**	**10**	**100.00**
			Cycle 8	**10**	**10**	**10**	**10**	**10**	**10**	**100.00**
E4.MTS	P4.6	Session 1	Cycle 1	2	8	2	6	9	7	56.67
			Cycle 2	1	5	5	5	4	3	38.33
		Session 2	Cycle 3	3	3	4	4	2	3	31.67
			Cycle 4	2	1	2	5	6	8	40.00
		Session 3	Cycle 5	2	1	3	7	7	8	46.67
			Cycle 6	4	7	7	**10**	**9**	**9**	76.67
		Session 4	Cycle 7	3	7	5	**9**	**10**	**9**	71.67
			Cycle 8	6	7	**10**	**10**	**10**	**10**	88.33
E4.MTS.INST	P4.10	Session 1	Cycle 1	0	1	6	**10**	8	0	41.67
			Cycle 2	7	5	**10**	**10**	0	4	60.00
		Session 2	Cycle 3	**10**	8	**10**	**10**	3	8	81.67
			Cycle 4	**10**	**10**	**10**	**10**	**10**	**10**	**100.00**
		Session 3	Cycle 5	**10**	**10**	**10**	**10**	**10**	**10**	**100.00**
			Cycle 6	**10**	**10**	**10**	**10**	**10**	**10**	**100.00**
		Session 4	Cycle 7	**10**	**10**	**10**	**10**	**10**	**10**	**100.00**
			Cycle 8	**10**	**10**	**10**	**10**	**10**	**10**	**100.00**
E4.MTS.INST	P4.11	Session 1	Cycle 1	4	2	4	5	2	7	40.00
			Cycle 2	**10**	0	**10**	**9**	4	9	70.00
		Session 2	Cycle 3	7	**0**	**9**	8	**10**	**10**	73.33
			Cycle 4	**10**	3	**10**	**10**	**9**	**10**	86.67
		Session 3	Cycle 5	**9**	**9**	**10**	**10**	**10**	**10**	**96.67**
			Cycle 6	8	**9**	**10**	**10**	**10**	**10**	95.00
		Session 4	Cycle 7	**9**	**10**	**10**	**10**	**10**	**10**	**98.33**
			Cycle 8	**9**	**10**	**10**	**10**	**10**	**10**	**98.33**

Table 4.7 continued.

Group	Participant	Session	Cycle	Sample and correct comparison stimuli						Total % correct
				A1-B2	A2-B3	A3-B1	B1C2	B2-C3	B3-C1	
E4.MTS.INST	P4.12	Session 1	Cycle 1	3	2	5	4	6	0	33.33
			Cycle 2	**9**	7	**10**	8	**10**	8	86.67
		Session 2	Cycle 3	**10**	**10**	**10**	**10**	**10**	**10**	**100.00**
			Cycle 4	_	_	_	_	_	_	_
		Session 3	Cycle 5	_	_	_	_	_	_	_
			Cycle 6	_	_	_	_	_	_	_
		Session 4	Cycle 7	_	_	_	_	_	_	_
			Cycle 8	_	_	_	_	_	_	_

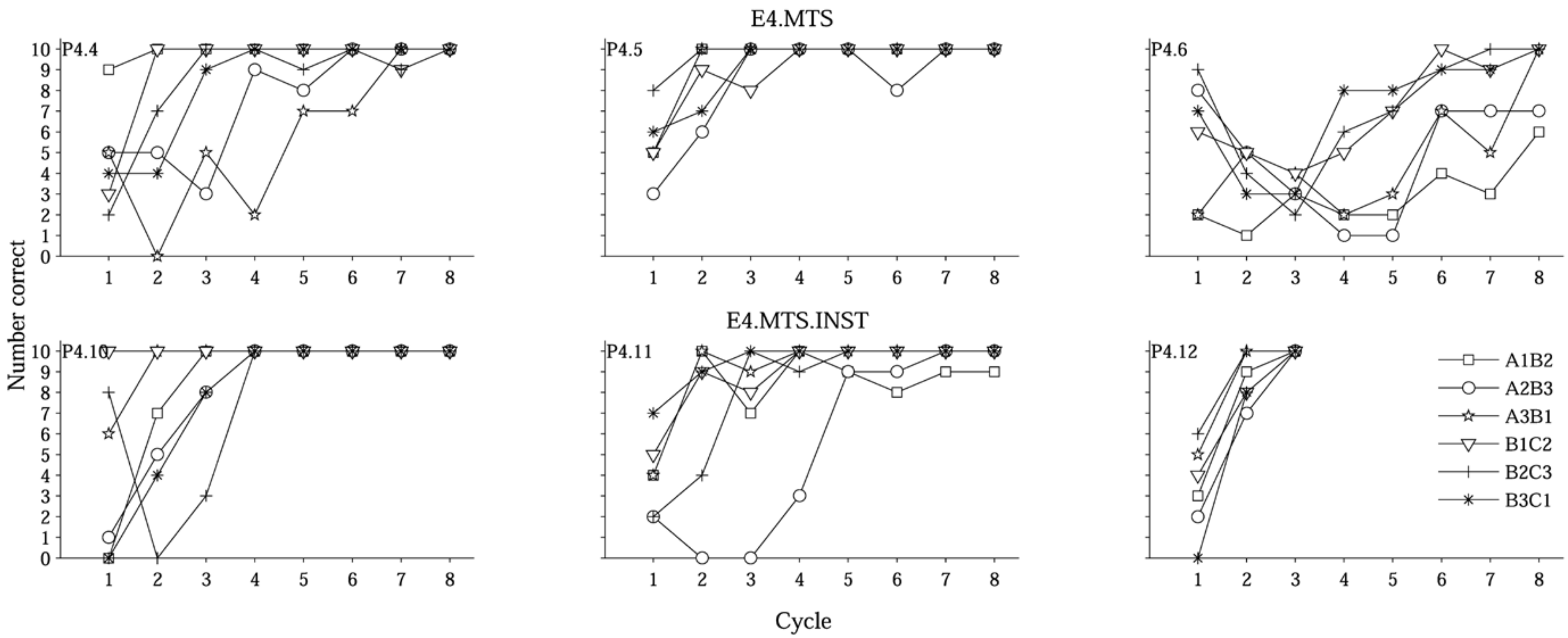

*Figure 4.*2 Number of correct responses made on each trained relation during MTS training across all sessions for all participants in Experiment 4.

Table 4.8

Chi-square tests for the distribution of correct and incorrect responses (compared to responses expected by chance (correct (20/60), incorrect (40/60) by the participants who completed MTS training (Groups E4.MTS and E4.MTS.INST).

Group	Participant	Session 1						Session 2						Session 3						Session 4					
		Cycle 1			Cycle 2			Cycle 3			Cycle 4			Cycle 5			Cycle 6			Cycle 7			Cycle 8		
		Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2
E4.MTS	P3.1	28	32	4.800*	36	24	19.200*	47	13	54.675*	51	9	72.075*	54	6	86.700*	57	3	102.675*	58	2	108.300*	60	0	120.000*
E4.MTS	P3.2	32	28	10.800*	52	8	76.800*	58	2	108.300*	60	0	120.000*	60	0	120.000*	58	2	108.300*	60	0	120.000*	60	0	120.000*
E4.MTS	P3.3	34	26	14.700*	23	37	0.675	19	41	0.075	24	36	1.200	28	32	4.800*	46	14	50.700*	43	17	50.475*	53	7	81.675*
E4.MTS.INST	P4.10	25	35	1.875	36	24	19.200*	49	11	63.075*	60	0	120.000*	60	0	120.000*	60	0	120.000*	60	0	120.000*	60	0	120.000*
E4.MTS.INST	P4.11	24	36	1.200	42	18	36.300*	44	16	43.200*	52	8	76.800*	58	2	76.800*	57	3	102.675*	59	1	114.075*	59	1	114.075*
E4.MTS.INST	P4.12	20	40	0.000	52	8	76.800*	60	0	120.000*	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_

Discussion

This experiment examined the effect of instructions that were more specific to the task than had been used in Experiments 1 to 3 and Leader and Barnes-Holmes (2001b). The use of a between-subjects research design meant that the maximum number of training and testing cycles completed per procedure was increased also. All of the participants who received the more specific instructions and the comprehension questions were able to answer the questions correctly.

As with Experiments 1 to 3, there was no difference here in the effectiveness of the SPO and MTS training procedures and the participants performed best on the symmetry relations. In this experiment the difference in performance on the symmetry and equivalence relations was statistically significant, and the effect size was larger than seen in Experiments 1 to 3. Overall, percent correct increased across sessions for participants in all groups, and two participants achieved the session criterion. Both of these participants had received the more specific instructions. This suggested that adding the more specific instructions increased the likelihood of meeting the session criterion. There was no significant difference in performance between the groups who had received the more specific instructions and those who had received the original instructions during the final session; however, the effect size was above Ferguson's (2009) RMPE. This effect size is small which shows that the differing instructions had a slight effect on performance and the effect was similar for both the SPO and MTS procedures. However, while the instructions had a slight effect, accuracy on the equivalence relations was still poor for most participants who received the more specific instructions. The more specific instructions appeared to also have little effect on performance during training. Therefore, further increasing the instructional specificity is unlikely to result in a substantial improvement in performance on the equivalence relations.

Leader et al. (1996) found no effect of more specific instructions on equivalence class formation with an SPO procedure. However, the instructions used by Leader et al. (1996) and here differ. In particular, the instructions used during MTS testing in the present study outlined that the participants were to choose the comparison that 'goes with' the sample. This instruction was not given by Leader et al. (1996). Sigurdardottir et al. (1990) considered the words 'go with' to be an important component of instructions used to increase the likelihood of equivalence. The findings of the present study give little support for this. As the instructions used by both Leader et al. (1996) and Sigurdardottir et al. (1990) both differ from those used here, and neither included a comprehension task, any comparisons with the present study should be treated with caution. However, it appears that, as with previous studies, the results of the present study do little to clarify the role of instructions in stimulus equivalence.

The less specific instructions in the present study were identical to those used by Leader and Barnes-Holmes (2001b). The more specific instructions included the addition of one task-specific phrase to each set of instructions. The instructions provided prior to SPO training told the participants to pay attention and outlined the relevance of the task to a later stage of the experiment. The instructions provided prior to MTS training and testing (following both training procedures) included a directive to choose the syllable that 'goes with' the sample. While the findings of the present experiment suggest a slight advantageous effect of the more specific instructions, Sidman (1992) suggests that the procedures used in equivalence tasks should be considered carefully as "if we tell our subject that stimuli "go with" each other...the data may then tell more about the subject's verbal history than about the effects of current experimental operations" (p.22). While this statement was made a number of years ago, the lack of agreement between the findings of studies exploring the effect of instructions on the formation of equivalence relations suggests that it is still relevant. The introduction of a particular statement may have differing effects based on each participant's history with that verbal statement.

The use of a between-subjects design in this experiment allowed the number of training and testing cycles to be increased. While most participants failed to reach the session criterion, overall percent correct generally increased across the sessions and was still increasing by the final session. This suggests that increasing the number of training trials, or the number of training and testing cycles further may aid in the formation of equivalence relations.

In summary, the specificity of the instructions had little effect on performance on the equivalence relations in the present experiment. This finding, when taken in conjunction with the range of procedures and conflicting findings in the existing literature, and with Sidman's (1992) comments about the effect of individual verbal histories suggests this is an area that requires more attention. Although this would be an interesting avenue of research, the findings to date do not identify a clear direction for research. There are, however, (as outlined in the *Introduction),* other procedural factors that could have affected the effectiveness of this procedure in facilitating the formation of equivalence relations. One of these factors is the arrangement of the stimuli in the trained relations (e.g., Fields et al., 1999; Saunders & Green, 1999). The effects of the arrangement of stimuli were explored further in the next two experiments (Experiments 5 and 6). Comparisons of the present data with earlier experiments suggested that more training trials or more training and testing cycles might also aid the development of equivalence relations. This was also examined further in Experiment 6.

EXPERIMENT 5

One factor that may have affected the performance of the participants in Experiments 1 to 4 of the present study is the arrangement of the stimuli during training the relations. The present study used a linear arrangement (A-B, B-C). That is, when the sample stimulus was from the A-stimulus set, the comparison stimuli were always the B-stimulus set, and when the sample stimulus was from the B-stimulus set, the comparison stimuli were always the C-stimulus set. However, there are two other stimuli arrangements that have been used in equivalence research. The first of these, called a many-to-one (MTO) arrangement, also known as, multiple-sample single-comparison (Saunders, Wachter & Spradlin, 1988), and comparison-as-node (CaN) (Fields et al, 1999) arrangements, involves the same set of stimuli being used as comparison stimuli on all training trials. For example, a B-A, C-A arrangement has the A set of stimuli as the comparison stimuli on all trials, while the sample stimuli are from either the B or C set. The second alternative is termed a one-to-many (OTM) arrangement, also known as single-sample multiple-comparison and sample-as-node (SaN) (Fields et al., 1999) arrangements. In a OTM arrangement with three sets of stimuli, the comparison stimuli are from either of two sets while the sample stimuli are always from the same set. In this case, an A-B, A-C arrangement would involve a sample stimulus from the A set of stimuli on every trial and the comparison stimuli would be either the B or C sets.

Some research has examined the comparative effectiveness of these different arrangements at facilitating the formation of equivalence relations. While most of the research examining the differential effect of the linear, MTO or OTM arrangements has used a MTS procedure (Fields, Hobbie-Reeve, Adams & Reeve, 1999; Hove, 2003; Spradlin & Saunders, 1986; Saunders, Wachter & Spradlin, 1988a; Saunders et al., 1993; Saunders et al., 1999; Smeets, Barnes-Holmes & Roche, 2001), a SPO procedure has also been used (Smeets et al., 1997; Leader et al., 2000).

A number of the MTS studies have reported the MTO arrangement to be more effective than linear or OTM (Fields et al., 1999; Hove, 2003; Spradlin & Saunders, 1986; Saunders et al., 1988; Saunders et al., 1993; Saunders et al., 1999; Smeets et al., 2001;). An early study by Spradlin and Saunders (1986) found the MTO procedure to be more effective than the OTM procedure in facilitating the formation of equivalence relations. This was followed by a study by Saunders et al. (1988), who examined the formation of equivalence relations by six intellectually disabled adolescents and young adults. They reported that the MTO procedure resulted in the formation of equivalence relations by more of the participants than the OTM procedure. The greater effectiveness of the MTO procedure was supported by the findings of Saunders et al. (1993) with adolescent and adult participants who had mild intellectual disabilities. Saunders et al. (1988) used a procedure similar to that used by Saunders et al. (1999) to study the effectiveness of the MTO and OTM arrangements in facilitating equivalence with preschool children. They reported that the MTO arrangement was the most effective. Hove (2003) compared performance on MTO and OTM arrangements using a MTS procedure with college students, and reported the MTO arrangement to be the most effective. The participants in this study who did not demonstrate equivalence were generally more accurate on tests for symmetry than for equivalence (Hove, 2003).

Fields et al. (1999) hypothesised the greater effectiveness of the MTO procedure was true for participants who had “limited behavioural repertoires” (p.707), including children, and adults with intellectual or developmental disabilities. They suggested that no difference in effectiveness was seen between the methods when studies involved normally developing adults and older children because the task was not challenging. This hypothesis was also raised by Saunders et al. (1993) who refer to an unpublished study that found no difference in the effectiveness of the MTO and OTM procedures in three normally developing children aged 8 to 14 years. To test this hypothesis, Fields et al. (1997) used equivalence classes that used a large number of stimuli to produce a more challenging task for normally developed adults. The results of this study show that differences in the effectiveness of the MTO and OTM arrangements were evident when the equivalence task was more difficult, as measured by the number of stimuli in each equivalence class (5- or 7-stimuli classes). As with the previous studies, the MTO arrangement was found to be more effective than the OTM arrangement (Fields et al., 1999).

Saunders and Green (1999) argue that the greater effectiveness of the MTO compared to the OTM and linear arrangements is due to the failure of either of these later arrangements to present all of the training discriminations that would be “subsequently required for consistently positive outcomes on all tests for the properties of equivalence” (p.117). Saunders et al. (1993) and Saunders and Green (1999) explain that the MTO training procedure involves learning discriminations between each stimulus and every other stimulus in the experiment. However, the OTM procedure only involves discriminations within each set of stimuli, while between set discriminations are required to respond correctly on the symmetry and equivalence relations. Saunders and Green (1999) state that the linear arrangement provides the discriminations required to respond correctly on the symmetry but not equivalence relations. Saunders and McEntee (2004) extended this hypothesis further, and found that performance on linear equivalence tasks by adults with mild intellectual disabilities could be improved when alterations to the procedure result in all of the required discriminations to be learnt. That study reported similar findings with a linear arrangement as would be expected with a OTM arrangement, based on Saunders and Green's (1999) hypothesis. When the procedure with the linear arrangement was altered to include more of the discriminations required to respond correctly on the equivalence task, performance was similar to that expected using a MTO arrangement. Fields et al. (1999) also support Saunders and Green's (1999) hypothesis, but suggest that the differences in performance observed between the OTM and MTO arrangements are also a function of the number of stimuli in each class (which alters the number of trained relations), and are inversely related to the behavioural repertoires of the participants. Therefore, when equivalence classes involving a small number of stimuli are used with participants who have larger behavioural repertoires (normally developed adults and older children) the differences in effectiveness between the OTM and MTO arrangements become negligible (Fields et al., 1999). This suggestion is not supported by the data from Hove's (2003) study where normally developing adults (college students) were more likely to develop equivalence between 3-stimuli classes with a MTO than a OTM arrangement. With college students and a small class size, Fields et al's (1999) hypothesis suggests that no difference in effectiveness between the two arrangements would be found.

Arntzen and Holth (1997) examined the effectiveness of the linear, MTO and OTM arrangements on facilitating equivalence tests only. They also included a group of participants who experienced a linear arrangement, and were tested for both symmetry and equivalence. This study concluded that the OTM arrangement was more effective at facilitating equivalence than the MTO arrangement, and that both of these arrangements were more effective than either of the linear arrangements (Arntzen & Holth, 1997). Most of Arntzen and Holth's (1997) participants who completed both symmetry and equivalence tests performed best on the symmetry relations but failed to demonstrate equivalence. The differences in findings of this and previous studies that found the MTO arrangement to be superior are suggested by Fields et al. (1999) to be due to procedural differences between the studies.

The poor effectiveness of the linear arrangement was supported by a later study (Arnzten & Holth, 2000) that used a repeated measures design to assess the effectiveness of the three procedures within each subject. This study showed that OTM was more likely than

MTO to facilitate equivalence, and that both of these arrangements were more likely to result in equivalence than the linear arrangement. Also, the superiority of the OTM arrangement was still seen when the number of stimuli in the classes were increased (Arntzen & Holth, 2000). A more recent study (Smeets & Barnes-Holmes, 2005) reported no difference in the effectiveness of the MTO and OTM procedures in facilitating equivalence performances by preschool children. This finding conflicts with Fields et al.'s (1999) suggestion the MTO is more effective than OTM when the participants are young children. Saunders, Chaney and Marquis (2005), in a study with senior citizens, found little difference in the effectiveness of the linear, MTO and OTM procedures using a delayed-MTS (DMTS) procedure. In a second experiment in this study, a 0 s DMTS procedure was used. With this procedure, the MTO and OTM arrangements were slightly more effective than the linear arrangement. Arntzen (2006) also examined the effect of delays in MTS procedures across MTO and OTM arrangements. That study reported that participants who experience a MTO arrangement were more likely to achieve equivalence as the delay increased, while the participants who completed OTM training all demonstrated equivalence regardless of the delay used.

As outlined above, much of the research in this area using MTS has identified MTO as the most effective arrangement to facilitate the formation of equivalence relations. However, not all studies have agreed with these findings. Some studies have found that no difference in the effectiveness of the MTO and OTM arrangements (Smeets & Barnes-Holmes, 2005), while others have found the OTM arrangement to be the most effective (Arntzen, 2006; Arntzen & Holth, 1997). It has generally been concluded that both the MTO and OTM arrangements are more effective than the linear arrangement (Arntzen & Holth, 1997).

The use of MTO and OTM procedures has also been examined in studies that used SPO procedures. The effect of the stimuli arrangement on performance on an SPO equivalence task was examined by Smeets et al. (1997). This study reported conflicting findings across four experiments. Experiment 1 of their study found the MTO arrangement resulted in greater performance than the OTM procedure. However, Experiments 2 and 3 of that study showed little difference in effectiveness between the three procedures as demonstrated by either poor performance by all groups, or good performance by all groups respectively. The removal of tests for the trained relations in Experiment 4 resulted in poorer performance on the linear and MTO procedures compared to the OTM procedure. Smeets et al. (1997) suggest that the MTO and OTM procedures are not equally effective at facilitating the formation of equivalence classes and that their conflicting results indicate a need for further study in this area. However, a study by Leader et al. (2000) reported little difference in performance by children on equivalence tasks that used linear, MTO and OTM procedures when they were used in conjunction with a training procedure that tests for the symmetry and equivalence relations separately (known as a simple to complex procedure).

In summary, most research has found the MTO arrangements to be the most effective, and it is generally agreed that both the MTO and OTM arrangements are more effective than the linear arrangement at facilitating the formation of equivalence relations. As the experiments in this study, thus far, have used a linear arrangement it is possible that this is, at least in part, responsible for the failure of most participants to demonstrate equivalence in either the SPO or MTS procedures. Therefore, Experiment 5 examined the effectiveness of the MTO and OTM arrangements in an equivalence task. The tested relations used were the symmetry and combined (C-B) symmetry/transitivity (equivalence) relations. These tests were chosen to provide consistency with the tested relations in the previous experiments. The inclusion of the linear procedure was not required here as, in all but the stimulus arrangement, the procedure was identical to that completed by the participants in Experiment 4 who received the less specific instructions. Therefore, the results for those participants could be compared directly to the results for the participants here who experienced the MTO or OTM stimulus arrangement. The number of training trials, and all other procedural aspects were the same as in Experiment 4 to allow this comparison.

Method

Participants, Ethics, Apparatus and Setting

Participant recruitment, ethics procedures, apparatus and setting were identical to those used in Experiments 3 and 4. This experiment involved 12 participants, three in each of four groups (E5.SPO.MTO, E5.MTS.MTO, E5.SPO.OTM, and E5.MTS.OTM). Group membership was determined quasi-randomly on the order of recruitment and is shown in Table 5.1.

Stimuli

The stimuli used in both the SPO and MTS procedures were the same nine nonsense syllables that were used in Experiments 2 and 4.

Table 5.1

Group assignment, procedure (SPO or MTS) and stimulus arrangement (MTO or OTM) experienced by each participant in Experiment 5.

Participant	Group	Procedure	Arrangment
P5.1	E5.SPO.MTO	SPO	MTO
P5.2	E5.SPO.MTO	SPO	MTO
P5.3	E5.SPO.MTO	SPO	MTO
P5.4	E5.MTS.MTO	MTS	MTO
P5.5	E5.MTS.MTO	MTS	MTO
P5.6	E5.MTS.MTO	MTS	MTO
P5.7	E5.SPO.OTM	SPO	OTM
P5.8	E5.SPO.OTM	SPO	OTM
P5.9	E5.SPO.OTM	SPO	OTM
P5.10	E5.MTS.OTM	MTS	OTM
P5.11	E5.MTS.OTM	MTS	OTM
P5.12	E5.MTS.OTM	MTS	OTM

Procedure

This procedure was nearly identical to that of Experiment 4, however the stimuli pairs used in training (and subsequent testing) were ordered in either a MTO or OTM arrangement. The trained and tested stimuli pairs for each training procedure and stimuli arrangement are shown in Table 5.2. The instructions provided to the participants in the present experiment were the same as those used in Experiments 1 to3, and were the ‘less-specific’ instructions used in Experiment 4.

Session criterion

The session criterion was the same as used in Experiment 4.

Table 5.2

Trained and tested relations in the linear (Experiment 4), MTO and OTM (Experiment 5) arrangements.

SPO Training			MTS training		
Linear	MTO	OTM	Linear	MTO	OTM
A1-B1	B1-A1	A1-B1	A1-B2	B2-A1	A1-B2
A2-B2	B2-A2	A2-B2	A2-B3	B3-A2	A2-B3
A3-B3	B3-A3	A3-B3	A3-B1	B1-A3	A3-B1
B1-C1	C1-A1	A1-C1	B1-C2	C3-A1	A1-C3
B2-C2	C2-A2	A2-C2	B2-C3	C1-A2	A2-C1
B3-C3	C3-A3	A3-C3	B3-C1	C2-A3	A3-C2

SPO Testing			MTS testing		
Linear	MTO	OTM	Linear	MTO	OTM
Symmetry			Symmetry		
B1-A1	A1-B1	B1-A1	B1-A3	A1-B2	B1-A3
B2-A2	A2-B2	B2-A2	B2-A1	A2-B3	B2-A1
B3-A3	A3-B3	B3-A3	B3-A2	A3-B1	B3-A2
C1-B1	A1-C1	C1-A1	C1-B3	A1-C3	C1-A2
C2-B2	A2-B2	C2-A2	C2-B1	A2-C1	C2-A3
C3-B3	A3-B3	C3-A3	C3-B2	A3-C2	C3-A1
Equivalence			Equivalence		
C1-A1	C1-B1	C1-B1	C1-A2	C1-B3	C1-B3
C2-B2	C1-B2	C2-B2	C2-A3	C2-B1	C2-B1
C3-B3	C1-B3	C3-B3	C3-A1	C3-B2	C3-B2

Results

Number of sessions

Nine of the 12 participants met the session criterion. These included two participants in each of the groups who experienced the MTO arrangement (E5.SPO.MTO and E5.MTS.MTO). The participants in these two groups all met the session criterion in their second training and testing cycle. The session criterion was also met by two participants who completed SPO training with a OTM arrangement in their third training and testing cycle. All of the participants who completed MTS with a OTM arrangement met the session criterion, and did so in 2 to 7 training and testing cycles.

Testing

Number correct on each tested relation and overall percent correct. Tables 5.3-5.6 show the number of correct responses made on each tested relation and the overall percentage of correct responses made on the symmetry and equivalence relations by each participant. As shown in Tables 5.3-5.6, the number of relations on which nine or more responses were correct and the percentage of correct responses on the symmetry and equivalence relations increased across the training and testing cycles, with most participants meeting the session criterion prior to the end of the eighth cycle. Two of the three participants (P5.2 and P5.9) who did not meet the criterion showed no clear trend in the number of trials answered correctly for each relation or percentage of correct response on the symmetry and equivalence relations across the sessions. The remaining participant (P5.6) was very close to meeting the session criterion in their final training and testing cycle. One participant (P5.9) failed to achieve nine or more responses on any tested relation during any training and testing cycle.

Figure 5.1 shows the percentage of correct responses achieved on the symmetry and equivalence relations during each training and testing cycle for all participants. Of the participants who failed to meet the session criterion, the data for two (P5.6 in E5.MTS.MTO and E5.SPO.OTM) show no clear difference in percent correct on the symmetry and equivalence relations during each cycle. The remaining participant (P5.2 in E5.SPO.MTO) achieved greater percentages of correct responses on the symmetry than the equivalence relations during most training and testing cycles.

Table 5.3

Number of correct responses for each tested relation of each training and testing cycle in each experimental session for the participants who completed SPO training with a MTO arrangement (E5.SPO.MTO)

Participant	Session	Condition	Tested relations - sample and correct comparison stimuli											Overall % Correct	Session criterion met
			Symmetry relations						% Correct Symmetry	Equivalence relations			% Correct Equivalence		
			A1-B1	A2-B2	A3-B3	A1-C1	A2-C2	C3-C3		C1-B1	C2-B2	C3-B3			
P5.1	Session 1	Cycle 1	1	8	**9**	6	7	**10**	68.33	1	5	4	33.33	56.67	N
		Cycle 2	**9**	**10**	**10**	**10**	**10**	**10**	**98.33**	**10**	**10**	**10**	**100.00**	**98.89**	**Y**
	Session 2	Cycle 3	–	–	–	–	–	–	–	–	–	–	–	–	–
		Cycle 4	–	–	–	–	–	–	–	–	–	–	–	–	–
	Session 3	Cycle 5	–	–	–	–	–	–	–	–	–	–	–	–	–
		Cycle 6	–	–	–	–	–	–	–	–	–	–	–	–	–
	Session 4	Cycle 7	–	–	–	–	–	–	–	–	–	–	–	–	–
		Cycle 8	–	–	–	–	–	–	–	–	–	–	–	–	–
P5.2	Session 1	Cycle 1	0	5	2	6	5	1	31.67	3	4	4	36.67	33.33	N
		Cycle 2	1	**9**	2	**10**	0	7	48.33	1	4	4	30.00	42.22	N
	Session 2	Cycle 3	3	**9**	3	**10**	2	7	56.67	2	6	2	33.33	48.89	N
		Cycle 4	1	8	4	**10**	**9**	**10**	70.00	1	5	1	23.33	54.44	N
	Session 3	Cycle 5	2	8	3	**10**	7	**10**	66.67	3	2	1	20.00	51.11	N
		Cycle 6	0	7	1	**10**	7	**10**	58.33	0	5	2	23.33	46.67	N
	Session 4	Cycle 7	1	5	3	**10**	6	**10**	58.33	3	4	3	33.33	50.00	N
		Cycle 8	1	3	2	**10**	4	**10**	50.00	1	0	0	3.33	34.44	N
P5.3	Session 1	Cycle 1	1	1	3	**10**	**9**	8	53.33	6	8	2	53.33	53.33	N
		Cycle 2	**10**	**10**	**9**	**10**	**10**	**10**	**98.33**	**10**	**10**	**10**	**100.00**	**98.89**	**Y**
	Session 2	Cycle 3	–	–	–	–	–	–	–	–	–	–	–	–	–
		Cycle 4	–	–	–	–	–	–	–	–	–	–	–	–	–
	Session 3	Cycle 5	–	–	–	–	–	–	–	–	–	–	–	–	–
		Cycle 6	–	–	–	–	–	–	–	–	–	–	–	–	–
	Session 4	Cycle 7	–	–	–	–	–	–	–	–	–	–	–	–	–
		Cycle 8	–	–	–	–	–	–	–	–	–	–	–	–	–

Table 5.4

Number of correct responses for each tested relation of each training and testing cycle in each experimental session for the participants who completed MTS training with a MTO arrangement (E5.MTS.MTO)

Participant	Session	Condition	Tested relations - sample and correct comparison stimuli											Overall % Correct	Session criterion met
			Symmetry relations						% Correct Symmetry	Equivalence relations			% Correct Equivalence		
			A1-B2	A2-B3	A3-B1	A1-C3	A2-C1	A3-C2		C1-B3	C2-B1	C3-B2			
P5.4	Session 1	Cycle 1	**10**	0	**10**	**10**	**10**	**10**	83.33	1	**10**	4	50.00	72.22	N
		Cycle 2	**10**	**10**	**10**	**10**	**10**	**10**	**100.00**	**10**	**10**	**10**	**100.00**	**100.00**	**Y**
	Session 2	Cycle 3	–	–	–	–	–	–	–	–	–	–	–	–	–
		Cycle 4	–	–	–	–	–	–	–	–	–	–	–	–	–
	Session 3	Cycle 5	–	–	–	–	–	–	–	–	–	–	–	–	–
		Cycle 6	–	–	–	–	–	–	–	–	–	–	–	–	–
	Session 4	Cycle 7	–	–	–	–	–	–	–	–	–	–	–	–	–
		Cycle 8	–	–	–	–	–	–	–	–	–	–	–	–	–
P5.5	Session 1	Cycle 1	**10**	**9**	7	3	**10**	5	73.33	5	6	4	50.00	65.56	N
		Cycle 2	**10**	**10**	**10**	**10**	**10**	**10**	**100.00**	**10**	**9**	**10**	**96.67**	**98.89**	**Y**
	Session 2	Cycle 3	–	–	–	–	–	–	–	–	–	–	–	–	–
		Cycle 4	–	–	–	–	–	–	–	–	–	–	–	–	–
	Session 3	Cycle 5	–	–	–	–	–	–	–	–	–	–	–	–	–
		Cycle 6	–	–	–	–	–	–	–	–	–	–	–	–	–
	Session 4	Cycle 7	–	–	–	–	–	–	–	–	–	–	–	–	–
		Cycle 8	–	–	–	–	–	–	–	–	–	–	–	–	–
P5.6	Session 1	Cycle 1	2	7	3	0	0	**9**	35.00	7	8	0	50.00	40.00	N
		Cycle 2	0	10	1	0	0	**10**	35.00	6	**9**	0	50.00	40.00	N
	Session 2	Cycle 3	8	8	0	0	4	**10**	50.00	6	2	2	33.33	44.44	N
		Cycle 4	7	**9**	**10**	6	5	**10**	78.33	7	**10**	2	63.33	73.33	N
	Session 3	Cycle 5	6	7	**10**	7	7	**10**	78.33	7	**10**	2	63.33	73.33	N
		Cycle 6	7	**10**	**10**	8	5	**10**	83.33	7	**10**	0	56.67	74.44	N
	Session 4	Cycle 7	4	**9**	**10**	3	**10**	**10**	76.67	**9**	7	**10**	86.67	80.00	N
		Cycle 8	**10**	**10**	**10**	8	**10**	**10**	96.67	**10**	**10**	**10**	**100.00**	97.78	N

Table 5.5

Number of correct responses for each tested relation of each training and testing cycle in each experimental session for the participants who completed SPO training with a OTM arrangement (E5.SPO.OTM)

Participant	Session	Condition	Tested relations - sample and correct comparison stimuli											Overall % Correct	Session criterion met
			Symmetry relations						% Correct Symmetry	Equivalence relations			% Correct Equivalence		
			A1-B2	A2-B3	A3-B1	A1-C3	A2-C1	A3-C2		C1-B3	C2-B1	C3-B2			
P5.7	Session 1	Cycle 1	5	**9**	6	1	5	1	45.00	0	**9**	2	36.67	42.22	N
		Cycle 2	6	**10**	**9**	**10**	**10**	**10**	91.67	6	**10**	5	**70.00**	**84.44**	N
	Session 2	Cycle 3	**10**	**10**	**10**	**9**	**10**	**10**	**98.33**	**10**	**10**	**10**	**100.00**	**98.89**	**Y**
		Cycle 4	–	–	–	–	–	–	–	–	–	–	–	–	–
	Session 3	Cycle 5	–	–	–	–	–	–	–	–	–	–	–	–	–
		Cycle 6	–	–	–	–	–	–	–	–	–	–	–	–	–
	Session 4	Cycle 7	–	–	–	–	–	–	–	–	–	–	–	–	–
		Cycle 8	–	–	–	–	–	–	–	–	–	–	–	–	–
P5.8	Session 1	Cycle 1	2	4	2	2	3	2	25.00	6	5	5	53.33	34.44	N
		Cycle 2	8	**10**	**10**	**10**	**10**	**10**	**96.67**	**10**	**9**	**9**	**93.33**	**95.56**	N
	Session 2	Cycle 3	**10**	**10**	**10**	**10**	**10**	**10**	**100.00**	**10**	**10**	**10**	**100.00**	**100.00**	**Y**
		Cycle 4	–	–	–	–	–	–	–	–	–	–	–	–	–
	Session 3	Cycle 5	–	–	–	–	–	–	–	–	–	–	–	–	–
		Cycle 6	–	–	–	–	–	–	–	–	–	–	–	–	–
	Session 4	Cycle 7	–	–	–	–	–	–	–	–	–	–	–	–	–
		Cycle 8	–	–	–	–	–	–	–	–	–	–	–	–	–
P5.9	Session 1	Cycle 1	3	4	9	6	3	5	50.00	5	4	5	46.67	48.89	N
		Cycle 2	5	5	7	4	7	5	55.00	4	3	3	33.33	47.78	N
	Session 2	Cycle 3	3	1	2	4	4	5	31.67	4	4	2	33.33	32.22	N
		Cycle 4	2	5	5	4	4	2	36.67	4	1	5	33.33	35.56	N
	Session 3	Cycle 5	4	2	2	3	2	3	26.67	4	2	1	23.33	25.56	N
		Cycle 6	3	2	1	2	5	1	23.33	2	7	4	43.33	30.00	N
	Session 4	Cycle 7	3	3	3	1	1	5	26.67	5	6	2	43.33	32.22	N
		Cycle 8	5	3	2	1	6	1	30.00	4	4	2	33.33	31.11	N

Table 5.6

Number of correct responses for each tested relation of each training and testing cycle in each experimental session for the participants who completed MTS training with a OTM arrangement (E5.MTS.OTM)

Participant	Session	Condition	Tested relations - sample and correct comparison stimuli											Overall % Correct	Session criterion met
			Symmetry relations						% Correct Symmetry	Equivalence relations			% Correct Equivalence		
			B1-A3	B2-A1	B3-A2	C1-A2	C2-A3	C3-A1		C1-B3	C2-B1	C3-B2			
P5.10	Session 1	Cycle 1	3	1	2	6	2	3	28.33	5	2	3	33.33	30.00	N
		Cycle 2	5	2	4	5	5	3	40.00	4	2	5	36.67	38.89	N
	Session 2	Cycle 3	5	6	5	4	3	3	43.33	4	4	3	36.67	41.11	N
		Cycle 4	4	4	3	2	5	3	35.00	4	1	4	30.00	33.33	N
	Session 3	Cycle 5	6	7	3	6	6	8	60.00	6	6	6	60.00	60.00	N
		Cycle 6	6	**9**	3	6	8	5	61.67	6	5	5	53.33	58.89	N
	Session 4	Cycle 7	**10**	**10**	**10**	**10**	**9**	**10**	**98.33**	**10**	**9**	**9**	**93.33**	**96.67**	**Y**
		Cycle 8	–	–	–	–	–	–	–	–	–	–	–	–	–
P5.11	Session 1	Cycle 1	6	0	1	5	0	0	20.00	2	1	5	26.67	22.22	N
		Cycle 2	0	0	0	0	0	2	3.33	0	1	0	3.33	3.33	N
	Session 2	Cycle 3	3	2	5	3	4	1	30.00	1	4	0	16.67	25.56	N
		Cycle 4	6	2	3	2	1	3	28.33	6	5	2	43.33	33.33	N
	Session 3	Cycle 5	4	5	6	3	2	3	38.33	3	5	5	43.33	40.00	N
		Cycle 6	3	2	7	5	4	1	36.67	5	3	3	36.67	36.67	N
	Session 4	Cycle 7	**10**	**10**	**10**	**10**	**10**	**10**	**100.00**	**10**	**10**	**10**	**100.00**	**100.00**	**Y**
		Cycle 8	–	–	–	–	–	–	–	–	–	–	–	–	–
P5.12	Session 1	Cycle 1	**10**	**10**	**10**	**10**	**9**	**10**	**98.33**	**9**	8	**10**	90.00	95.56	N
		Cycle 2	**10**	**10**	**10**	**10**	**10**	**10**	**100.00**	**10**	**10**	**10**	**100.00**	**100.00**	**Y**
	Session 2	Cycle 3	–	–	–	–	–	–	–	–	–	–	–	–	–
		Cycle 4	–	–	–	–	–	–	–	–	–	–	–	–	–
	Session 3	Cycle 5	–	–	–	–	–	–	–	–	–	–	–	–	–
		Cycle 6	–	–	–	–	–	–	–	–	–	–	–	–	–
	Session 4	Cycle 7	–	–	–	–	–	–	–	–	–	–	–	–	–
		Cycle 8	–	–	–	–	–	–	–	–	–	–	–	–	–

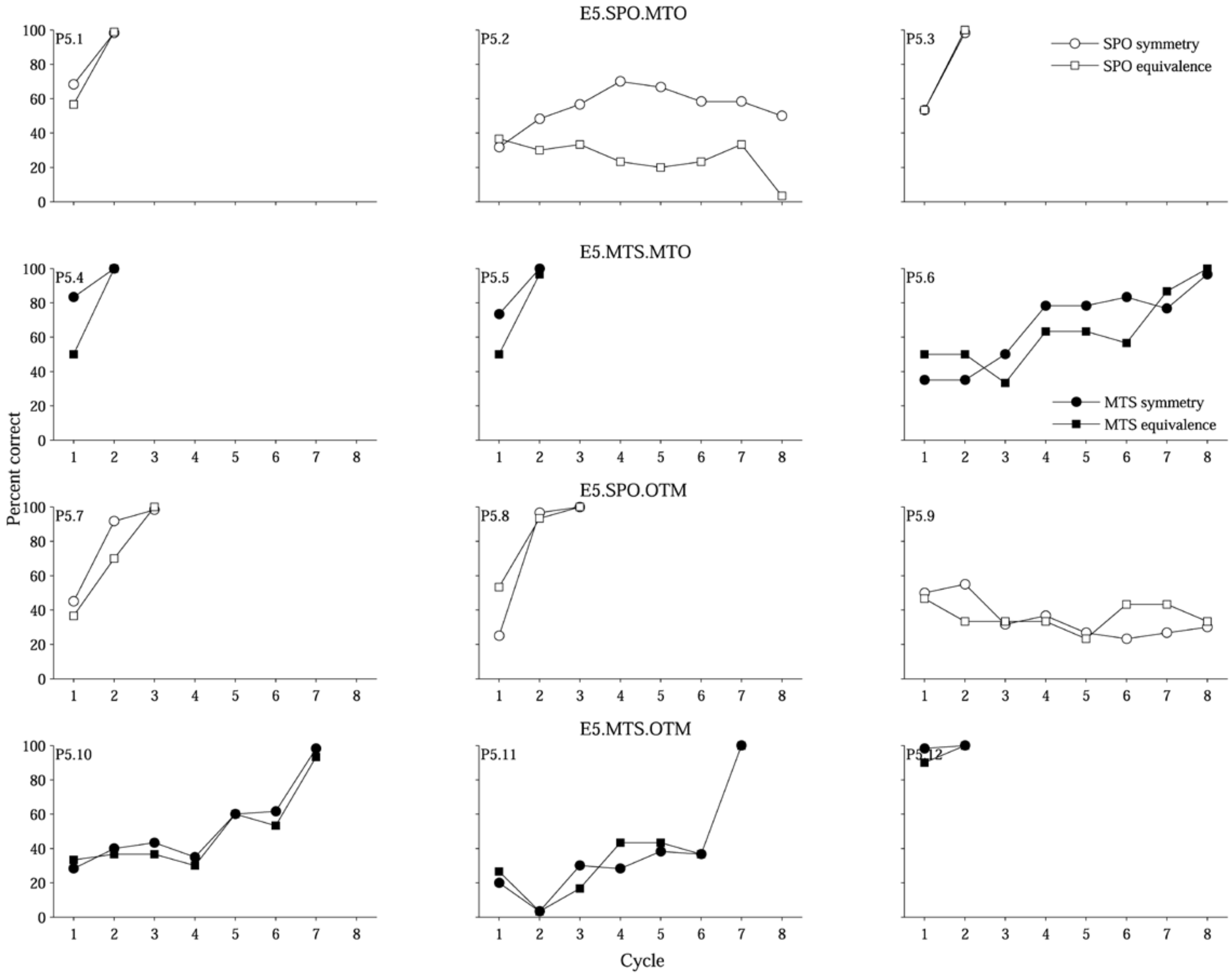

Figure 5.1 Percent correct on the symmetry and equivalence relations following SPO and MTS training across all sessions for all participants in Experiment 5.

A factorial repeated-measures ANOVA was conducted to compare the percentage of correct responses made by each participant in the present experiment and the participants in Experiment 4 who were given the standard instructions during testing of their final training and testing cycle (E4.SPO and E4.MTS). The inclusion of the participants in Groups E4.SPO and E4.MTS allowed a comparison between the linear, MTO and OTM procedures. The results of this ANOVA are given in Table 5.7. The significant within-subjects effect of the type of relation shows a greater percentage of correct responses on the symmetry than equivalence relations. This result had a moderate effect size (0.557). There was a significant within-subjects interaction between the type of relation and the stimuli arrangement which also showed a moderate effect size (0.602).

Table 5.7

Results of an ANOVA to compare percent correct on the symmetry and equivalence relations of the final session for both procedures for all groups in Experiment 5, and the groups who received the less-specific instructions with a linear arrangement in Experiment 4.

Source	*df*	*F*	$\eta^2_{partial}$
Within-subjects effects			
Type of relation	1,12	15.115*	0.557
Type of relation x Procedure	1,12	0.324	0.026
Type of relation x Stimuli Arrangement (Linear/MTO/OTM)	1,12	9.066*	0.602
Type of relation x Procedure x Stimuli Arrangement	1,12	0.452	0.070
Between-subjects effects			
Procedure	1,12	0.666	0.053
Stimuli Arrangement	1,12	1.912	0.242
Procedure x Stimuli Arrangement	1,12	1.089	0.154

*=significant at $p<0.05$

Correct vs. Incorrect (compared to chance) during testing. As in Experiments 1-4, χ^2 tests were conducted to compare the number of correct and incorrect responses made by each participant to a response distribution predicted by chance. Table 5.8 shows the number of correct and incorrect responses made during each training and testing cycle by each participant, and the χ^2 statistic for each test. As shown in Table 5.7, most participants were making a greater number of correct responses than predicted by chance by the end of their final training and testing cycle. The exception to this was P.5.2.

MTS training

Number correct on each tested relation and overall percent correct. Table 5.9 and Figure 5.2 show the number of trials on which correct responses were made for each trained relation by all participants. As shown in Table 5.7 and Figure 5.2, all but one (P5.6 in Group E5.MTS.MTO) achieved nine or more correct on all trained relations by the end of the final session.

Correct vs. incorrect (compared to chance) during MTS training. χ^2 tests were conducted to compare the distribution of correct and incorrect responses to that predicted by chance. The distribution predicted by chance was the same as in Experiments 1 to 4. Table 5.10 shows the number of correct and incorrect responses made by each participant during each training and testing cycle, and the χ^2 statistic for each test. All participants achieved a greater number of correct responses than predicted by chance in testing of most training and testing cycles. By the end of the final session the number of incorrect responses was few, or none, for all participants.

Table 5.8

Chi-square tests for the distribution of correct (Corr.) and incorrect (Incor.) responses (compared to responses expected by chance (correct (30/90), incorrect (60/90)) for all participants in Experiment 5.

Group	Participant	Session 1						Session 2						Session 3						Session 4					
		Cycle 1			Cycle 2			Cycle 3			Cycle 4			Cycle 5			Cycle 6			Cycle 7			Cycle 8		
		Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2
E5.SPO.MTO	P5.1	51	39	22.050*	89	1	174.050*	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_
E5.SPO.MTO	P5.2	30	60	0.000	38	52	3.200	44	45	10.283*	49	41	18.050*	46	44	12.800*	42	48	7.200*	45	45	11.250*	31	59	0.050
E5.SPO.MTO	P5.3	48	42	16.200*	89	1	174.050*	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_
E5.MTS.MTO	P5.4	65	25	61.250*	90	0	180.000*	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_
E5.MTS.MTO	P5.5	59	31	42.050*	89	1	174.050*	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_
E5.MTS.MTO	P5.6	36	54	1.800	36	54	1.800	40	50	5.000*	66	24	64.800*	66	24	64.800*	67	23	68.450*	72	18	88.200*	88	2	168.200*
E5.SPO.OTM	P5.7	38	52	3.200*	76	14	105.800*	89	1	174.050*	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_
E5.SPO.OTM	P5.8	31	59	0.050*	86	4	156.800*	90	0	180.000*	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_
E5.SPO.OTM	P5.9	44	46	9.800*	43	47	8.4500*	29	61	0.050	32	58	0.200	23	67	2.450	27	63	0.450	29	61	0.050	28	62	0.200
E5.MTS.OTM	P5.10	27	63	0.450	35	55	1.250	37	53	2.450	30	60	0.000	54	36	28.800*	53	37	26.450*	87	3	162.450*	_	_	_
E5.MTS.OTM	P5.11	20	70	5.000*	3	87	36.45*	23	67	2.450*	30	60	0.000	36	54	1.800	33	57	0.450	90	0	180.000*	_	_	_
E5.MTS.OTM	P5.12	86	4	156.800*	90	0	180.000*	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_

*=significant at $p < .05$

Table 5.9

Number of correct responses for each trial type during MTS training of each experimental session for all participants in Groups E5.MTS.MTO and E5.MTS.OTM

Group	Participant	Session	Cycle	Sample and correct comparison stimuli						Total % correct
				A1-B2	A2-B3	A3-B1	B1C2	B2-C3	B3-C1	
E5.MTS.MTO	P5.4	Session 1	Cycle 1	**9**	3	1	**4**	**10**	6	55.00
			Cycle 2	**10**	**10**	**10**	**10**	**10**	**10**	**100.00**
		Session 2	Cycle 3	–	–	–	–	–	–	–
			Cycle 4	–	–	–	–	–	–	–
		Session 3	Cycle 5	–	–	–	–	–	–	–
			Cycle 6	–	–	–	–	–	–	–
		Session 4	Cycle 7	–	–	–	–	–	–	–
			Cycle 8	–	–	–	–	–	–	–
E5.MTS.MTO	P5.5	Session 1	Cycle 1	6	4	4	9	6	4	55.00
			Cycle 2	**10**	**10**	**10**	**10**	**9**	8	95.00
		Session 2	Cycle 3	–	–	–	–	–	–	–
			Cycle 4	–	–	–	–	–	–	–
		Session 3	Cycle 5	–	–	–	–	–	–	–
			Cycle 6	–	–	–	–	–	–	–
		Session 4	Cycle 7	–	–	–	–	–	–	–
			Cycle 8	–	–	–	–	–	–	–
E5.MTS.MTO	P5.6	Session 1	Cycle 1	5	2	3	2	2	5	31.67
			Cycle 2	1	8	3	3	9	8	53.33
		Session 2	Cycle 3	3	8	4	5	**10**	7	61.67
			Cycle 4	6	3	6	6	**10**	5	60.00
		Session 3	Cycle 5	9	9	8	4	**10**	**9**	81.67
			Cycle 6	8	**10**	**9**	**9**	**10**	8	90.00
		Session 4	Cycle 7	8	**10**	10	**9**	8	**10**	91.67
			Cycle 8	8	**10**	**9**	7	**10**	**9**	88.33
E5.MTS.OTM	P5.10	Session 1	Cycle 1	2	2	3	2	3	1	21.67
			Cycle 2	1	5	0	**4**	3	3	26.67
		Session 2	Cycle 3	6	5	4	**10**	4	2	51.67
			Cycle 4	7	**9**	**9**	**10**	**9**	**10**	90.00
		Session 3	Cycle 5	7	7	6	**10**	7	8	75.00
			Cycle 6	**10**	6	6	**9**	**9**	8	80.00
		Session 4	Cycle 7	**9**	**9**	**10**	**10**	**10**	**10**	**96.67**
			Cycle 8	–	–	–	–	–	–	–
E5.MTS.OTM	P5.11	Session 1	Cycle 1	2	5	1	4	3	**4**	31.67
			Cycle 2	**10**	**9**	**9**	8	1	10	78.33
		Session 2	Cycle 3	4	5	6	7	2	6	50.00
			Cycle 4	4	7	7	**9**	5	4	60.00
		Session 3	Cycle 5	4	4	6	4	6	5	48.33
			Cycle 6	2	8	6	4	3	5	46.67
		Session 4	Cycle 7	**10**	**9**	**10**	**10**	**9**	**10**	**96.67**
			Cycle 8	–	–	–	–	–	–	–

Table 5.9 continued.

Group	Participant	Session	Cycle	Sample and correct comparison stimuli						Total % correct
				A1-B2	A2-B3	A3-B1	B1C2	B2-C3	B3-C1	
E5.MTS.OTM	P5.12	Session 1	Cycle 1	8	6	5	6	5	7	61.67
			Cycle 2	**10**	**9**	**10**	**10**	**10**	**10**	**98.33**
		Session 2	Cycle 3	–	–	–	–	–	–	–
			Cycle 4	–	–	–	–	–	–	–
		Session 3	Cycle 5	–	–	–	–	–	–	–
			Cycle 6	–	–	–	–	–	–	–
		Session 4	Cycle 7	–	–	–	–	–	–	–
			Cycle 8	–	–	–	–	–	–	–

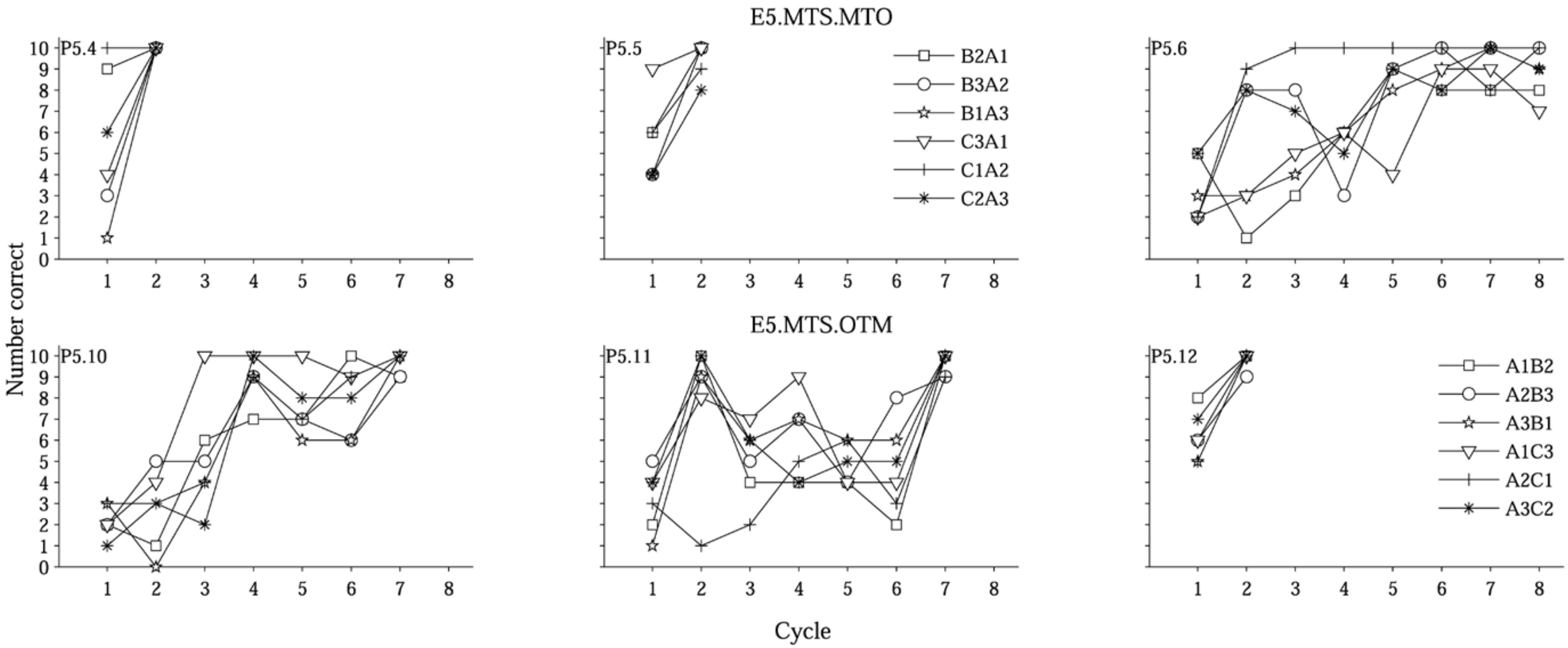

Figure 5.2 Number of correct responses made on each trained relation during MTS training across all sessions for all participants in Experiment 5.

Table 5.10

Chi-square tests for the distribution of correct (Corr.) and incorrect (Incor.) responses during MTS training (compared to responses expected by chance (correct (20/60), incorrect (40/60)) for all participants in E5.MTS.MTO and E5.MTS.OTM.

Group	Participant	Session 1						Session 2						Session 3						Session 4					
		Cycle 1			Cycle 2			Cycle 3			Cycle 4			Cycle 5			Cycle 6			Cycle 7			Cycle 8		
		Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2	Corr.	Incor.	χ^2
E5.MTS.MTO	P5.4	33	27	12.675*	60	0	120.000*	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_
E5.MTS.MTO	P5.5	33	27	12.675*	57	3	102.675*	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_
E5.MTS.MTO	P5.6	19	41	0.075	32	28	10.800*	37	23	21.675*	36	24	19.200*	49	11	63.075*	54	6	86.700*	55	5	91.875*	53	7	81.675*
E5.MTS.OTM	P5.10	13	47	3.675	16	44	1.200	31	29	9.075*	54	6	86.700*	45	15	46.875*	48	12	58.800*	58	2	108.300*	_	_	_
E5.MTS.OTM	P5.11	19	41	0.075	47	13	54.675*	30	30	7.500*	36	24	19.200*	29	31	6.075*	28	32	4.800*	58	2	108.300*	_	_	_
E5.MTS.OTM	P5.12	37	23	21.675*	59	1	114.075*	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_

*=significant at p <.05

Discussion

The present experiment compared the effectiveness of the MTO and OTM stimulus arrangements in facilitating the formation of equivalence relations. As all other procedural factors were held constant, the results from these groups could be compared directly to those of the participants in Experiment 4 who experienced a linear stimuli arrangement and received the 'less-specific' instructions. In the present experiment, three quarters of the participants achieved the session criterion during testing. Therefore, a greater proportion of participants demonstrated equivalence in the present study than in any of the previous experiments.

As in the previous experiments, the findings of the present experiment and for the participants in Experiment 4 who received the 'less-specific' instructions show no difference in the effectiveness of the SPO and MTS procedures, and a significantly better performance on the symmetry ($\bar{x}$ = 86.49%, *SD* = 22.01%) than the equivalence relations ($\bar{x}$ = 71.30%, *SD* = 35.41%), with a moderate effect size. However, there was also a significant interaction between the type of relation (symmetry or equivalence) and the stimuli arrangement (Linear, MTO, or OTM). This interaction showed a moderate effect size by Ferguson's (2009) conventions. On closer inspection this interaction revealed that the difference between accuracy on the symmetry and equivalence relations was greater when the linear arrangement was used than when the MTO and OTM arrangements were used. It appears, then, that the significant result obtained for the main within-subject effect of the type of relation is due to the differential performance on the symmetry and equivalence relations by the participants who experienced the linear arrangement.

Better performance on the symmetry than equivalence relations was also seen in the studies by Arntzen and Holth (1997) and Hove (2003). Most of the participants in Arntzen and Holth's (1997) study who completed both the symmetry and equivalence tests performed best on the tests for symmetry. As the other groups in that study did not complete the symmetry tests, differential performance on the symmetry and equivalence relations within that study cannot be assessed. However, the data for the participants who experienced a linear arrangement and the symmetry tests show similarities to the findings of Experiments 1 to 4 of this study where performance on the symmetry relations was better than on the equivalence relations for most participants. Hove's (2003) participants experienced either MTO or OTM arrangements, suggesting that uneven performance on symmetry and equivalence relations is not an effect that is specific to the linear arrangement.

However, in the present study, the differential performance on the symmetry and equivalence relations was found for the linear arrangement only. Therefore, the MTO and OTM arrangements helped to improve performance on the equivalence relations relative to the previous experiments. This is further supported by an effect size greater than Ferguson's (2009) RMPE for the between-subjects effect of stimuli arrangements (Linear $\bar{x}$ = 61.94%; MTO $\bar{x}$ = 86.94%; OTM $\bar{x}$ = 87.78%). This finding is consistent with Fields et al.'s (1999) hypothesis that for studies involving small stimulus classes, and normally developed adults, the MTO and OTM procedures would be equally effective. However, this finding is contrary to Hove's (2003) study which used a similar participant group and found the MTO procedure to be most effective.

By Fergusson's (2009) conventions, there was also a small interaction effect (just above the RMPE) between the training procedure (SPO or MTS) and the stimuli arrangement. This interaction is the result of the greater percentage of correct responses made during testing following MTS than SPO training for the MTO and OTM arrangements, which was reversed when a linear arrangement was used. Therefore, it is possible that the MTS procedure may be more effective than the SPO procedure when a MTO or OTM arrangement is used. There is only small evidence for this effect here, but it does suggest an avenue for future research.

Most of the participants who met the session criterion did so within four training and testing cycles (240 training trials). This is equivalent to the total number of training trials that could be completed by the participants in Experiments 1to 3, where nearly all of the participants failed to meet the session criterion. The session criterion was also not met by any of the participants in Experiment 4 who received the less specific instructions. Therefore, those participants completed the maximum of 480 training trials across the eight training and testing cycles. The only procedural difference experienced by the participants in the present experiment and those who received the 'less-specific' instructions in Experiment 4 was the stimulus arrangement. Therefore, the MTO and OTM procedures were more likely to result in equivalence. Some of the participants in the present study, and the 'less-specific' instruction group of Experiment 4 who did not meet the session criterion did improve in overall accuracy across the training and testing cycles. This suggests that the addition of more training trials, or cycles, may have aided the development of equivalence relations for these participants.

In summary, the present study found the MTO and OTM arrangements to be equally effective, and both were more effective than the linear arrangement, as used in Experiments

1to 4, at facilitating the formation of equivalence relations. Most of the participants who demonstrated equivalence also did so within fewer trials than were completed by the participants in Experiments 1to 4. There is some evidence to suggest that increasing the number of training trials, or cycles, would have made equivalence formation more likely for some of the participants who failed to meet the session criterion, particularly when a linear procedure was used.

EXPERIMENT 6

As discussed previously, there is some evidence from Experiments 1 to 5 to suggest that the addition of more training trials or of more training and testing cycles might increase the likelihood of equivalence. In Experiments 1 to 3 there were 60 training trials, 10 for each of six trained relations, prior to testing in each training and testing cycle with both SPO and MTS training. This gave a maximum of 240 training trials or stimulus presentations in each training procedure across the sessions. In Experiments 4 and 5 of the present study the participants completed a maximum of 480 training trials or stimulus presentations with either SPO or MTS training. However, equivalence was demonstrated by few participants in Experiments 1 to 4 of the present study. In Experiment 1 it was concluded that as there was no clear trend in the data for the participants who failed to demonstrate equivalence there would be little benefit from increasing the number of training trials or stimulus presentations. However, the data for some participants in the later experiments of this study showed an increase in accuracy during testing across the sessions. For these participants, more training trials may have resulted in the demonstration of equivalence.

The findings of the experiments thus far raised questions regarding the effect of the number of training trials on the formation of equivalence classes. Most studies that use a MTS procedure required that the participants reach a criterion during training prior to completing tests for symmetry and equivalence; this is not possible with the SPO procedure as no responses are required. Using such a criterion means participants complete different numbers of training trials prior to the test. The number of trials each participant experienced is not always reported (e.g., Barnes-Holmes et al., 2000), but where it is reported it differs both within and across experiments. For example, Hayes, Thompson, and Hayes's (1989) participants required between 28 and 250 training trials to reach criterion during training, and all of these participants developed equivalence. The participants in Experiment 1 by Roche and Barnes (1997) required 54 and 174 training trials to reach the criterion within one or two training and testing cycles. All of those participants demonstrated equivalence. Most of Hayes et al.'s (1991) participants met the training criterion after 132 to 162 trials, and all but two then demonstrated equivalence. The two participants who failed the equivalence test then received further training, receiving up to 196 trials in total, and then they successfully completed the equivalence test. Holth and Arntzen's (1998) participants met the training criterion after between 90 and 334 trials. Of the 40 participants in that study, 32 demonstrated equivalence. The data presented suggest that for the group where nearly all failures of equivalence occurred, better performances followed greater numbers of training trials.

Markham et al.'s (2002) participants completed between 7 and 60 blocks of 18 trials (i.e., 126 to1080 trials) involving compound stimuli prior to meeting the training criterion and all then demonstrated equivalence. The examples given above demonstrate the wide range in the number of training trials required by participants to meet the training criterion under MTS procedures. The number of training trials in the present study fall within the ranges outlined above, however, much lower rates of success on the equivalence tests were found than reported by those studies. While the total number of training trials required across and within studies varies greatly, there is some suggestion that a larger numbers of training trials or re-exposure to training conditions following an initial failure to demonstrate equivalence (and therefore exposure to more training trials in total) may make the development of equivalence more likely.

The number of presentations of stimulus pairs that have been used also varies across SPO studies. For example, Leader et al. (1996) and Smeets et al. (1997) used 10 repetitions of each of the six trained relations (60 stimulus pairings) prior to each testing condition. This is the same number of pairings as used in Experiments 1 to 5 of the present study. Twenty-two of Leader et al.'s (1996) 35 participants demonstrated equivalence following between one and six exposures to training and testing (60 to 360 stimulus pairings in total). As did nine of Smeets et al.'s (1997) 10 adult participants. Smeets et al.'s (1997) participants who demonstrated equivalence completed between 60 to180 stimulus pairings (one to three blocks). The participant who did not demonstrate equivalence experienced 240 stimulus pairings (four blocks) in total. These studies report higher rates of equivalence than seen in the present study. Also using an SPO procedure, Layng and Chase (2001) used either 12 trials (2 repetitions of six relations) or 144 trials (12 repetitions of six relations) across different experimental conditions. Ten of the 13 participants in that study demonstrated equivalence after 72 to 480 trials. The remaining three participants had not done so after 984 to 1296 trials. In the experiments of that study, participants achieved equivalence within fewer stimulus presentations when all training and testing cycles involved 12 stimulus pairings than when the first cycle contained 144 stimulus pairings, suggesting that the repeated exposures to the MTS testing was necessary to facilitate the formation of equivalence classes. Fields et al.'s (1997) SPO study involved a criterion on yes or no responding and so the number of training trials varied across participants, with participants taking between 2 and 8 blocks of 16 trials to learn each trained relation. All of the participants met the training criterion and 10 of the 18 participants demonstrated equivalence during testing. These studies show that, as with MTS training, the number of trials used in SPO varies greatly across experiments. The number of training trials used in present study falls with this range but the study found much lower rates of equivalence than reported for these SPO procedures. It is not clear if more trials would have helped when using an SPO procedure.

The two studies that have compared the SPO and MTS procedures (Clayton & Hayes, 2004; Leader & Barnes-Holmes, 2001b) have also used differing numbers of training trials. Experiment 1 by Leader and Barnes-Holmes (2001b), which has formed the procedural basis for all of the experiments in the present study used 60 training trials or stimulus pairings prior to testing in each training and testing cycle. However, unlike the results found in this study, nearly all of Leader and Barnes-Holmes's (2001b) participants demonstrated equivalence following SPO training, and half demonstrated equivalence following MTS training. As mentioned previously, it is not clear why the results of the present study differ from those found by Leader and Barnes-Holmes (2001b). In Experiments 2 to 4 of that study, a criterion on MTS training performance was introduced, requiring 12 consecutive correct responses prior to testing in each training and testing cycle. In SPO training of those experiments, only 12 SPO trials were presented in each training and testing cycle. Therefore, unless the participants achieved correct responses on the first 12 MTS training trials, participants experienced more training trials during MTS than SPO training. However, across those experiments, more participants met the equivalence criterion following SPO than MTS training.

Clayton and Hayes's (2004) participants completed sets of 12 MTS training trials until they reached a criterion of 100% correct in a set. They were also presented with 12 SPO trials. This training was completed once prior to testing for symmetry and then again prior to testing for equivalence. Clayton and Hayes (2004) do not give the number of MTS training trials required prior to meeting the training criterion. However, as with Experiments 2 to 4 by Leader and Barnes-Holmes (2001b), unless the participants responded correctly on all of the first set of 12 trials, the number of trials completed would have been greater than the number of SPO presentations. In this study, the MTS procedure was shown to be more effective. As outlined in the *Introduction*, conflicts with the findings of Leader and Barnes-Holmes (2001b).

Thus, as with the studies that used either MTS or SPO training, the two studies that have compared these procedures do not provide any clear conclusion on the effect of the number of training trials on the likelihood of equivalence formation.

In summary, the studies outlined above do not indicate a clear conclusion on the effect of the number of training trials on equivalence class formation. However, there is some evidence to suggest that increasing the number of training trials would result in a greater number of participants reaching the equivalence criterion during testing. In most of those studies, more than half (and sometimes all) of the participants demonstrated equivalence. A similar rate of equivalence was seen here in Experiment 5 only. This finding suggested that the MTO and OTM procedures may make equivalence more likely compared to a linear procedure involving the same number of training trials or stimulus presentations. Of the three participants in Experiment 5 who failed to demonstrate equivalence, one had shown an increase in accuracy across the sessions, suggesting that exposure to more training trials would have resulted in this participant achieving equivalence. This response pattern was also noted for some participants in Experiments 2 to 4. Therefore, although it does not account for all failures, there appears to be a group of participants who would have been likely to demonstrate equivalence following further training. With the SPO procedure, the findings of Layng and Chase (2001) suggested that exposure to the MTS test was necessary for equivalence to develop, and increasing the number of trials prior to this test did not aid in the formation of equivalence relations.

Given the findings of the previous experiments in this study, and the great range in the number of training trials that have been used in previous studies, the present study investigated the effect of the number of training trials on equivalence formation across the SPO and MTS procedures, and across the three stimulus arrangements (linear, MTO, and OTM). Experiment 5 showed the MTO and OTM arrangements to result in more participants demonstrating equivalence than the linear arrangement. However, a linear arrangement was experienced by some participants here to allow comparison back to the previous experiments in this study. The participants in the present experiment experienced the same procedures and arrangements as the participants in Experiment 5, and the 'less-specific' instruction group from Experiment 4, however, the number of training trials per cycle was increased from 60 to 120. Therefore, the participants in Experiment 6 could complete a maximum of 960 trials in total across the training and testing cycles of the experiment. Increasing the number of trials per cycle allowed an examination of the effect of the number of trials per cycle and the number of cycles. Within the first four training and testing cycles completed in the present experiment, the participants completed the same number of trials as during eight training and testing cycles in previous experiments. Therefore, the effect of the same number of training trials across either four or eight training and testing cycles could be assessed. In all other respects, the experimental procedure was identical to that used in the previous experiments in this study. Maintaining a procedure consistent with the earlier experiments allowed an experimental examination of the effect of the number of training trials across both the procedure used (SPO or MTS) and the three stimulus arrangements (linear, MTO, or OTM).

METHOD

Participants, Ethics, Apparatus and Setting

The ethics procedures, participant recruitment, apparatus and setting were the same as those used in Experiments 3 to 5. This experiment involved 60 participants, 10 in each of six groups (E6.SPO.LIN, E6.MTS.LIN, E6.SPO.MTO, E6.MTS.MTO, E6.SPO.OTM, and E6.MTS.OTM). As with the previous experiments, group assignment was determined quasi-randomly on order of recruitment. Group membership is shown in Table 6.1.

Table 6.1

Group assignment, procedure (SPO or MTS) and stimulus arrangement (MTO or OTM) experienced by each participant in Experiment 6.

Participant	Group	Procedure	Arrangment
P6.1-6.10	E6.SPO.LIN	SPO	Linear
P6.11-6.20	E6.MTS.LIN	MTS	Linear
P6.21-6.30	E6.SPO.MTO	SPO	MTO
P6.31-6.40	E6.MTS.MTO	MTS	MTO
P6.41-6.50	E6.SPO.OTM	SPO	OTM
P6.51-6.60	E6.MTS.OTM	MTS	OTM

Stimuli

The stimuli used in all procedures were the same nine nonsense syllables that were used in Experiments 2 to 5 of the present study.

Procedure

The procedure used in this experiment was identical to that used in Experiment 5 for the participants who experienced the MTO and OTM arrangements. The linear arrangement was identical to that used in Experiment 4, with the less specific instructions. The only difference was that all participants in the present study completed 120 training trials prior to each testing condition.

Session criterion

The session criterion was the same as used in Experiments 4 and 5.

Results

Number of sessions

The session criterion was met by 42 of the 60 participants. When the participants are split into groups based on the training procedure experienced, 21 out of 30 participants in each group (SPO or MTS) achieved the session criterion. When these participants are organised by stimulus arrangement (20 each experienced linear, MTO, or OTM), equivalence was demonstrated by 11 participants who experienced the linear arrangement, 14 who experienced the MTO arrangement, and 17 who experienced the OTM arrangement. All but five of these 42 participants met the equivalence criterion within 4 training and testing cycles (480 trials).

Testing

Number correct on each tested relation and overall percent correct. The present experiment involved many more participants than did the previous experiments in this study. Therefore, unlike the previous experiments, the individual data across all sessions are not presented here. Tables and graphs consistent with the presentation of data in the previous experiments are included in Appendix D. The data presented here focus on performance during the final cycle completed by each participant in Experiment 6. Tables 6.2 - 6.4 show the number of cycles and training trials completed, and the number of correct responses on each tested relation and the overall number correct on the symmetry and equivalence relations during the final cycle for each participant. The results of a χ^2 goodness of fit tests comparing the distribution of correct and incorrect responses to that predicted by chance are also shown in these tables for the final cycle by the participants who did not meet the session criterion. As in previous experiments, numbers that appear in bold are indicative of nine or more correct responses on that relation. Tables 6.2 - 6.4, show that most participants performed well on the symmetry and equivalence tests and met the session criterion, thereby demonstrating equivalence. Figure 6.1 and Tables 6.2 and 6.4 show the percentage of correct responses made on the symmetry and equivalence trials during the final training and testing cycle by each of the participants. As shown in Figure 6.1, participants in Groups E6.SPO.LIN and E6.MTS.LIN who failed to meet the session generally performed better on the symmetry than the equivalence relations. Most participants in the other groups met the session criterion, however those participants who did not meet the criterion performed more similarly on the symmetry and equivalence relations.

A factorial repeated-measures ANOVA was conducted to compare the performance of each group on the symmetry and equivalence relations in the final training and testing cycle. The within-subjects factor was the type of relation (symmetry or equivalence). The between-subjects factors were the procedure (SPO or MTS) and the stimulus arrangement (linear, MTO, or OTM). The results of this are in Table 6.5. There was a significant within-subjects difference in performance on the symmetry ($\bar{x}$ = 88.36%) and equivalence ($\bar{x}$ = 82.11%) trials with an effect size that is below Ferguson's (2009) RMPE. None of the remaining within-subjects effects, or any of the between-subjects effects were significant, and all effect sizes were below the RMPE as defined by Ferguson (2009).

Correct vs. Incorrect (compared to chance) during testing. χ^2 tests for goodness of fit were conducted to compare the distributions of correct and incorrect responses to that predicted by chance. As most participants' performances met the session criterion, which requires a response distribution that shows few errors and so must be clearly different from that predicted by chance, only the response distributions for those participants who failed to meet the session criterion were assessed here. Tables 6.2- 6.4 show overall number of correct responses made during testing of the final training and testing cycle by all participants. The number of errors is not shown, however, the number of trials completed was always 90. The results of $\chi 2$ tests comparing the distribution of correct and incorrect responses to that predicted by chance (30/90 correct, 60/90 incorrect) are shown for the 18 participants who failed to meet the session criterion. These results show that 12 of these participants were making more correct responses than predicted by chance during testing of most cycles.

MTS training

Number correct on each trained relation and overall percent correct. Table 6.6 shows the number of correct responses (out of 20) made on each trained relation by the participants who completed MTS training (Groups E6.MTS.LIN, E6.MTS.MTO, and

E6.MTS.OTM). As in the previous experiments, no performance criterion was used during training. However, a criterion-level performance was said to have occurred a participant achieved 18 or more correct responses on each trained relation. This is the same percentage of correct responses as required to meet the session criterion during testing.

Table 6.6 shows that 12 of the 21 participants who met the session criterion during testing following MTS training had also achieved 18 or more (90%) correct responses on each trained relation. The remaining nine participants who met the session criterion during testing did so without achieving a criterion-level performance during training. Two response patterns emerged for the participants who did not meet the session criterion during testing. Seven of these participants (P6.13, P6.17, P6.18 and P6.20 in Group E6.MTS.LIN, P6.39 and P6.40 in Group E6.MTS.MTO, and P6.55 in Group E6.MTS.OTM) achieved 18 or more correct on all trained relations by the end of their final session, but did not meet the session criterion during testing. Two participants (P6.16 in Group E6.MTS.LIN and P6.56 in Group MTS.OTM) who failed to meet the session criterion had also not achieved a criterion-level performance during training.

Correct vs. incorrect (compared to chance) during MTS training. χ^2 goodness of fit tests were conducted to assess the distribution of correct and incorrect responses during training compared to that expected by chance (40/120 correct, 80/120 incorrect), for the participants who completed MTS training and the χ^2 statistics are given Table 6.6. All participants achieved a significantly greater number of correct responses than predicted by chance during training in the final training and testing cycle.

Table 6.2

Number of cycles and training trials completed, and number of correct responses for each tested relation of the final training and testing cycle completed by the participants who completed SPO (E6.SPO.LIN) or MTS (E6.MTS.LIN) training with a linear arrangement. Chi square results comparing the number of correct and incorrect responses to a distribution predicted by chance are shown for participants who did not meet the session criterion.

Participant	Number of cycles	Number of training trials	Tested relations - sample and correct comparison stimuli											Overall Number Correct	Session criterion met	χ2
			Symmetry relations						Symmetry: Number Correct /60	Equivalence relations			Equivalence Number Correct /30			
E6.SPO.LIN			B1-A1	B2-A2	B3-A3	C1-B1	C2-B2	C3-B3		C1-A1	C2-A2	C3-A3				
P6.1	**4**	**480**	**10**	**10**	**10**	**10**	**10**	**10**	**60**	**10**	**10**	**10**	**30**	**90**	**Y**	_
P6.2	8	960	**10**	0	0	0	**10**	0	20	**10**	0	0	10	30	N	0.00
P6.3	**4**	**480**	**10**	**10**	**10**	**10**	**10**	**10**	**60**	**10**	**9**	**9**	**28**	**88**	**Y**	_
P6.4	**3**	**360**	**10**	**10**	**10**	**10**	**9**	**10**	**59**	**9**	**10**	**10**	**29**	**88**	**Y**	_
P6.5	8	960	**9**	**10**	**10**	**10**	**10**	**10**	59	0	0	**10**	10	69	N	76.05*
P6.6	8	960	**10**	**10**	**10**	**10**	**10**	**10**	60	**10**	0	0	10	70	N	80.00*
P6.7	**3**	**360**	**10**	**10**	**10**	**10**	**10**	**9**	**59**	**10**	**9**	**10**	**29**	**88**	**Y**	_
P6.8	**6**	**720**	**10**	**10**	**10**	**10**	**10**	**10**	**60**	**10**	**10**	**10**	**30**	**90**	**Y**	_
P6.9	**1**	**120**	**10**	**10**	**10**	**10**	**10**	**10**	**60**	**10**	**9**	**10**	**29**	**89**	**Y**	_
P6.10	3	960	5	2	5	2	2	4	20	5	2	4	11	31	N	0.05
E6.MTS.LIN			B1-A3	B2-A1	B3-A2	C1-B3	C2-B1	C3-B2		C1-A2	C2-A3	C3-A1				
P6.11	**2**	**240**	**10**	**10**	**10**	**10**	**10**	**10**	**60**	**10**	**10**	**10**	**30**	**90**	**Y**	_
P6.12	**2**	**240**	**10**	**10**	**10**	**10**	**10**	**10**	**60**	**10**	**10**	**10**	**30**	**90**	**Y**	_
P6.13	8	960	0	0	0	0	**10**	0	10	0	**9**	0	9	19	N	6.050*
P6.14	**2**	**240**	**10**	**10**	**10**	**10**	**10**	**10**	**60**	**10**	**10**	**10**	**30**	**90**	**Y**	_
P6.15	**5**	**600**	**10**	**10**	**10**	**10**	**10**	**10**	**60**	**10**	**10**	**10**	**30**	**90**	**Y**	_
P6.16	8	960	5	4	4	6	4	2	25	2	2	4	8	33	N	0.45
P6.17	8	960	**10**	**9**	**10**	**10**	**10**	**10**	**59**	4	6	6	16	75	N	101.250*
P6.18	8	960	**9**	**10**	**9**	**10**	**10**	**10**	**58**	7	8	**9**	24	82	N	135.200*
P6.19	**3**	**360**	**10**	**10**	**10**	**10**	**10**	**10**	**60**	**9**	**10**	**10**	**29**	**89**	**Y**	_
P6.20	8	960	**8**	6	**10**	6	**10**	**10**	**50**	4	**9**	0	13	63	N	54.450*

*=significant at p<0.05

Table 6.3

Number of cycles and training trials completed, and number of correct responses for each tested relation of the final training and testing cycle completed by the participants who completed SPO (E6.SPO.MTO) or MTS (E6.MTS.MTO) training with a MTO arrangement. Chi square results comparing the number of correct and incorrect responses to a distribution predicted by chance are shown for participants who did not meet the session criterion.

Participant	Number of cycles	Number of training trials	Tested relations - sample and correct comparison stimuli											Overall Number Correct	Session criterion met	χ2
			Symmetry relations						Symmetry: Number Correct /60	Equivalence relations			Equivalence Number Correct /30			
E6.SPO.MTO			B1-A1	B2-A2	B3-A3	C1-A1	C2-A2	C3-A3		C1-B3	C2-B1	C3-B2				
P6.21	4	**480**	**10**	**10**	**10**	**10**	10	**10**	**60.00**	**10**	**10**	**10**	**30.00**	**90.00**	**Y**	_
P6.22	1	**120**	**10**	**10**	**10**	**10**	**10**	**9**	**59.00**	**10**	**10**	**9**	**29.00**	**88.00**	**Y**	_
P6.23	8	960	**10**	**10**	3	1	0	3	27.00	**9**	0	0	9.00	36.00	N	1.80
P6.24	5	**600**	**10**	**10**	**10**	**10**	**10**	**10**	**60.00**	**10**	**10**	**10**	**30.00**	**90.00**	**Y**	_
P6.25	2	**240**	**10**	**10**	**10**	**10**	**10**	**10**	**60.00**	**10**	**10**	**10**	**30.00**	**90.00**	**Y**	_
P6.26	8	**960**	**10**	**10**	**10**	**10**	**10**	**10**	**60.00**	**10**	**10**	**10**	**30.00**	**90.00**	**Y**	_
P6.27	8	960	2	4	1	8	7	7	29.00	8	4	3	15.00	44.00	N	9.80*
P6.28	8	960	**9**	**10**	5	8	8	6	46.00	**8**	**9**	8	25.00	71.00	N	84.05*
P6.29	8	960	0	5	3	6	2	6	22.00	0	5	7	12.00	34.00	N	0.80
P6.30	2	**240**	**10**	**10**	**10**	**10**	**10**	**10**	**60.00**	**10**	**10**	**10**	**30.00**	**90.00**	**Y**	_
E6.MTS.MTO			B1-A1	B2-A2	B3-A3	C1-A1	C2-A2	C3-A3		C1-B3	C2-B1	C3-B2				
P6.31	1	**120**	**10**	**10**	**10**	**10**	**10**	**10**	**60.00**	**10**	**10**	**10**	**30.00**	**90.00**	**Y**	_
P6.32	3	**360**	**10**	**10**	**10**	**10**	**9**	**10**	**59.00**	**10**	**9**	**10**	**29.00**	**88.00**	**Y**	_
P6.33	3	**360**	**9**	**10**	**10**	**10**	**10**	**9**	**58.00**	**10**	**10**	**10**	**30.00**	**88.00**	**Y**	_
P6.34	2	**240**	**10**	**10**	**10**	**10**	**10**	**10**	**60.00**	**10**	**10**	**10**	**30.00**	**90.00**	**Y**	_
P6.35	1	**120**	10	**10**	**10**	**10**	**10**	**10**	**60.00**	**10**	**10**	**10**	**30.00**	**90.00**	**Y**	_
P6.36	3	**360**	**10**	**9**	**9**	**10**	**10**	**10**	**58.00**	**9**	**9**	**10**	**28.00**	**86.00**	**Y**	_
P6.37	3	**360**	**10**	**9**	**10**	**10**	**10**	**10**	**59.00**	**10**	**10**	**10**	**30.00**	**89.00**	**Y**	_
P6.38	1	**120**	**10**	**10**	**9**	**10**	**10**	**10**	**59.00**	**10**	**10**	**10**	**30.00**	**89.00**	**Y**	_
P6.39	8	960	**10**	**10**	**10**	**10**	**10**	**10**	60.00	**10**	0	0	10.00	70.00	N	80.00*
P6.40	8	960	**10**	0	0	0	**10**	0	20.00	0	0	0	0.00	20.00	N	5.00*

*=significant at $p<0.05$

Table 6.4

Number of cycles and training trials completed, and number of correct responses for each tested relation of the final training and testing cycle completed by the participants who completed SPO (E6.SPO.OTM) or MTS (E6.MTS.OTM) training with a OTM arrangement. Chi square results comparing the number of correct and incorrect responses to a distribution predicted by chance are shown for participants who did not meet the session criterion.

Participant	Number of cycles	Number of training trials	Tested relations - sample and correct comparison stimuli											Overall Number Correct	Session criterion met	$\chi 2$
			Symmetry relations						Symmetry: Number Correct /60	Equivalence relations			Equivalence Number Correct /30			
E6.SPO.OTM			A1-B2	A2-B3	A3-B1	A1-C3	A2-C1	A3-C2		C1-B3	C2-B1	C3-B2				
P6.41	2	240	10	10	10	10	10	10	60	10	10	10	30	90	Y	_
P6.42	5	600	9	9	10	10	10	10	58	10	10	10	30	88	Y	_
P6.43	3	360	10	10	10	10	10	10	60	10	10	9	29	89	Y	_
P6.44	8	960	8	4	4	6	7	8	37	6	5	6	17	54	N	51.20*
P6.45	1	120	10	10	10	10	10	9	59	10	10	9	29	88	Y	_
P6.46	2	240	10	10	10	10	9	10	59	9	10	10	29	88	Y	_
P6.47	4	480	10	10	10	9	10	9	58	10	10	10	30	88	Y	_
P6.48	7	840	10	10	10	10	10	10	60	10	10	10	30	90	Y	_
P6.49	3	360	10	10	10	10	10	10	60	10	9	10	29	89	Y	_
P6.50	4	480	10	10	10	10	10	10	60	10	10	10	30	90	Y	_
E6.MTS.OTM			A1-B2	A2-B3	A3-B1	A1-C3	A2-C1	A3-C2		C1-B3	C2-B1	C3-B2				
P6.51	2	240	10	9	10	10	10	9	58	10	10	10	30	88	Y	_
P6.52	2	240	10	10	10	10	10	10	60	10	10	10	30	90	Y	_
P6.53	3	360	10	10	10	10	10	10	60	10	9	10	29	89	Y	_
P6.54	1	120	10	10	10	10	10	10	60	10	10	10	30	90	Y	_
P6.55	8	960	10	10	8	10	8	7	53	9	7	8	24	77	N	110.45*
P6.56	8	960	4	7	4	4	3	4	26	5	4	3	12	38	N	3.20
P6.57	2	240	10	10	10	10	10	10	60	9	10	10	29	89	Y	_
P6.58	1	120	10	10	10	10	10	10	60	10	10	10	30	90	Y	_
P6.59	2	240	9	9	10	10	10	10	58	9	10	10	29	87	Y	_
P6.60	2	240	10	10	10	10	10	10	60	10	10	10	30	90	Y	_

*=significant at $p<0.05$

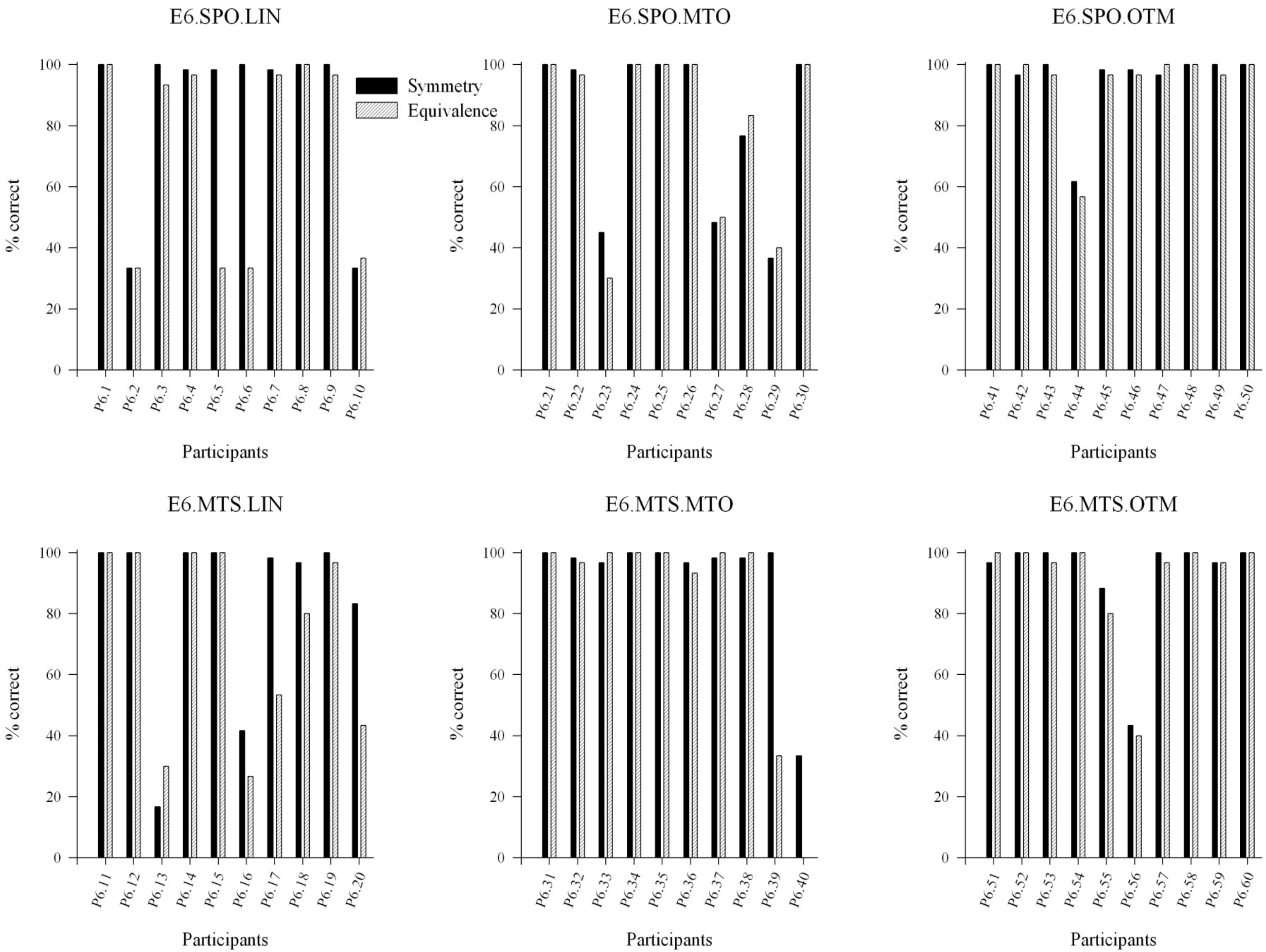

Figure 6.1 Percent correct on the symmetry and equivalence relations during the final training and testing cycle for each of the participants.

Table 6.5

Results of an ANOVA to compare percent correct on the symmetry and equivalence relations of the final session for both procedures for all groups in Experiment 6.

Source	*df*	*F*	$\eta^2_{partial}$
Within-subjects effects			
Type of relation	1,12	8.449*	0.135
Type of relation x Procedure	1,12	0.254	0.005
Type of relation x Stimuli Arrangement (Linear/MTO/OTM)	1,12	2.345	0.080
Type of relation x Procedure x Stimuli Arrangement	1,12	0.772	0.028
Between-subjects effects			
Procedure	1,12	0.030	0.001
Stimuli Arrangement	1,12	1.849	0.064
Procedure x Stimuli Arrangement	1,12	0.233	0.009

*=significant at p<0.05

Number of training trial completed prior to meeting the session criterion during testing. As the majority of participants demonstrated equivalence, a factorial ANOVA was conducted to compare the number of training trials completed prior to meeting the session criterion across the procedure (SPO or MTS) and stimulus arrangement (linear, MTO, or OTM). The total number of training trials (and the number of training and testing cycles) completed by each participant are shown in Tables 6.2-6.4. Tables 6.2-6.4 show there was a lot of within-group variation in the number of training trials completed; however, overall most of the participants who completed MTS training and met the criterion did so following fewer total training trials than the participants who completed SPO training. A factorial ANOVA was conducted to compare the number of training trials completed prior to meeting the session criterion across the procedures and arrangements. The results of the ANOVA (in Table 6.7) show that this difference was significant, and the effect size was above Fergusson's (2009) RMPE. That is, the participants who met the criterion following MTS training completed significantly fewer training trials ($\bar{x}$ = 272.00) than those who met the criterion following SPO training ($\bar{x}$ = 424.44). There were no significant differences in the number of training trials to criterion across the stimuli arrangements, and no significant interaction between the training procedures and the stimuli arrangements. For these results the effect sizes were both very small and below Ferguson's (2009) RMPE.

Table 6.6

Number of cycles completed, number of correct responses for each trial type, and chi square results comparig number of correct and incorrect performace to that predicted by chance (correct (30/90),

Participant	Cycle	A1-B2	A2-B3	A3-B1	B1-C2	B2-C3	B3-C1	Number correct /120	$\chi 2$
E6.MTS.LIN									
P6.11	Cycle 1	9	9	8	17	10	8	61.00	
	Cycle 2	**19**	**20**	**20**	**20**	**20**	**20**	**119.00**	222.34*
P6.12	Cycle 1	9	9	11	**18**	**20**	**18**	85.00	
	Cycle 2	**20**	**19**	**20**	**20**	**20**	**20**	**119.00**	199.84*
P6.13	Cycle 1	12	8	9	12	16	10	67.00	
	Cycle 2	**20**	17	17	**20**	**20**	**20**	114.00	
	Cycle 3	**20**	**20**	**20**	**20**	**20**	**20**	**120.00**	
	Cycle 4	**20**	**20**	**20**	**20**	**20**	**20**	**120.00**	
	Cycle 5	**20**	**20**	**20**	**20**	**20**	**20**	**120.00**	
	Cycle 6	**19**	**20**	**20**	**20**	**20**	**20**	**119.00**	
	Cycle 7	**20**	**19**	**20**	**20**	**20**	**20**	**119.00**	
	Cycle 8	**20**	**20**	**20**	**20**	**20**	**20**	**120.00**	240.00*
P6.14	Cycle 1	17	15	**18**	15	13	14	92.00	
	Cycle 2	**20**	**20**	**20**	**20**	**20**	**20**	**120.00**	216.60*
P6.15	Cycle 1	17	8	8	11	**19**	11	74.00	
	Cycle 2	**18**	17	**20**	**19**	**20**	**20**	114.00	
	Cycle 3	**19**	**20**	**20**	**19**	**20**	**19**	**117.00**	
	Cycle 4	**20**	**19**	**20**	**20**	**19**	**20**	**118.00**	
	Cycle 5	**18**	**20**	**20**	**20**	**20**	**20**	**118.00**	
	Cycle 6	**20**	**20**	**20**	**20**	**19**	**20**	**119.00**	234.04*
P6.16	Cycle 1	9	6	7	8	6	6	42.00	
	Cycle 2	9	4	13	8	4	4	42.00	
	Cycle 3	8	4	14	3	7	13	49.00	
	Cycle 4	4	5	12	5	10	10	46.00	
	Cycle 5	9	3	11	6	10	3	42.00	
	Cycle 6	6	6	4	8	5	9	38.00	
	Cycle 7	10	5	16	6	4	12	53.00	
	Cycle 8	8	5	15	10	6	8	52.00	5.40*
P6.17	Cycle 1	14	8	2	14	15	10	63.00	
	Cycle 2	15	9	3	20	12	8	67.00	
	Cycle 3	**20**	11	9	**20**	12	17	89.00	
	Cycle 4	**20**	**19**	**18**	**20**	**20**	**20**	**117.00**	
	Cycle 5	**20**	**18**	**19**	**20**	**20**	**18**	**115.00**	
	Cycle 6	**20**	**19**	**19**	**20**	**20**	**20**	**118.00**	
	Cycle 7	**20**	**20**	**20**	**19**	**20**	17	116.00	
	Cycle 8	17	**20**	**18**	**20**	**20**	**19**	114.00	205.35*

Table 6.6 continued.

E6.MTS.LIN		Sample and correct comparison stimuli						Number correct /120	χ2
Participant	Cycle	A1-B2	A2-B3	A3-B1	B1-C2	B2-C3	B3-C1		
P6.18	Cycle 1	13	6	7	6	10	10	52.00	
	Cycle 2	17	**19**	**19**	**18**	**20**	**20**	113.00	
	Cycle 3	**19**	**20**	**20**	**20**	**20**	**20**	**119.00**	
	Cycle 4	**20**	**20**	**20**	**20**	**20**	**20**	**120.00**	
	Cycle 5	**20**	**19**	**20**	**20**	**20**	**20**	**119.00**	
	Cycle 6	**20**	**20**	**20**	**20**	**20**	**20**	**120.00**	
	Cycle 7	**20**	**20**	**20**	**20**	**20**	**20**	**120.00**	
	Cycle 8	**20**	**20**	**20**	**20**	**20**	**20**	**120.00**	240.00*
P6.19	Cycle 1	13	13	9	16	17	16	84.00	
	Cycle 2	**20**	**20**	**20**	**20**	**20**	**20**	**120.00**	
	Cycle 3	**19**	**20**	**20**	**20**	**20**	**20**	**119.00**	234.04*
P6.20	Cycle 1	6	9	17	3	9	7	51.00	
	Cycle 2	17	17	**19**	10	15	13	91.00	
	Cycle 3	**20**	**20**	**20**	**18**	11	12	101.00	
	Cycle 4	**19**	**19**	**20**	15	16	16	105.00	
	Cycle 5	**20**	**19**	**19**	**20**	**19**	**20**	**117.00**	
	Cycle 6	**20**	**20**	**20**	**20**	**19**	**20**	**119.00**	
	Cycle 7	**19**	**18**	**20**	**20**	**20**	**20**	**117.00**	
	Cycle 8	**20**	**20**	**20**	17	**19**	**20**	**116.00**	216.60*

E6.MTS.MTO		Sample and correct comparison stimuli						Number correct /120	χ2
Participant	Cycle	B2-A1	B3-A2	B1-A3	C3-A1	C1-A2	C2-A3		
P6.31	Cycle 1	**20**	**19**	12	15	**20**	**19**	105.00	158.44*
P6.32	Cycle 1	7	8	15	14	17	10	71.00	
	Cycle 2	**20**	**20**	**20**	**20**	**20**	**20**	**120.00**	
	Cycle 3	**20**	**20**	**20**	**20**	**20**	**20**	**120.00**	240.00*
P6.33	Cycle 1	15	11	7	8	17	9	67.00	
	Cycle 2	**20**	16	**19**	**20**	**20**	**20**	**115.00**	
	Cycle 3	**20**	**20**	**20**	**20**	**20**	**20**	**120.00**	240.00*
P6.34	Cycle 1	**10**	15	12	8	11	15	71.00	
	Cycle 2	**20**	**19**	**20**	**20**	**20**	**20**	**119.00**	234.04*
P6.35	Cycle 1	15	17	11	15	9	12	79.00	57.04*
P6.36	Cycle 1	5	7	6	14	11	18	61.00	
	Cycle 2	**18**	**19**	14	**19**	15	**20**	105.00	
	Cycle 3	**20**	**20**	17	**20**	**20**	**20**	117.00	222.34*
P6.37	Cycle 1	**20**	7	6	5	**20**	13	71.00	
	Cycle 2	**20**	13	15	17	**20**	**20**	105.00	
	Cycle 3	**19**	17	17	**20**	**20**	**20**	113.00	199.84*

Table 6.6 continued.

E6.MTS.MTO		Sample and correct comparison stimuli						Number correct /120	χ2
Participant	Cycle	B2-A1	B3-A2	B1-A3	C3-A1	C1-A2	C2-A3		
P6.38	Cycle 1	4	12	**18**	9	12	14	69.00	31.54*
P6.39	Cycle 1	11	4	4	5	10	**20**	54.00	
	Cycle 2	7	17	13	14	**18**	**20**	89.00	
	Cycle 3	**20**	**20**	**20**	**19**	**20**	**20**	**119.00**	
	Cycle 4	**20**	**20**	**20**	**20**	**20**	**20**	**120.00**	
	Cycle 5	**20**	**20**	**20**	**20**	**20**	**20**	**120.00**	
	Cycle 6	**20**	**20**	**20**	**20**	**20**	**20**	**120.00**	
	Cycle 7	**20**	**20**	**20**	**20**	**20**	**20**	**120.00**	
	Cycle 8	**20**	**20**	**20**	**20**	**20**	**20**	**120.00**	240.00*
P6.40	Cycle 1	12	3	6	8	14	5	48.00	
	Cycle 2	15	16	11	**19**	**17**	15	93.00	
	Cycle 3	**19**	**20**	**19**	**18**	**20**	**18**	**114.00**	
	Cycle 4	**19**	**20**	**20**	**20**	**20**	**20**	**119.00**	
	Cycle 5	**16**	**20**	**19**	17	**20**	17	109.00	
	Cycle 6	**18**	**20**	**19**	**19**	**20**	**16**	112.00	
	Cycle 7	**18**	**20**	**20**	16	**19**	**16**	109.00	
	Cycle 8	**18**	**20**	**20**	**19**	**20**	**19**	**116.00**	216.60*

E6.MTS.OTM		Sample and correct comparison stimuli						Number correct /120	χ2
Participant	Cycle	A1-B2	A2-B3	A3-B1	A1-C3	A2-C1	A3-C2		
P6.51	Cycle 1	**18**	14	**18**	15	14	16	95.00	
	Cycle 2	**20**	**20**	**20**	**20**	**20**	**19**	**119.00**	234.04*
P6.52	Cycle 1	**14**	**18**	**19**	**18**	13	13	95.00	
	Cycle 2	**18**	**20**	**19**	**19**	**20**	**19**	**115.00**	210.94*
P6.53	Cycle 1	5	9	6	3	12	16	51.00	
	Cycle 2	11	9	7	8	**18**	6	59.00	
	Cycle 3	16	15	**20**	**20**	**19**	**18**	108.00	173.40*
P6.54	Cycle 1	**19**	**18**	17	17	16	17	104.00	153.60*
P6.55	Cycle 1	15	12	18	8	13	13	79.00	
	Cycle 2	**20**	19	**20**	**20**	**20**	18	117.00	
	Cycle 3	**20**	**20**	**20**	20	**20**	**20**	120.00	
	Cycle 4	**20**	**20**	**20**	**18**	**20**	**19**	**117.00**	
	Cycle 5	**19**	**20**	**20**	**19**	**20**	**20**	**118.00**	
	Cycle 6	**19**	**20**	**20**	**18**	**18**	14	109.00	
	Cycle 7	**20**	**20**	**20**	**20**	**19**	**20**	**119.00**	
	Cycle 8	**20**	**20**	**20**	**18**	**19**	**19**	**116.00**	216.60*

Table 6.6 continued.

E6.MTS.OTM		Sample and correct comparison stimuli						Number correct /120	$\chi 2$
Participant	Cycle	A1-B2	A2-B3	A3-B1	A1-C3	A2-C1	A3-C2		
P6.56	Cycle 1	4	6	10	6	6	2	34.00	
	Cycle 2	9	7	5	9	8	9	47.00	
	Cycle 3	10	9	5	9	8	3	44.00	
	Cycle 4	4	5	7	6	9	10	41.00	
	Cycle 5	5	5	1	7	1	7	26.00	
	Cycle 6	**18**	9	**18**	16	14	8	83.00	
	Cycle 7	**20**	15	16	**18**	17	16	102.00	
	Cycle 8	**20**	14	15	**20**	14	17	100.00	135.00*
P6.57	Cycle 1	13	11	10	16	13	**19**	82.00	
	Cycle 2	**20**	**19**	**20**	**20**	**20**	**20**	**119.00**	234.04*
P6.58	Cycle 1	**18**	15	**19**	13	17	**19**	101.00	
P6.59	Cycle 1	17	11	17	16	12	14	87.00	
	Cycle 2	**20**	**20**	**18**	**19**	**20**	**19**	**116.00**	216.60*
P6.60	Cycle 1	9	10	18	9	4	**19**	69.00	
	Cycle 2	17	14	**20**	17	13	**20**	101.00	139.54*

Table 6.7

Results of an ANOVA to compare the number of training trials completed prior to reaching criterion on the tested relations on the symmetry and equivalence relations of the final session for both procedures for all groups in Experiment 6.

Source	*df*	*F*	$\eta^2_{partial}$
Between-subjects effects			
Procedure	1,36	6.524*	0.153
Stimuli Arrangement	2,36	0.324	0.018
Procedure x Stimuli Arrangement	2,36	0.297	0.016

*=significant at p<0.05

Discussion

The present experiment compared the effectiveness of the linear, MTO, and OTM stimulus arrangements across the SPO and MTS procedures. The participants in this experiment completed twice the number of training trials per training and testing cycle as participants in Experiments 1 to 5. With the exception of the number of training trials per cycle, the procedure of this experiment was identical to that of Experiment 5 and to the procedure experienced by the participants in Experiment 4 who were presented with the less specific instructions (Groups E4.SPO and E4.MTS). Therefore, the data from these groups can be compared. This allowed the examination of the effect the number of training trials completed prior to each testing condition across the three stimulus arrangements.

As in the previous experiments of this study, the SPO and MTS procedures were equally effective at facilitating the formation of equivalence relations. In contrast to the previous experiments in this study, most of the participants here met the session criterion before the end of the final training and testing cycle. For the groups who completed SPO training, the OTM stimulus arrangement resulted in a greater number of participants meeting the session criterion than either of the MTO or linear arrangements. This finding does not agree with those of Experiment 5 where performance was similar for the SPO procedure with the MTO and OTM arrangements, and both were more effective than the findings for the linear procedure used in Experiment 4.

In the present experiment there was no difference in the number of participants who met the session criterion for the MTO and OTM arrangements when MTS training was used and both MTO and OTM were more effective than the linear arrangement. This is consistent with the results of Experiments 4 and 5 when the MTS procedure was used. As with the findings of Experiment 5, the similar effectiveness of the MTO and OTM procedures (when using MTS training) agree with Fields et al.'s (1999) hypothesis that these procedures should not result in differences in outcomes with normally developed adult participants using equivalence classes that involve only a small number of stimuli. It is not clear there were differences in the findings regarding the effectiveness of the MTO and OTM procedures across the experiments. However, a greater number of participants in Experiment 5 may have helped to clarify this finding here.

While there was a difference in the number of participants who achieved equivalence across the stimuli arrangements, when percent correct during testing of the final cycle for each participant was compared statistically, there was no difference in the effectiveness of the three stimulus arrangements was found. Therefore, while the OTM procedure resulted in a greater number of the participants meeting the session criterion, it did not result in greater accuracy during the final session. There was also no difference in the accuracy achieved during the final session across the SPO and MTS procedures.

As in the previous experiments, there was some evidence to suggest that the participants performed better on the symmetry than the equivalence relations. In the present experiment this difference during the final training and testing cycle was statistically significant, but the effect size was below Fergusson's (2009) RMPE. This effect size was smaller than those obtained for this difference in any of the previous experiments. This suggests that the increase in the number of training trials per cycle resulted in higher accuracy on the equivalence relations, thus decreasing the difference in accuracy between the symmetry and equivalence relations. Most of the participants who performed better on the symmetry than on the equivalence relations were in the groups that experienced a linear stimulus arrangement. This is similar to the findings across Experiments 4 and 5. Across these experiments, a difference in accuracy on the symmetry and equivalence trials was seen when the linear arrangement was used but not when the MTO or OTM arrangements were used. As outlined in the *Discussion* of Experiment 5, this finding is similar to those of Arntzen and Holth (1997) but differs from Hove's (2003) finding. Hove (2003) reported a difference in performance on the symmetry and equivalence relations with MTO and OTM stimulus arrangements. Unlike the findings across Experiments 4 and 5, the interaction between the type of relation and stimulus arrangement in the present experiment was not significant, and the effect size was below the RMPE. Therefore, it appears that the difference in performance on the symmetry and equivalence relations when a linear arrangement was used was greater in the earlier experiments where the participants completed 60 training trials per cycle than when there were 120 training trials per cycle.

To assess the number of training trials statistically, a factorial repeated measures ANOVA was conducted to compare the percentage of correct responses made on the symmetry and equivalence relations during the final session by each participant in this experiment, Experiment 5, and Groups 4.SPO and 4.MTS (who received the less-specific instructions) of Experiment 4. The between-subjects factors were the procedure (SPO or MTS) and the stimuli arrangement (linear, MTO or OTM). The results of this ANOVA are presented in Table 6.14. There was a significant within-subjects difference in performance on the symmetry and equivalence relations. The effect size for this difference was above Ferguson's (2009) RMPE indicating that this was a practically significant difference. Significant within-subject interactions were found between the type of relation and the stimulus arrangement, and the type of relation and the number of trials and a significant three-way interaction was found between the type of relation, the stimulus arrangement, and the number of trials. Of these, only the interaction between the type of relation and the stimulus arrangement showed a practically significant effect size. On closer inspection, this interaction indicates that there was a difference in performance on the symmetry and equivalence relations when the linear stimulus arrangement was used (symmetry = 86.90%, equivalence = 60.69%), but that this difference was not observed for the MTO and OTM arrangements (MTO symmetry $\bar{x}$ = 88.44%, MTO equivalence $\bar{x}$ = 82.25; OTM symmetry $\bar{x}$ = 90.81%, OTM equivalence $\bar{x}$ = 90.22%). Therefore, while all procedures were effective at facilitating the formation of symmetry relations, the linear arrangement failed to facilitate the development of the equivalence relations. As discussed above, when assessed separately this interaction was found to be significant when 60, but not 120, training trials were completed in each training and testing cycle. None of the remaining within-subjects interactions, or any of the between-subjects effects (procedure, stimulus arrangement, or number of trials) was significant, and all effect sizes were below Ferguson's (2009) RMPE.

To summarise the ANOVA findings across the present experiment, Experiment 5, and the groups of Experiment 4 who received the same instructions as in Experiments 5 and here, the OTM and MTO stimulus arrangements resulted in better performance in the final test on the equivalence, but not on the symmetry relations than the linear arrangement. Performance on the symmetry relations was good regardless of the stimulus arrangement; therefore, only performance on the equivalence relations was affected by the change in stimulus arrangement.

Table 6.8

Results of an ANOVA to compare percent correct on the symmetry and equivalence relations of the final session for both procedures for all participants in the E4.SPO, E4.MTS, Experiment 5 and Experiment 6.

Source	*df*	*F*	$\eta^2_{partial}$
Within-subjects effects			
Type of relation (symmetry/equivalence)	1,72	22.104*	0.235
Type of relation x Procedure (SPO/MTS)	1,72	0.023	0.000
Type of relation x Stimuli arrangement (Linear/MTO/OTM)	2,72	12.822*	0.263
Type of relation x Number of trials (60/120)	1,72	4.118*	0.054
Type of relation x Procedure x Stimulus arrangement	2,72	0.122	0.003
Type of relation x Procedure x Number of trials	1,72	0.097	0.001
Type of relations x Stimulus arrangement x Number of trials	2,72	4.420*	0.109
Type of relation x Procedure x Stimulus Arrangement x Number of trials	2,72	1.364	0.037
Between-subjects effects			
Procedure	1,72	1.001	0.014
Stimulus arrangement	2,72	3.011	0.077
Number of trials	1,72	0.438	0.006
Procedure x Stimulus arrangement	2,72	1.323	0.035
Procedure x Number of trials	1,72	0.677	0.009
Stimulus arrangement x Number of trials	2,72	0.411	0.011
Procedure x Stimulus arrangement x Number of trials	2,72	0.942	0.025

*=significant at $p<0.05$

As mentioned previously, SPO and MTS procedures were equally effective when measured in terms of accuracy on the final test. These procedures also resulted in the same number of participants demonstrating equivalence. However, the participants who completed MTS training and demonstrated equivalence did so within fewer training and testing cycles, and thus fewer total training trials, than the participants who completed SPO training. Therefore, while the training procedure did not affect either performance during the final session or the number of participants to reach criterion, MTS resulted in the formation of equivalence relations over fewer trials than SPO training for those participants who demonstrated equivalence.

It was not possible to conduct a statistical comparison of the number of training trials completed prior to meeting the criterion during testing in the previous experiments due to the low numbers of participants in these experiments who demonstrated equivalence. However, the use of 120 training trials per training and testing resulted in the demonstration of equivalence by approximately half of the participants who experienced the linear stimulus arrangement. In contrast, equivalence was not demonstrated with a linear arrangement by any of the participants who completed 60 training trials per cycle. This finding cannot be attributed solely to the greater total number of trials that could have been completed by the participants in the present experiment, as most of the participants who demonstrated equivalence with the linear arrangement did so in fewer than the number of trials (480) that were available for the participants in groups who completed 60 training trials per cycle. This finding was not as strong for the MTO and OTM groups across the experiments. For these groups, the proportion of participants who achieved equivalence was generally higher with 120 training trials per cycle. The small number of participants in the groups who experienced 60 training trials per cycle means that caution is required in interpreting these results. For those participants who completed MTO and OTM training and achieved equivalence, the total number of training trials completed prior to demonstrating equivalence criterion was similar regardless of the number of trials in a cycle. Nearly all of the participants in the groups in the present experiment who experienced the MTO or OTM arrangements and demonstrated equivalence did so within the maximum 480 trials that could have been completed by the participants in their respective groups in Experiment 5. Therefore, it appears that increasing the number of training trials in each training and testing cycle increased the likelihood of achieving equivalence when the linear, but not when the MTO or OTM arrangements were used, and that this effect was a result of the larger number of trials experienced prior to testing, not the total number of trials completed.

Most of the participants who completed MTS training and achieved equivalence also performed well during training. As outlined previously, a criterion was not used during MTS training as the aim was to keep trial numbers the same with both procedures. However, for the purposes of assessing performance during testing, accuracy was measured against a criterion-level performance of 9 out of 10 trials correct on each trained relation. Most participants who completed MTS training and demonstrated equivalence also achieved 90% correct on each trained relation during training. A few participants achieved equivalence without achieving 90% correct on each trained relation. As noted in Experiment 1, it is not generally considered possible for equivalence relations to develop without the prior formation of the trained relations from which the equivalence relations are derived (Sidman, 1994). This result was unexpected. However, the large number of training trials completed in each training and testing cycle means that it is possible that the participants had responded correctly on many trials without error prior to testing even if they did not respond correctly on 90% for each trained relation overall. For example, during each training and testing cycle here, the participants completed 20 trials with each trained relation (120 trials in total for the six trained relations). There were also several participants in the present experiment who achieved 90% correct during training but then failed to demonstrate equivalence during testing. Failure to demonstrate equivalence following a criterion-level performance during training is not uncommon, as evidenced by the numbers of participants who did not demonstrate equivalence after meeting the training criterion in the studies outlined in Experiment 1.

In summary, there was no difference in the effectiveness of the SPO and MTS procedures in terms of the number of participants who demonstrated equivalence, or accuracy on the tested relations during the final session. However, for those participants who demonstrated equivalence, the MTS procedure resulted in the acquisition of equivalence relations after fewer trials than did the SPO procedure. The OTM arrangement resulted in the greatest number of participants demonstrating equivalence when the SPO procedure was used and the OTM and MTO procedures were equally effective when the MTS procedure was used. All three stimulus arrangements resulted in similar accuracy on the symmetry relations; however the MTO and OTM procedures were more effective at facilitating the formation of equivalence relations. Increasing the number of training trials per cycle resulted in a greater number of participants achieving equivalence with the linear procedure and the data suggest that it was the number of training trials per cycle, not the total number of training trials completed that was responsible for this result. This finding was not seen when the OTM and MTO arrangements were used.

GENERAL DISCUSSION

Overall aim of the study

This thesis started by aiming to compare the effectiveness of the SPO and MTS procedures in facilitating the formation of equivalence relations. As a result of the outcomes of the initial experiments the thesis explored the effects of several procedural factors on the formation of equivalence relations using these two procedures. This next section will present a summary of the findings.

Summary

Experiment 1 was a procedural replication of the first experiment by Leader and Barnes-Holmes (2001b) but using Chinese characters. It included two groups of participants that could and two groups that could not read the Chinese characters. The results did not replicate those of Leader and Barnes-Holmes (2001b) or those of Clayton and Hayes (2004) as no differences were found between the effectiveness of the SPO and MTS procedures. Prior experience with the Chinese characters did not affect this outcome. In fact few of the participants demonstrated equivalence after either SPO or MTS training. However, nearly all participants achieved accuracy during testing that was greater than that predicted by chance and generally performed well on some tested relations. Where this occurred, the participants were sometimes choosing the comparison stimulus that was the correct response when presented with that sample stimulus following the other training procedure. This suggested that requiring the participants to learn conflicting relations was interfering with the development of equivalence. Overall, participants performed better on the symmetry than the equivalence relations regardless of the training procedure used. Most participants responded correctly on some trained relations but not others during MTS training. A few achieved a criterion-level performance by the end of their final in the MTS training session but nearly all of these did not achieve equivalence in the test session.

Given the failure to replicate Leader and Barnes-Holmes's (2001b) findings it was decided to repeat Experiment 1, replacing the Chinese characters with the nonsense syllables used originally by Leader and Barnes-Holmes (2001b). Despite this more direct replication of Leader and Barnes-Holmes's (2001b) experiment, no participants demonstrated equivalence. As in Experiment 1, greater accuracy was generally found with the symmetry than with the equivalence relations and participants performed accurately on some relations but not others. For some participants, good performance on a tested relation following one training condition was sometimes paired with poor performance on the tested relation involving a conflicting relation following the other training condition. In addition, performance during MTS training was poor, with only one participant demonstrating a 'criterion-level' performance. The similarity in the results of Experiments 1 and 2 suggested that it was not the stimuli used in Experiment 1 that gave rise to the failure to replicate Leader and Barnes-Holmes's (2001b) findings.

Experiments 1 and 2 results suggested that the use of the same stimuli in conflicting relations across the two training procedures may have interfered with the formation of equivalence relations. Therefore, Experiment 3 examined the effect of using different stimuli in each training procedure. As in the previous experiments, performance on the symmetry relations was better than performance on the equivalence relations. While most of the participants still failed to demonstrate equivalence, the accuracy achieved was generally greater than in Experiments 1 and 2. Thus, the use of different stimuli in each training procedure helped to improve performance in the test sessions. However, there was no difference in the effectiveness of the SPO and MTS procedures. Most of the participants achieved a criterion-level performance during MTS training, suggesting that it was the use of the same stimuli in conflicting relations during training that was responsible for the poor performance seen in Experiments 1 and 2.

Experiment 2 suggested that the instructions given and the number of training trials were two variables that might have contributed to the failure of the participants to demonstrate equivalence. Several participants in each of the first three experiments were uncertain about the task that they were required to complete. The instructions used in Experiments 1 to 3 were the same as those used by Leader and Barnes-Holmes (2001b). These had remained the same throughout Experiments 1 to 3 to allow the effect of the type of stimuli, and the use of the same stimuli in conflicting relations to be examined in isolation. However, this anecdotal evidence obtained during the debriefing following completion of the experiment suggested that instructions that were more specific to the task might aid in the formation of the equivalence classes. Therefore, Experiment 4 replicated Experiment 3 with half of the participants receiving instructions that outlined the task in more detail. The participants who received these modified instructions were also required to complete a comprehension task. This was to ensure that they understood the task required. Experiment 4 used a between-subjects design in which participants experienced only one training procedure (SPO or MTS) and this gave enough time for participants to complete more training and testing cycles. This allowed the question of whether experiencing only one procedure and allowing more training and testing cycles to be completed would aid in the development of equivalence. The more specific instructions resulted in more participants achieving equivalence during the final session compared to the participants who received the less-specific instructions. The participants who demonstrated equivalence did so in fewer than four cycles and so did not complete any more training and testing cycles than were completed in the early experiments. Some of the participants who did not achieve equivalence showed increases in accuracy across the sessions suggesting that they might have demonstrated equivalence after further training and testing. As with the previous experiments, better performance on the symmetry than the equivalence relations was seen for those participants who did not demonstrate equivalence, and there was no difference in the effectiveness of the SPO and MTS procedures. As in Experiment 3, most participants were performing well in the MTS training sessions by the end of the MTS training. While the instructions resulted in a small number of participants achieving equivalence, the lack of clarity around the effect of instructions on equivalence formation and Sidman's (1992) comments regarding the interaction of instructions with previous verbal histories resulted in the use of the original 'less-specific' instructions in Experiments 5 and 6. At the end of Experiment 4, the number of training trials completed across the experiment and the arrangement of stimuli in the trained (and tested) relations were chosen as areas to explore further.

Experiment 5 examined the effect of stimulus arrangement. Previous research suggested that MTO and OTM arrangements were more effective at facilitating equivalence class formation than the linear arrangement used in Experiments 1 to 4 and by Leader and Barnes-Holmes (2001b). Therefore, Experiment 5 compared the MTO and OTM arrangements. In all other respects Experiment 5 was a replication of the procedure used in Experiment 4 with the 'less-specific' instructions. Therefore the results from the MTO and OTM arrangements could be compared directly to those from the linear arrangement used in Experiment 4. Across both the SPO and MTS procedures, the MTO and OTM stimulus arrangements resulted in greater numbers of participants achieving equivalence than did the linear procedure. As with the previous experiments, the SPO and MTS procedures were equally effective at facilitating the formation of equivalence classes regardless of the stimulus arrangement, and participants who did not achieve equivalence generally performed better on the symmetry than the equivalence relations. Nearly all participants achieved a ‘criterion-level’ performance during MTS training by their final session.

Across Experiments 3 to 5, some of the participants who failed to demonstrate equivalence showed increases in the accuracy they achieved during testing across the training and testing cycles. The results of Experiment 4 showed that adding further training and testing cycles had not aided in equivalence formation. Therefore, in Experiment 6, the number of training trials per training and testing cycle was increased from 60 to 120. As the results of Experiment 5 showed the MTO and OTM stimuli arrangements to aid in equivalence formation, all three arrangements were used in Experiment 6. In all other respects the procedure was the same as Experiment 5. The inclusion of the linear arrangement allowed the findings to be compared with the previous experiments in the study.

Most of the participants in Experiment 6 demonstrated equivalence. Equal numbers of participants demonstrated equivalence following the SPO and MTS procedures. The two procedures also resulted in similar accuracy on the tested relations. However, the MTS procedure resulted in participants demonstrating equivalence within fewer trials than the SPO procedure. When the SPO procedure was used, participants were more likely to demonstrate equivalence with the OTM arrangement than with the MTO or linear arrangements. When the MTS procedure was used, the OTM and MTO arrangements were equally effective and were both more effective than the linear arrangement. This differs from the findings of Experiment 5 where the OTM and MTO arrangements were similarly effective and more effective than the linear arrangement (in Experiment 4) for both the SPO and MTS procedures. In Experiment 6, some of the participants who did not demonstrate equivalence performed better on the symmetry than the equivalence relations. Most of these participants had experienced the linear arrangement. Just over half of the participants who demonstrated equivalence following MTS training also produced a ‘criterion-level’ performance during MTS training. This is a different finding from Experiments 3 to 5, where most participants performed well during MTS training even if they did not then demonstrate equivalence. It was suggested that this was possibly due to the larger number of trials completed during MTS training of Experiment 6. A number of the participants who did not demonstrate equivalence also achieved a criterion-level performance during training. As outlined in Experiment 6, failing to achieve equivalence following a criterion-level performance during training is not uncommon.

Discussion

Comparison of the SPO and MTS procedures

Over all of the experiments here there was no difference in the effectiveness of the SPO and MTS procedure in terms of the accuracy achieved in the test sessions. As outlined in the *Introduction*, Leader and Barnes-Holmes (2001b) found the SPO, and Clayton and Hayes (2004) found MTS procedure to be most effective. The present results do not agree with the findings of either of these studies and so the relative effectiveness of these two procedures is still unclear. Experiment 6 showed the MTS procedure facilitated the formation of equivalence relations in fewer training trials than the SPO procedure. Thus it could be argued that MTS was the more effective procedure in the present study. Clayton and Hayes (2004) do not give the number of trials that their participants completed prior to demonstrating equivalence so no comparison can be made to the present result. Leader and Barnes-Holmes's (2001b) participants demonstrated equivalence following both training procedures after fewer training and testing cycles if they experienced SPO training first in each session than if they experienced MTS training first in each session. However, as their participants completed both MTS and SPO training this finding is not directly comparable to those in Experiment 6, here. In the present study, where the participants completed both procedures, few, or no participants demonstrated equivalence following either procedure. Thus, this thesis does not agree conclusively with either of the previous studies, and the comparative effectiveness of the SPO and MTS procedures is an area that requires further attention.

Accuracy on symmetry and equivalence relations

When participants achieve equivalence, accuracy on both the symmetry and equivalence relations must be high as they must have met the criterion required to demonstrate equivalence. It is, therefore, possible to compare performance on the two types of test trials only when participants have not demonstrated equivalence. Across all experiments in the present study, many of the participants who failed to demonstrate equivalence performed better on the symmetry than on the equivalence relations. In Experiment 6 this pattern was present mainly when a linear procedure was used. This pattern is not evident for the MTO or OTM procedures in Experiment 6 as most participants who experienced those stimulus arrangements performed well on both the symmetry and equivalence relations.

Clayton and Hayes (2004) is the only one of the two studies that compared the SPO and MTS procedures to provide their findings for the symmetry and equivalence relations separately. They reported no difference in performance on the different relations. However, better performance on the symmetry than on the equivalence relations has been reported in other studies (e.g., Leader et al., 1996; Pilgrim & Galizio, 1990; 1995; Rehfeldt, 2003). As symmetry relations are generally considered to be precursor relations to equivalence relations (Sidman, 1994; Sidman & Tailby, 1982) it is not surprising that, where participants have failed to develop equivalence, some still performed well on the symmetry relations.

The unequal performance on the symmetry and equivalence relations seen here and in some other studies, as mentioned above, suggests that research that aims to improve the effectiveness of methods that are used to facilitate the formation of equivalence relations should assess performance on the symmetry and equivalence relations separately. As performance on these two sets of relations can vary greatly, an overall measure of accuracy does not provide a good measure of what is being learnt. The prevalence of this pattern also suggests that future research should focus on identifying procedural modifications that improve performance on the equivalence relations, such as increasing the number of training trials per cycle and the use of the MTO and OTM stimulus arrangements here.

The use of the same stimuli in conflicting relations

One of the factors that affected the results of Experiments 1 and 2 was the use of the same stimuli in conflicting relations. In those experiments the same stimuli were used in both training procedures and so participants were required to learn different equivalence classes involving the same stimuli. A response pattern emerged for some participants where, during testing, they were most likely to choose the comparison stimulus that was the correct choice when presented with that same sample stimulus in testing following the other training condition. It seems then that the conflicting relations hindered the formation of equivalence classes for these participants. This response pattern was also reported by Leader and Barnes-Holmes (2001b). In that study, SPO training was found to be more effective than MTS training, and participants were likely to respond in accordance with the correct responses for the tested relations following SPO training in testing following MTS training. However, in the present study this effect occurred in both directions and both procedures resulted in similar overall accuracy on the tested relations. As noted in Experiment 3, Clayton and

Hayes (2004) outlined how consistent good performance on one of the conflicting relations over another would indicate the superior performance of one training procedure over another. In the present experiment, participants did, at times, respond correctly on some relations but they choose the comparison stimulus that was the correct response following the other training procedure for other relations within the equivalence test. However, the relations on which performance was good did not occur consistently as the result of one training procedure. This resulted in reduced accuracy for both training procedures overall, therefore, rendering both ineffective.

As outlined in Experiment 3, while some studies have shown that reversal of equivalence relations is possible (e.g., Smeets et al., 2003; Spradlin et al., 1973, cited in Spradlin et al., 1992; Spradlin et al., 1992; Wirth & Chase, 2001), the use of the same stimuli in conflicting relations has been shown to interfere with the formation of equivalence relations (e.g., Pilgrim & Galizio, 1990; 1995; Saunders et al., 1988). Two of these studies (Pilgrim & Galizio, 1990; 1995) identified that a reversal procedure was more likely to result in correctly reversed symmetry than equivalence or transitivity relations in the testing condition. This present study did not test for transitivity, however, the findings agree with those of Pilgrim and Galizio (1990; 1995) in that participants performed better on the symmetry than on the equivalence relations. Therefore, the greater accuracy on the symmetry than the equivalence relations in Experiments 1 and 2 of the present study may be partly the result of learning the reversed symmetry but not the equivalence relations.

Instructions

Increasing the specificity of the instructions in Experiment 4 resulted in a small increase in the effectiveness of the both the SPO and MTS procedures. As only one study (Leader et al., 1996) has examined the effect of instructional specificity with the SPO procedure, the present study has added to the body of research in this area.

This avenue of inquiry was not taken further in this thesis. It is a complex area and it is possible that any effect attributed to the instructions could be the result of the individual verbal histories of the participants. What is apparent from reviewing the research is that the role of instructions in equivalence tasks is far from clear. As outlined in the *Introduction* to Experiment 4, the instructions used in equivalence research vary greatly across studies. The studies differ in how instructions are worded, how they are presented, and in the information that they provide. While some studies (as outlined in Experiment 4) have examined instructional effects in equivalence research, they have varied across a number of procedural factors. This makes the comparison of the findings across these studies difficult. Therefore, there is a need for further research that systematically examines the effects of different components, content and presentation of instructions.

Stimulus Arrangement

Experiments 5 and 6 found the OTM and MTO arrangements to be equally effective with the SPO and MTS training procedures when there were 60 training trials per cycle, and with the MTS procedure when there were 120 training trials per cycle. With SPO training and 120 training trials per cycle, the OTM arrangement was more effective than the MTO or linear arrangements. These findings suggest a possible interaction between the number of training trials and the stimulus arrangement. That is, the effectiveness of the stimuli arrangements differed depending on the training procedure used. However, as the sample sizes in Experiment 5 are small, this suggestion is made with caution and requires further research.

The equal effectiveness of the MTO and OTM arrangements under most conditions does not agree with much of the research which has found the MTO procedure to be the most effective. The finding does agree with Fields et al.'s (1999) suggestion that both of these arrangements should be equally likely to result in equivalence formation with normally developed adult participants such as those used here.

Smeets et al. (1997) and Smeets and Barnes-Holmes (2005) state that most research comparing the effectiveness of the different stimulus arrangements on the formation of equivalence relations fail to do so as they test for symmetry relations prior to testing for equivalence relations. Their argument is that if the equivalence relations are tested after the participants have completed tests for symmetry then their responses during equivalence may not be based solely on the trained discriminations (Smeets et al., 1997). Instead responding during the equivalence tests may "be based on any demonstrated relations, that is, the trained relations...the tested symmetry relations...or a combination of trained and tested relations" (Smeets & Barnes-Holmes, 2005, p.282). In the present study, the tested symmetry and equivalence relations occurred within the same testing condition and in a mixed quasi-random order. It is unclear whether intermixing the symmetry and equivalence trials would have affected performance. However, this could be addressed by presenting only the equivalence tests following training, or presenting the equivalence tests initially, followed by the symmetry tests if equivalence is not achieved.

Number of training trials

The participants in Experiments 1 to 3 of this study completed a maximum of four training and testing cycles with each training procedure. As there were 60 training trials per cycle and all participants failed to meet the session criterion - all participants completed 240 training trials in each training condition within a session. As outlined in the *Discussion* for Experiment 1, the participants in Leader and Barnes-Holmes (2001b) could have completed a maximum of 360 training trials. However, all of their participants who met the session criterion did so within four training and testing cycles (240 training trials) with each procedure. Therefore, the smaller number of cycles available to the participants here cannot account for the difference in findings of the studies.

The increased number of training trials per training and testing cycle in Experiment 6 resulted in more participants achieving equivalence. The fact that most of these participants demonstrated equivalence within the total number of trials available to participants in the earlier experiments suggests that it was the number of training trials per training and testing cycle, and not the total number of training trial*s* completed that was responsible for this result. This finding is contrary to the finding by Layng and Chase (2001), that a large block of trials prior to the first equivalence test is less effective at facilitating the formation of equivalence relations than the same number of trials spread across multiple training and testing cycles. However, it has been suggested that repeated exposures to training and MTS testing are required in order for equivalence to develop (e.g., Sidman, 1992). This may account for the failure of Layng and Chase’s (2001) participants to demonstrate equivalence when all of the trials occurred prior to the first equivalence test. The necessity of multiple training and testing cycles is also supported by the findings that very few participants in the presents study achieved equivalence in the first equivalence test, regardless of the number of training trials completed, or the stimulus arrangement.

The number of training trials used in equivalence research varies greatly both between and within experiments (as outlined in the *Introduction* to Experiment 6). Also the effects of different stimulus arrangements of the training and testing trials are not clear. Thus further research is required to examine the effect of both of these factors on the formation of equivalence tasks.

Performance during MTS training

As outlined in Experiment 6, most equivalence research using MTS training uses a criterion that participants must achieve during training prior to beginning tests for equivalence. A criterion was not used here as the same number of exposures to the stimuli was used with both the MTS and the SPO procedures both within and across participant, and also because it is not possible to have a performance criterion with SPO. If a criterion of 90% correct had been used with MTS training, few participants in Experiments 1 and 2 would have achieved this during the MTS training. As it is highly unlikely that participants would perform well on equivalence tests without having first learnt the baseline relations, it is not surprising that few or no participants in those experiments demonstrated equivalence following MTS training. Response patterns suggesting that the use of the same stimuli in conflicting relations was interfering with performance during MTS training were seen for some participants. The greater accuracy achieved during MTS training by the participants in Experiments 3 to 6 confirm the conflicting relations as the likely cause of the poor performance during MTS training of Experiments 1 and 2.

As most participants in Experiments 3 to 5 achieved more than 90% correct during MTS training, it is unlikely that the lack of a training criterion hindered the formation of equivalence relations in these experiments. As mentioned earlier, and outlined in Experiment 1, other studies have reported similar findings with participants producing a criterion-level performance during training but not demonstrating equivalence.

Procedures in equivalence research

One problem with assessing the effect of different procedural factors on equivalence class formation is the number of procedural differences that are present between studies. As outlined in the *Introduction*, as well as varying in terms of MTS, SPO or pREP procedures, within each of these procedures there are many different factors that can vary (e.g., the type of stimuli used, the number of equivalence classes trained, the number of training and testing trials, pre-experimental or familiarisation procedures, the stimuli arrangement (linear, MTO, OTM), and the instructions given). The great variation between studies of equivalence means that any comparisons between studies must be made with caution. This thesis has provided an examination of the effect of some of these factors across the SPO and MTO procedures.

One feature of the procedure used in this study differs from that used in many other studies. This was that the alphanumeric designations of the stimuli were not balanced across the stimuli in the current study. That is, all of the participants experienced the same stimuli in the same trials (e.g., A1 referred to the same stimulus for all participants). This differs from the common practice of balancing the stimuli so that each pair of stimuli (e.g., A1-B1) does not involve the same stimuli for each participant. However, as outlined by Underwood

(1949) this practice of counterbalancing does not eliminate order effects but disguises them. Therefore, holding constant the alphanumeric designation of the stimuli allows the identification of stimuli, or pairs of stimuli that result in different response patterns. No consistent response patterns to particular stimulus pairs or individual stimuli were observed across the experiments in this study. Therefore, we can conclude that the participants were not showing biases towards particular stimuli.

Another procedural factor on which studies can differ is the range of derived relations that are tested. For example, Hayes et al. (1991) tested for both symmetry and equivalence relations, Barnes-Holmes et al. (2000) tested for transitivity and equivalence, but not symmetry, and Arntzen (2006) tested for equivalence only.

The demonstration of equivalence classes infers that a participant would respond correctly on tests for symmetry and transitivity. Therefore, which relations were tested is only questioned when participants fail to demonstrated equivalence. The present study tested for symmetry and equivalence, but not transitivity. A number of the participants who failed to demonstrate equivalence here did perform well on the symmetry relations. However, it is not possible to say whether the participants in this study who did not demonstrate equivalence would have responded correctly on tests for transitivity. As such, the inclusion of tests for transitivity would have allowed the identification of transitive responding where equivalence was not demonstrated.

One other procedural factor that is of particular relevance here is the simultaneous presentation of the sample and comparison stimuli in MTS training. In research involving MTS procedures, the comparison are either presented on the screen with the sample stimulus (simultaneous presentation), or they are presented at varying delays following the removal of the sample stimulus (successive presentation). In the present study, the sample and comparison appeared on the screen together. In the present study, the sample stimulus was presented alone for 1.5 s, after which it was joined by the three comparison stimuli. The sample and comparison stimuli remained on the screen until a response was made. This procedure was used in the present study as the initial experiments were procedural replications of Leader and Barnes-Holmes (2001b). Subsequent experiments used the same procedures to allow the results of the experiments to be compared directly.

While such simultaneous presentation is not uncommon (e.g., Holth & Arntzen, 1998; Markham et al., 2002), it may be a confounding factor in a study that aims to compare the differential effects of operant (MTS) and associative (SPO) procedures as it introduces an associative learning element to the MTS procedure. That is, as the sample and comparison stimuli are presented together on the screen, it can be argued that resulting relations may be learned more easily as their relation is not the result of just the contingencies in the MTS procedure but also because they are paired together in each trial. However, it could be argued that a simultaneous protocol also pairs the sample stimulus with the two incorrect comparison stimuli. It is not clear whether successive presentation of the sample and comparison stimuli may have resulted in different findings.

Arnzten (2006) studied the effect of different delays between the removal of the sample stimuli and the presentation of the comparison stimuli on equivalence formation with a MTS procedure. That study included a condition where the sample and comparison stimuli were presented together. As outlined in Experiment 5, the findings Arntzen (2006) suggested that when a MTO stimulus arrangement was used equivalence was more likely when larger delays were used than when the stimuli were presented simultaneously. That is, simultaneous presentation was less likely to result in equivalence than a delay between the presentation of the sample stimulus and the comparison stimuli. When a OTM stimulus arrangement was used the delay did not affect the likelihood of equivalence. That is, whether the comparison stimuli were presented with the sample stimulus simultaneously, or after a delay, did not affect the likelihood of equivalence.

As mentioned previously, Leader and Barnes-Holmes (2001b) used a simultaneous presentation. That study found the SPO procedure to be more effective than the MTS procedure. In their final experiment they removed the incorrect comparison stimuli from the response options in the MTS training, leaving only the sample stimulus and the correct comparison stimulus. It could be argued that this reduced the MTS procedure to a SPO procedure that required a response from the participant. Performance on the equivalence task following MTS training was better when that procedure was used, than when three comparison stimuli were presented. It seems then, in that study, that associative components of the MTS procedure aided in the formation of equivalence classes, when only the correct comparison stimulus was paired (on the screen) with the sample stimulus. Clayton and Hayes (2004) also presented the sample and comparison stimuli on the monitor together. That study found the MTS procedure to be the more effective, contrary to Leader and Barnes-Holmes's (2001b) findings. Together with Clayton and Hayes's (2004) findings, the improvement in the effectiveness of the MTS procedure when it was made to be more like a SPO procedure in Leader and Barnes-Holmes's (2001b) study suggest that the associative elements of the MTS procedure may aid in the formation of equivalence classes when MTS procedures are used.

Theoretical implications of this research.

Operant and associative procedures. As mentioned previously, Rehfeldt and Hayes (1998) have argued that the role of associative learning in stimulus equivalence procedures needs more recognition, and that the processes involved are separate from the procedures used. Specifically, they argue that while a procedure can be classed as either operant or associative, this does not mean that the processes involved in learning the equivalence classes are also only operant or associative. Instead it is likely that both operant and associative processes are involved.

In all of the experiments of the present study, the MTS procedure involved all of the basic events of the SPO procedure. That is, the MTS procedure involved pairing two stimuli (the sample stimulus and the correct comparison) together across a number of trials, as did the SPO procedure. These procedures all involve association. The presence of operant events in the SPO procedure are not so clear, as no response is required nor is reinforcement available on a trial by trial basis. However, in the wider context, the training/testing procedure has similarities to other events, such as practice and test taking, with which university students (the primary participant pool) are familiar. Therefore, the behaviour that occurred under these conditions could have been a generalised operant. Also, it is possible that the behaviour observed during training and testing was at least partly rule-governed as instructions were provided at the start of each training and testing condition. Instructions allow the occurrence of behaviour that has not contacted a reinforcement contingency. Behaviour such as observing the stimuli in the SPO procedure may have been rule governed. It is also possible that the participants generated novel, idiosyncratic, rules that may have affected performance.

Therefore, while the SPO and MTS procedures are labelled as operant and associative respectively, some of the behaviours that these procedures evoke may not actually differ greatly. Overall, across all of the experiments in this study, the SPO and MTS procedures proved equally effective in facilitating the formation of equivalence relations. Performance changed equally with the two procedures when other factors, such as the number of training trials and the stimulus arrangement, were varied. Therefore, it seems that even though the arrangement of experimental events appear to differ the outcomes do not.

Implications of this research for the theories of stimulus equivalence

There are three main theories that inform equivalence relations. Horne and Lowe's (1996) naming hypothesis suggests that being able to name objects is necessary for equivalence to develop. In the present experiment, all of the participants were capable of naming; therefore this hypothesis was not tested directly. However, the inclusion in Experiment 1 of students who could read the Chinese characters provides an interesting comparison that is relevant to this argument. For these participants, the Chinese characters were easily nameable, using the words they represent. In contrast, these stimuli would not have been as easily nameable for the participants who reported that they could not read them. No consistent differences were observed between these groups and, overall, the participants in Experiment 1 performed poorly on the equivalence relations in most cases.

If being able to name the stimuli was advantageous to the development of equivalence, it would also be expected that the introduction of the nonsense syllables in Experiment 2 would have made equivalence more likely. These stimuli were arbitrary, as with the Chinese characters in Experiment 1 for the participants who could not read them. However, unlike the Chinese characters, the nonsense syllables could be pronounced phonetically, and so could be easily named. However, despite this, none of the participants in Experiment 2 demonstrated equivalence with either the SPO or MTS procedure. Therefore, there is no evidence that being able to name the stimuli affected the development of equivalence in the present study.

Sidman's (1994) theory of equivalence, based on mathematical set theory, is outlined in the *Introduction* of this thesis. However, regarding naming, Sidman (1994) argues that it is not clear whether naming is necessary for the development of equivalence relations, although being able to name the stimuli makes equivalence more likely. The findings of the present study do not support this.

The third theory that informs stimulus equivalence is Relational Frame Theory (RFT). Within RFT, equivalence relations are one type of derived relation. RFT also acknowledges that both operant stimulus-reinforcer relations and associative stimulus-stimulus relations can be involved in the formation of derived relations. As pointed out in the *Introduction,* proponents of RFT have noted that much of the learning involving derived relations involves the association of stimuli, rather than a response-reinforcer relation (e.g., Blackledge, 2004; Dixon et al., 2006). In this sense, RFT includes consideration of both types of relations and so can account for the findings of this study where training procedures involving stimulus-stimulus and stimulus-reinforcer relations have resulted in the formation of equivalence relations.

Conclusion

Overall, the SPO and MTS procedures were found to be equally effective at facilitating the formation of equivalence relations in terms of accuracy achieved, however, the MTS procedure resulted in the development of equivalence over fewer training trials than did the SPO training. The formation of equivalence relations was hindered when an attempt was made to compare the SPO and MTS procedures using the same stimuli in each training condition, resulting in the participants having to learn conflicting relations. The use of different stimuli with each training procedure resulted in greater accuracy achieved, but this was not enough on its own to result in the formation of equivalence classes and participants failed to demonstrate equivalence with both procedures. Detailed instructions made equivalence slightly more likely, although this effect was minimal. Variations in the procedures used and the complex and unknown effect of differing verbal histories make this an area of equivalence research that requires much more attention. Two procedural factors were shown by this study to affect equivalence formation. These were the arrangement of the stimuli, and the number of training trials experienced within each training and testing cycle. In general, equivalence research is made more difficult by the lack of consistency in the procedures used. As such, the experiments in this thesis have helped to isolate the individual effects of a number of the procedural differences that may affect the findings of equivalence research.

REFERENCES

Adcock, A.C., Merwin, R.M., Wilson, K.G., Drake, C.E., Tucker, C.I. & Elliot, C. (2010). The problem is not learning: Facilitated acquisition of stimulus equivalence classes among low-achieving college students. *The Psychological Record, 60,* 43-56.

American Psychological Association (2010). *Publication manual of the American psychological association (6th ed.).* Washington: American Psychological Association.

Arntzen, E. (2006). Delayed matching to sample: probability of responding in accord with equivalence as a function of different delays. *The Psychological Record, 56*(1), 135-167.

Arntzen, E. and Holth, P. (1997). Probability of stimulus equivalence as a function of training design. *The Psychological Record, 47*(2), 309-320.

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

Bach, P. & Hayes, S.C. (2002). The use of acceptance and commitment therapy to prevent the rehospitalization of psychotic patients: A randomized controlled trial. *Journal of Consulting and Clinical Psychology, 20*(5), 1129-1139.

Balluerka, N., Gomez, J. & Hildago, D. (2005). The controversy over null hypothesis significance testing revisited. *Methodology, 1*(2), 55-70.

Barnes, D. & Keenan, M. (1993). A transfer of functions through derived arbitrary and nonarbitrary stimulus relations. *Journal of the Experimental Analysis of Behavior, 59*(1), 61-81.

Barnes-Holmes, D., Barnes-Holmes, Y., Smeets, P.M., Cullinan, V. & Leader, G. (2004). Relational frame theory and stimulus equivalence: conceptual and procedural issues. *International Journal of Psychology and Psychological Therapy, 4*(2), 181-214.

Barnes-Holmes, D., Keane, J. Barnes-Holmes, Y. & Smeets et al., P.M. (2000). A derived transfer of emotive functions as a means of establishing differential preferences for soft drinks. *The Psychological Record, 50,* 493-511.

Barnes, D., Lawlor, H., Smeets, P.M and Roche, B. (1996). Stimulus equivalence and academic self-concept among mildly mentally handicapped and nonhandicapped children. *The Psychological Record, 46,* 87-107.

Blackledge, J.T. (2003). An introduction to relational frame theory: Basics and applications. *The Behavior Analyst Today, 3*(4), 421-433.

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.

Carr, D. & Blackman, D.E. (2001). Relations among equivalence, naming, and conflicting baseline control. *Journal of the Experimental Analysis of Behavior, 75*(1), 55-76.

Clayton, M.C, & Hayes, L.J. (2004). A comparison of match-to-sample and respondent-type training of equivalence classes. *The Psychological Record, 54,* 579-602.

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd ed.). New

Cohen, J. (1994). The earth is round (p<.05). *American Psychologist, 49*(12), 997-1003.

Cullinan, V.A, Barnes, D. & Smeets, P.M. (1998) A precursor to the relational evaluation procedure: Analyzing stimulus equivalence. *The Psychological Record, 48*(1), 121-145.

Cullinan, V.A., Barnes-Holmes, D. & Smeets, P.M (2000). A precursor to the relational evaluation procedure: Analyzing stimulus equivalence II. *The Psychological Record, 50*(3), 467-492.

de Medeiros, C.A., de Freitas Ribeiro, A. & de Faria Galvao, O. (2003). Effect of instructions on the demonstration of equivalence among locations. *Psicologia: Teoria e Pesquisa, 19*(2), 175-171.

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 behaviour, 50*(1), 1-20.

Dibbets, P. Maes, J.H.R., & Voessen, J.M.H., (2002). Contextual dependencies in a stimulus equivalence paradigm. *The Quarterly Journal of Experimental Psychology, 55B*(2), 97-119.

Dickins, D.W., Bentall, R.P. and Smith, A.B. (1993). The role of individual stimulus names in the emergence of equivalence relations: The effects of interpolated paired-associates training of discordant associations between names. *The Psychological Record, 43*(4), 713-724.

Dixon, M.R., Rehfeldt, R.A., Zlomke, K.R. and Robinson, A. (2006). Exploring the development and dismantling of equivalence classes involving terrorist stimuli. *The Psychological Record, 56,* 83-103.

Dougher, M.J., Auguston, E., Markham, M.R. & Greenway, D.E. (1994). The transfer of respondent eliciting and extinction functions through stimulus equivalence classes. *Journal of the Experimental Analysis of Behavior, 62*(3), 331-351.

Drake, C.E. & Wilson, K.G. (2008). Instructional effects on performance in a matching-to-sample study. *Journal of the Experimental Analysis of Behavior, 89*(3), 333-340.

Duarte, A.M., Eikeseth, S., Rosales-Ruiz, J. & Baer, D.M. (1998). The effects of a can't-answer response on option and instructions on stimulus equivalence. *The Psychological Record, 48*(4), 631-646.

Dube, W.V., McIlvane, W.J., Maguire, R.W., Mackay, H.A., & Stoddard, L.T. (1989). Stimulus class formation and stimulus-reinforcer relations. *Journal of the Experimental Analysis of Behavior, 51*(1), 65-76.

Dymond, S. & Barnes, D. (1997). The effects of prior equivalence testing and verbal instructions on derived self-discrimination transfer: a follow-up study. *The Psychological Record, 47*(1), 147-170.

Eifert, G.H., Forsyth, J.P., Arch, J., Espejo, E., Keller, M. & Langer, D. (2009). Acceptance and commitment therapy for anxiety disorders: three case studies exemplifying a unified treatment protocol. *Cognitive and Behavioral Practice*, *16,* 368-385.

Eikeseth, S., & Baer, D.M. (1997). Using a preexisting verbal relation to prevent the properties of stimulus equivalence from emerging in new relations. In, Baer, D,M. & Pinkston, E.M. (Ed.), *Environment and behavior* (pp.138-144). Boulder: Westview Press.

Eikeseth, S., Rosales-Ruiz, J., Duarte, A., & Baer, D.M. (1997). The quick development of equivalence classes in a paper-and-pencil format through written instructions. *The Psychological Record*, *47*(2), 275-284.

Elias, N.C., Goyos, C. Saunders, M. & Saunders, R. (2008) Teaching manual signs to adults with mental retardation using matching-to-sample procedures and stimulus equivalence. *Analysis of Verbal Behavior, 24,* 1-13.

Ferguson, C.J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, *40*(5), 532-538.

Fields, L., Hobbie-Reeve, S.A., Adams, B.A., & Reeve, F. (1999). Effects of training directionality and class size on equivalence class formation by adults. *The Psychological Record, 49,* 703-724.

Fields, L. & Verhave, T. (1987). The structure of equivalence classes. *Journal of the Experimental Analysis of Behavior, 48*(2), 317-322.

Fields, L. Reeve, K.F., Adams, B.J., Brown, J.L. & Verhave, T. (1997). Predicting the extension of equivalence classes from primary generalization gradients: the merger of equivalence classes and perceptual classes. *Journal of the Experimental Analysis of Behavior*, *68*(1), 67-91.

Fields, L., Reeve, K.F., Varelas, A., Rosen, D. & Belanich, J. (1997). Equivalence class formation using stimulus-pairing and yes-no responding. *The Psychological Record, 47*(4), 661-686.

Forman, E.M., Herbert, J.D., Moitra, E. Yeomans, P.D. & Geller, P.A. (2007). A randomized controlled effectiveness trial of acceptance and commitment therapy and cognitive therapy for anxiety and depression. *Behavior Modificatioin*, *31*(6), 772-799.

Green, G., Sigurdardottir, 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.

Hayes, L.J., Thompson, S., and Hayes, S.C. (1989). Stimulus equivalence and rule following. *Journal of the Experimental Analysis of Behavior, 52*(3), 275-291.

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.

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

Hove, O. (2003). Differential probability of equivalence class formation following a one-to-many versus a many-to-one training structure. *The Psychological Record*, 53, 617-634.

Juarascio, A.S., Forman, E.M., & Herbert, J.D. (2010). Acceptance and commitment therapy versus cognitive therapy for the treatment of comorbid eating pathology. *Behavior Modification, 34*(2), 175-190.

Kirk, R.E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement, 56*(5), 746-759.

Layng, M.P. & Chase, P.N. (2001). Stimulus-stimulus pairing, matching-to-sample, and emergent relations. *The Psychological Record, 51*, 605-628.

LeBlanc, L.A., Miguel, C.F., Cummings, A.R., Goldsmith, T.R. & Carr, J.E. (2003). The effects of three stimulus-equivalence testing conditions on emergent US geography relations of children diagnosed with autism. *Behavioral Interventions, 18*(4), 279-289.

Leader, G., Barnes, D. & Smeets, P.M. (1996). Establishing equivalence relations using a respondent-type training procedure. *The Psychological Record, 46*(4), 685-706.

Leader, G., & Barnes-Holmes, D. (2001b). Matching-to-sample and respondent-type training as methods for producting eqivalence relations: Isolating the critical variable. . *The Psychological Record, 51*(429-444).

Leader, G. & Barnes-Holmes, D. (2001a). Establishing fraction-decimal equivalence using a respondent-type training procedure. *The Psychological Record, 51,* 151-165.

Leader, G., Barnes-Holmes, D., & Smeets, P.M., (2000). Establishing equivalence relations using a respondent-type training procedure III. *The Psychological Record, 50*(1), 63-78.

Leslie, J.C., Tierney, K.J., Robinson, C.P. & Keenan, M. (1993). Differences between clinically anxious and non-anxious subjects in a stimulus equivalence training task involving threat words. *The Psychological Record, 43*(1), 153-161.

Lunde, L. & Nordhus, I.H., & Stale, P. (2009). The effectiveness of cognitive and behavioural treatment of chronic pain in the elderly: A quantitative review. *Journal of Clinical Psychology in Medical Settings, 16(3),* 254-262.

Markham, M.R., Dougher, M.J. & Auguston, E.M. (2002). Transfer of operant discrimination and respondent elicitation via emergent relations of compound stimuli. *The Psychological Record*, *52*(3), 325-350.

Merwin, R.M. & Wilson, K.G. (2005). Preliminary findings on the effects of self-referring and evaluative stimuli on stimulus equivalence class formation. *The Psychological Record, 55,* 561-575.

Michael, R.L. & Bernstein, D.J. (1991). Transient effects of acquisition history on generalization in a matching-to-sample task. *Journal of the Experimental Analysis of Behavior, 56*(1), 335-347.

Minster, S.T., Jones, M. Eliffe, D., & Muthukumaraswamy, D. (2006). Stimulus equivalence: testing Sidman's (2000) theory. *Journal of the Experimental Analysis of Behavior*, *85*(3), 371-391.

Moxon, P.D., Keenan, M., and Hine, L. (1993). Gender-role stereotyping and stimulus equivalence. *The Psychological Record*, *45,* 207-222.

Nakagawa, S. and Foster, T.M. (2004). The case against retrospective statistical power analyses with an introduction to power analysis. *Acta Ethologica*, *7*(2), 95-101.

O'Toole, C., Barnes-Holmes, D. & Smith, S. (2007). A derived transfer of functions and the implicit association test. *Journal of the Experimental Analysis of Behavior, 88*(2), 263-283.

Pilgrim, C., Chambers, L., & Galizio, M. (1995b). Reversal of baseline relations and stimulus equivalence: II. Children. *Journal of the Experimental Analysis of Behavior, 63(3), 239-254.*

Pilgrim, C. & Galizio, M. (1990). Relations between baseline contingencies and equivalence probe performances. *Journal of the Experimental Analysis of Behavior*, *54*(3), 214-224.

Pilgrim, C. & Galizio, M. (1995). Reversal of baseline relations and stimulus equivalence: I. Adults. *Journal of the Experimental Analysis of Behavior, 63*(3), 225-238.

Plaud, J.J., Gaither, G.A., Franklin, M., Weller, L.A., & Barth, J. (1998). The effects of sexually explicit words on the formation of stimulus equivalence classes. *The Psychological Record*, *48*(1), 63-79.

Peoples, M., Tierney, K.J., Bracken, M. & McKay, C. (1998). Prior learning and equivalence class formation. *The Psychological Record, 48*(1), 111-120.

Randell, T. & Remington, B. (2006). Equivalence relations, contextual control, and naming. *Journal of the Experimental Analysis of Behavior, 86*(3), 337-354.

Rehfeldt, R.A. (2003). Establishing contextual control over generalized equivalence relations. *The Psychological Record,* 53(3), 415-428.

Rehfeldt, R.A. & Hayes, L.J. (1998). The operant-repondent distinction revisited: toward an understanding of stimulus equivalence. *The Psychological Record, 48*(2), 187-211.

Roche, B. & Barnes, D. (1997). A transformation of respondently conditioned stimulus function in accordance with arbitrarily applicable relations. *Journal of the Experimental Analysis of Behavior,* 67(3), 275-301.

Roche, B., Barnes, D. and Smeets, P.M. (1997). Incongruous stimulus pairing and conditional discrimination training: effects on relational responding. *Journal of the Experimental Analysis of Behavior, 68*(2), 143-160.

Rosnow, R.L. and Rosenthal, R. (2003). Effect sizes for experimenting psychologists. *Canadian Journal of Experimental Psychology, 57*(3), 221-237.

Rosales-Ruiz, J. Eikeseth, S. Duarte, A. & Baer, D.M. (2000). Verbs and verb phrases as instructional stimuli in the control of stimulus-equivalence effects. *The Psychological Record, 50,* 173-187.

Saunders, R.R., Chaney, L. and Marquis, J.G. (2005). Equivalence class establishment with two-, three-, and four choice matching to sample by senior citizens. *The Psychological Record, 55*(4), 539-559.

Saunders, R.R., Drake, K.M. & Spradlin, J.E. (1999). Equivalence class establishment, expansion, and modification in preschool children. *Journal of the Experimental Analysis of Behavior, 71*(2), 195-214.

Saunders, R.R. & Green, G. (1999). A discrimination analysis of training-structure effects on stimulus equivalence outcomes. *Journal of the Experimental Analysis of Behavior, 72*(1), 117-137.

Saunders, R.R. & McEntee, J.E. (2004). Increasing the probability of stimulus equivalence with adults with mild mental retardation. *The Psychological Record, 54*(3), 423-435.

Saunders, R.R., Saunders, K.J., Kirby, K.C. and Spradlin, J.E. (1988). The merger and development of equivalence classes by unreinforced conditional selection of comparison stimuli. *Journal of the Experimental Analysis of Behavior, 50*(2), 145-162.

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. Special Issue: Stimulus Equivalence, 43*(4), 725-744.

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.

Spradlin, J.E., & Saunders, R.R. (1986). The development of stimulus classes using match-to-sample procedures: Sample classification versus comparison classification. *Analysis & Intervention in Developmental Disabilities. Special Issue: Stimulus control research and developmental disabilities, 6*(1-2), 41-58.

Schenk, J.J. (1994). Emergent relations of equivalence generate by outcome-specific consequences in conditional discrimination. *The Psychological Record, 44,* 537-558.

Sidman, M. (1971). Reading and auditory equivalence. *Journal of Speech & Hearing Research, 14*(1), 5-13.

Sidman, M. (1992). Equivalence relations: some basic considerations. In, Hayes, S.C. & Hayes, L.J. (Ed.). *Understanding Verbal Relations.* Nevada: Context Press (pp.15-28).

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

Sidman. M, & Cresson, O. (1973). Reading and crossmodal transfer of stimulus equivalence in severe retardation. *American Journal of mental Deficiency, 77*(5), 515-523.

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.

Sigurdardottir, Z.G., Green, G. and Saunders, R.R. (1990). Equivalence classes generated by sequence training. *Journal of the Experimental Analysis of Behavior, 53*(1), 47-63

Smeets, P.M., Barnes-Holmes, D. & Roche, B. (2001). Derived stimulus-response and stimulus-stimulus relations in children and adults: Assessing training order effects. *Journal of Experimental Child Psychology, 78,* 130-154.

Smeets, P.M. & Barnes-Holmes, D. (2005). Establishing equivalence classes in preschool children with one-to-many and many-to-one training protocols. *Behavioural Processes, 69*(3), 281-293.

Smeets, P.M., Barnes-Holmes, Y., Akpinar, D. & Barnes-Holmes, D. (2003). Reversal of equivalence relations. *The Psychological Record, 53*, 91-119.

Smeets, P.M., Barnes-Holmes, D. & Striefel, S. (2006). Establishing and reversing equivalence relations with a precursor to the relational evaluation procedure. *The Psychological Record, 56*(2), 267-286.

Smeets, P.M., Dymond, S. & Barnes-Holmes, D. (2000). Instructions, stimulus equivalence and stimulus sorting: Effects of sequential testing arrangements and a default option. *The Psychological Record, 50*(2), 339-354.

Smeets, P.M., Leader, G. & Barnes, D. (1997). Establishing stimulus classes in adults and children using a respondent-type training procedure: A follow-up study. *The Psychological Record, 47*(2), 285-308.

Smeets, P.M., van Wijngaarden, M., Barnes-Holmes, D., & Cullinan, V. (2004). Assessing stimulus equivalence with a precursor to the relational evaluation procedure. *Behavioral Processes, 65*(3), 241-251.

Smyth, S., Barnes-Holmes, D. & Forsyth, J.P. (2006). A derived transfer of simple discrimination and self-reported arousal functions in spider fearful and non-spider-fearful participants. *Journal of the Experimental Analysis of Behavior, 85*(2), 223-246.

Spradlin, J.E., Saunders, K.J. and Saunders, R.R. (1992). Stability of equivalence relations. In. Hayes, S.C. and Hayes, L.J. (Eds.), *Understanding Verbal Relations* (pp.29-42). Reno, Nevada: Context Press.

Stewart, I., Barnes-Holmes, D., Roche, B. and Smeets, P.M. (2002). Stimulus equivalence and non-arbitrary relations. *The Psychological Record, 52*, 77-88.

Tonneau, F. & Gonzalez, C. (2004). Function transfer in human operant experiments: The role of stimulus pairing. *Journal of the Experimental Analysis of Behavior, 81*(3), 239-255.

Twohig, M.P. & Crosby, J.M. (2010). Acceptance and commitment therapy as a treatment for problematic internet pornography viewing. *Behavior Therapy, 41*(3), 285-295.

Underwood, B.J. (1949). *Experimental psychology*. New York: Appleton-Century Crofts, Inc.

Watt, A., Keenan, M., Barnes, D. and Cairns, E. (1991). Social categorization and stimulus equivalence. *The Psychological Record, 41,* 33-50.

Wilkinson, L. and the Task Force on Statistical Inference (1999). Statistical methods in psychology journals. *American Psychologist 54*(8), 594-604.

Wilson, K.G. and Hayes, S.C. (1996). Resurgence of derived stimulus relations. *Journal of the Experimental Analysis of Behavior, 66*(3), 267-281.

Wirth, O. and Chase, P.N. (2002). Stability of functional equivalence and stimulus equivalence: effects of baseline reversals. *Journal of the Experimental Analysis of Behavior, 77*(1), 29-47.

Ybarra Sagarduy, J.L., Soriano, M.C.L. and Gomez Martin, S. (2002). Relaciones de equivalencia: Competitividad entre la preexperimental y experimental. / Equivalence relations: Competition between pre-experimental and experimental history. *International Journal of Clinical and Health Psychology, 2*(1), 137-152.