

The Role of Equivalence-Based Procedures in Establishing Knowledge of Carbohydrate Content in Food Items

O Papel de Procedimentos de Ensino Baseados em Equivalência de Estímulos no Estabelecimento de Conhecimentos sobre Quantidades de Carboidratos em Itens Alimentícios

Erik Arntzen¹, Ina Charlotte Spilling¹

[1] Oslo Metropolitan University | **Título abreviado:** | **Endereço para correspondência:** Oslo Metropolitan University, Department of Behavioral Science, Postbox 4 St. Olavs Plass, 0130 Oslo, Norway | **Email:** erik.arntzen@equivalence.net | **doi:** 10.18761/PAC2025_01a

Abstract: The study aimed to train participants on the carbohydrate values of various food items and evaluate the effect over a three-week period. Twenty-seven participants were assigned to three different groups. Groups 1 and 2 aimed to establish three-member classes, while Group 3 aimed for five-member classes. Initially, participants pre-sorted cards labeled with food items in categories based on the amount of carbohydrates: “less than 20,” “20–40,” and “more than 40” grams per 100 grams. Additionally, participants in Groups 2 and 3 could sort items into a “don’t know” category. Any incorrectly sorted items were identified and customized for each participant. Thus, tailored stimuli, incorrectly sorted stimuli, were then used for each participant in Many-to-One conditional discrimination training. Maintenance of the learned categories was tested after one week. Participants who did not meet the mastery criterion received additional training and testing. A second test of maintenance was conducted after three weeks. Both tests included tests for equivalence class formation in MTS and sorting format. The results showed that most participants responded correctly in both test formats. The majority of participants maintained the equivalence classes across all groups over the three-week period.

Keywords: carbohydrates, maintenance, sorting, stimulus equivalence, tailoring of stimuli

Resumo: O estudo teve como objetivo treinar participantes a identificar os valores de carboidratos de diversos alimentos e avaliar o efeito desse treino ao longo de um período de três semanas. Vinte e sete participantes foram distribuídos em três grupos diferentes. Os Grupos 1 e 2 tinham como meta estabelecer classes de três membros, enquanto o Grupo 3 tinha como meta estabelecer classes de cinco membros. Inicialmente, os participantes pré-classificaram cartões rotulados com itens alimentares em categorias baseadas na quantidade de carboidratos: “menos de 20”, “20–40” e “mais de 40” gramas por 100 gramas. Além disso, os participantes dos Grupos 2 e 3 puderam classificar itens na categoria “não sei”. Todos os itens classificados incorretamente foram identificados e personalizados para cada participante. Assim, os estímulos personalizados — ou seja, os estímulos classificados incorretamente — foram utilizados no treino de discriminação condicional do tipo muitos-para-um. A manutenção das categorias aprendidas foi testada após uma semana. Os participantes que não atingiram o critério de domínio receberam treinamento e testagem adicionais. Um segundo teste de manutenção foi conduzido após três semanas. Ambos os testes incluíram provas de formação de classes de equivalência em procedimento MTS (matching to sample) e em formato de classificação. Os resultados mostraram que a maioria dos participantes respondeu corretamente em ambos os formatos de teste. A maior parte dos participantes manteve as classes de equivalência em todos os grupos ao longo do período de três semanas.

Palavras-chave: carboidratos, manutenção, classificação, equivalência de estímulos, personalização de estímulos

Stimulus equivalence means that stimuli within a class are mutually interchangeable, such that replacing one stimulus with another does not alter the probability of a specific response (Green & Saunders, 1998) and is defined by three key properties: reflexivity, symmetry, and transitivity (Sidman, 1992; Sidman & Tailby, 1982). Reflexivity means that a stimulus has a relation with itself, such that ‘if A, then A.’ Symmetry refers to the reversibility of trained relations between stimuli. In other words, ‘if A, then B,’ it follows that ‘if B, then A.’ Transitivity involves a third stimulus, where if ‘A leads to B’ and ‘B leads to C’ are trained, the relations ‘if A, then C’ emerges without direct training.

Conditional discrimination training typically occurs in a matching-to-sample (MTS) format. In a computerized MTS format, a sample is presented on the screen, and clicking or touching the sample stimulus is followed by the presentation of two or more comparisons. A programmed consequence follows a response to one of the comparison stimuli. The MTS format is typically used in Equivalence-based Instruction (EBI). EBI is based on the principles of stimulus equivalence (e.g., Marin & Fienup, 2024) and was previously referred to as Stimulus Equivalence Technology (Critchfield & Fienup, 2010). In EBI, individuals are trained to respond to meaningful stimuli that are arbitrarily related and later tested for emergent relations. EBI enhances the accessibility of complex or abstract course content by establishing equivalence relations between concepts. Thus, EBI has been applied across various areas, including (a) academic skills (Albright et al., 2015; Fienup et al., 2010; Ong et al., 2018), (b) children’s skills (Bolanos et al., 2020; Hill et al., 2020; LaFond et al., 2021), (c) portion size estimation (Regan et al., 2018; Trucil et al., 2015; Vladescu et al., 2021), and (d) training nutritional content (Arntzen & Eilertsen, 2020; Hausman et al., 2017; Nastally et al., 2010).

For the training of nutritional facts, Nastally et al. (2010) utilized two OTM training structures (AB and AC; CD and CE) to teach calorie content and promote healthier food choices via a computerized stimulus equivalence procedure. Six students aged 20 to 28 participated. The A stimuli consisted of three different calorie levels (A1, below 500 calories; A2, between 500–1000 calories; and A3, above

1000 calories). B, C, D, and E stimuli included images of fast food and logos. In the first part, a pre- and post-preference test was conducted, ranking the food items according to the participants’ desire to eat them. In the next part, a calorie discrimination pre- and post-test was performed, where participants were asked to categorize pictures of food items based on their calorie level. The stimuli used in the conditional discrimination training were food items that participants had previously categorized incorrectly. In the third part, conditional discrimination training was conducted for AB and AC relations, followed by a mix of AB and AC relations, and finally, a test for BC and CB relations. In the fourth part, conditional discrimination training was conducted for CD and CE relations, followed by a mix of CD and CE relations, and finally, a test for DE and ED relations. The results indicated that participants improved their sorting into categories by an average of 45%. Four out of six participants self-reported an increase in healthier food choices, and it was found that Body Mass Index (BMI) had no correlation to correct responses. The authors suggest that future research could examine other nutritional content beyond calories to determine if the intervention yields similar results.

Arntzen and Eilertsen (2020) expanded Nastally et al. (2010) by focusing on carbohydrate knowledge using EBI. Carbohydrates, as the primary energy source for organs like the brain and kidneys, are divided into simple and complex types. Simple carbohydrates are sugars, and complex carbohydrates are starches. In Arntzen and Eilertsen (2020), 22 participants, aged 19–54, were randomly assigned to three groups, and the experiment involved sorting food items into categories based on their carbohydrate content. The experiment was divided into four phases. The first phase involved pre-sorting of food items in a table-top format. Group 1 sorted the stimuli into the categories ‘less than 20,’ ‘20-40,’ and ‘more than 40.’ Groups 2 and 3 had a fourth category, ‘don’t know.’ Groups 1 and 2 were presented with a pile of 21 stimulus cards with different food items, while Group 3 was presented with a pile of 33 stimulus cards with different food items. Participants were asked to sort the stimulus cards below the category cards. Incorrectly sorted items were used

as stimuli in subsequent phases; thus, the stimuli used in training and testing were tailored for each participant. The second phase involved a conditional discrimination procedure, arranged with an OTM training structure. Groups 1 and 2 trained on three 3-member classes (A1B1C1, A2B2C2, and A3B3C3). Group 3 trained on three 5-member classes (A1B1C1D1E1, A2B2C2D2E2, and A3B3C3D3E3). The category stimuli were designated as A1 (less than 20), A2 (20–40), and A3 (more than 20), respectively, while the B, C, D, and E stimuli represented different food items. The mastery criterion was 95% correct for all groups. The third phase consisted of testing, where all groups were presented with two test blocks. The test blocks for Groups 1 and 2 consisted of 54 trials, while the test block for Group 3 consisted of 180 trials. The test criterion was a minimum of 90% correct trials for all groups. For Groups 1 and 2, this meant a minimum of 49/54 correct trials, while for Group 3, it was a minimum of 162/180 correct trials. The fourth phase involved computerized post-sorting of the stimuli used in conditional discrimination training and testing, comprising nine stimuli for Groups 1 and 2, and 15 stimuli for Group 3. The results showed that six out of eight participants in Group 1, all participants in Group 2, and five out of six participants in Group 3 formed equivalence classes. All participants in Groups 1 and 2 correctly sorted the stimuli into their respective categories. In Group 3, all participants sorted correctly, except for the participant who did not demonstrate equivalence on the MTS test. Arntzen and Eilertsen suggest that further studies should include follow-up tests to examine the maintenance of equivalence classes over time.

Regaço et al. (2023) had 24 publications in their review on the maintenance of stimulus equivalence classes focusing on parameters as class size, training and testing parameters, the nature of stimuli, generalized relations, and equivalence-based instruction. The review highlights that multiple factors contribute to the maintenance of these classes, such as class size, training structures, and types of stimuli. The authors advocate for more focused research on these variables to better comprehend and enhance the durability of learned behaviors in both experimental and practical settings. Further stud-

ies are necessary to investigate the maintenance of derived repertoires following the application of a stimulus equivalence procedure.

Díaz et al. (2023) conducted a review on the use of EBI for learning portion sizes and nutritional content in food. One topic emphasized in the review was the necessity for more research focusing on retention and long-term follow-up. In their sample of reviewed articles, only two studies examined the maintenance of results—Hausman et al. (2014) addressed it after one week, and Trucil et al. (2015) after one and two weeks. Therefore, further investigations are needed to verify the effects of forming equivalence classes based on nutritional information on food categorization and choice preferences over different time intervals.

The goal of the present study was to replicate and extend Arntzen and Eilertsen (2020) by using a many-to-one instead of a one-to-many training structure when testing for the formation of small and large equivalence classes. Another extension was to include maintenance tests after 10 days and 17 days of class formation in the MTS test and the sorting test. Finally, we extended the study by increasing the number of stimuli in the sorting (pre- and post-tests). In the pre-test, we increased the number of stimuli to 60 compared to 21 and 33 stimuli, while in the post-test, we presented all 60 stimuli compared to 9 and 15 stimuli in the Arntzen and Eilertsen study.

Method

Participants

Twenty-six participants, all neurotypical, were recruited from a student pool of undergraduates at different programs at the university and through personal contacts. None of the participants reported having previously completed coursework on the foundational knowledge of carbohydrate content in food items. The group consisted of eight men and eighteen women, aged 20 to 54 years ($M=31.0$ years old, $SD=9.6$). Participants were randomly assigned to one of three groups: 10 in Group 1, 9 in Group 2, and 8 in Group 3. The availability of participants determined the number of participants in each group. None of the participants had prior

knowledge of stimulus equivalence. Before the study commenced, all participants read and signed a consent form. Upon completion, they were compensated with universal gift cards worth NOK 200.

Apparatus and Stimuli

The experimental sessions took place in a university laboratory within small cubicles designed for testing. Each cubicle, approximately 1.5 m² in size, was equipped with a chair, a table, and a computer featuring a 17-inch screen. Custom-built MTS program, running on Windows 10, presented all trials and recorded responses.

The stimuli used in the experiment are listed in Table 1. In the MTS program, each word and number was displayed as a black image on a white background within a 5 x 5 cm “invisible” square. The categories “Less than 20,” “20-40,” and “More than 40” were consistently used as A stimuli. The remaining stimuli were arranged as B and C stimuli for Groups 1 and 2, and as B, C, D, and E stimuli for Group 3, depending on the tailored selection (see Pre-sorting below). Participants used a mouse to click on the stimuli according to specific instructions.

In addition, 60 laminated cards representing food products and 4 laminated cards representing categories were used in a table-top arrangement. The food cards varied in size from 5 x 3 cm to 5.5 x 3.5 cm. The categories were labeled as “Less than 20,” “20-40,” “More than 40,” and “Don’t Know.” These category cards were always placed in a fixed order, with “Less than 20” at the top of the pile, followed by “20-40,” “More than 40,” and, for Groups 2 and 3, “Don’t Know.” The other 60 cards were shuffled and presented in a pile in the middle of the table below the category cards.

Design

We employed a group design, with Groups 1 and 2 receiving training on six food items, while Group 3 received training on twelve food items. Participants in Group 1 sorted items into three categories: “More than 40,” “20-40,” and “Less than 20.” In Groups 2 and 3, participants also had the option of sorting items into a fourth category: “Don’t Know.” The experimental arrangement included two tests for maintenance, one (Follow-up 1) and three (Follow-up 2) weeks after training and testing.

Since the primary focus in the present experiment was tailoring the stimuli for each participant, the dependent variables measured were the number of incorrectly sorted stimuli in the pre-sorting phase. Second, the number of trials required to reach the mastery criterion during conditional discrimination training. Third, we calculated the percentage of correct responses by the total number of trials during tests for equivalence relations. The fourth dependent variable was the number of correctly sorted stimuli during the post-sorting phase. The scoring in pre- and post-sorting phases of the stimuli used the same

Procedure

All groups were exposed to all phases. The procedure included eight phases (see Figure 1): (1) Pre-sorting of stimuli, (2) conditional discrimination training, (3) tests for emergent relations, (4) post-sorting of stimuli, (5-6) two follow-up tests for emergent relations, and (7-8) two follow-up sorting tests. If participants did not meet the test criterion during the follow-up tests, they received additional training and testing. Phases 1-4 were run successively in one session.

Phase 1: Pre-sorting and Tailoring of Stimuli

To be able to use stimuli that were not already established in class with the correct category, we used a pre-sorting phase for categorization. Participants were provided with a randomized stack of laminated cards representing food items, while the category cards remained fixed at the top. In Groups 2 and 3, the “Don’t Know” category was also included. Participants were instructed in Norwegian to “sort these cards as you wish. If you have any questions, I am sitting outside the room. Let me know when you are finished.”

After sorting, the second author documented the card positions through photographs. Incorrectly sorted cards with the names of the food items were selected for further training and testing—six items for Groups 1 and 2, and twelve for Group 3. For Groups 1 and 2, two cards from each category, and for Group 3, four cards from each category. After the stimuli were randomly selected, they were checked for similarity. If they belonged to the same group (for example, boiled basmati rice, cooked pasta, and cooked couscous), other stimuli were selected.

One participant was excluded because all items within one category were correctly sorted, and the participant was not replaced by another participant. After finishing the pre-sorting phase, the participants were exposed to the conditional discrimination training.

Phase 2: Conditional Discrimination Training

Participants completed MTS training, in which one sample stimulus appeared in the center of the screen, with three comparison stimuli displayed in the corners of the screen. The sample stimulus remained visible throughout the trial, utilizing a Simultaneous Matching-to-Sample (SMTS) protocol. The participant clicked one of the comparison stimuli, which was followed by a programmed consequence. If the choice was correct (for example, clicking on C1 and not C2 or C3 in the presence of A1), text such as 'awesome,' 'correct,' 'super,' etc., appeared on the screen, while for incorrect choices, the text 'incorrect' (for example, clicking on C3 or C1 and not C1 in the presence of B1) was displayed. The programmed consequences were displayed for 500 ms.

The following instruction in Norwegian was presented on the screen:

You will need to click on some stimuli that appear on the screen. The goal is to get as many correct as possible. When you move the mouse pointer to the stimulus in the center and click on it, more stimuli will appear on the screen. A mouse click on the correct stimulus in the corners will be followed by the text 'Correct' or something similar on the screen. Clicking on one of the incorrect ones will be followed by the text 'Incorrect.' This is how you find out what is correct and incorrect. After a while, you will no longer receive feedback on whether what you click is right or incorrect. It will always be necessary to click on the stimulus in the center before you click on the ones in the corners. Click 'Start' to begin the experiment.

An MTO structure was used, with baseline trials presented in serialized blocks (see Table 2). For Groups 1 and 2, three 3-member classes, A1B1C1, A2B2C2, and A3B3C3 were classes to-be formed. AC and BC trials were presented separately, then

mixed. For Group 3, three 5-member classes, A1B1C1D1E1, A2B2C2D2E2, and A3B3C3D3E3 were classes to-be formed. AE, BE, CE, and DE trials were trained in separate blocks, then mixed. The mastery criterion was set at 95% correct responses within the block before proceeding to the next teaching phase. If participants did not meet the mastery criterion, the block was repeated. The probability of programmed consequences was 100% when the relations were trained separately. The probability of programmed consequences was decreased gradually from 100% to 75%, 50%, and finally 0%, depending on meeting the mastery criterion as described above. Each trial type was presented five times per block, with a 500 ms intertrial interval. The test phase was implemented automatically when the participants met the mastery criterion.

Phase 3: Test for Emergent Relations

The test for emergent relations was conducted in an SMTS format with no programmed consequences and a 500 ms intertrial interval. Each trial type was presented five times. As illustrated in Table 2, Groups 1 and 2 had 90 total trials (30 trials each for baseline, symmetry, and equivalence), while Group 3 had 300 total trials (60 trials each for baseline and symmetry, and 180 equivalence trials). The experimenter-defined criterion for forming an equivalence class was 90% correct responses on all relations. This means that even if the average of the scores is 90% or above, it will not qualify for the experimenters' definition if the scores for one or two of the relation types are below 90% correct.

Phase 4: Post-sorting

Immediately after Phase 3, participants repeated the sorting task with the laminated cards, following the same procedure as in Phase 1. After completing this, participants were informed they had finished for the day.

Phase 5: Follow-up 1: Test for Emergent Relations

Seven to ten days after Phase 3, participants returned for maintenance testing. The test was identical to the test in Phase 3, and the participants were given the following instruction:

The goal is to get as many answers as possible correct. When you move the mouse pointer to the stimulus in the center and click on it, more stimuli will appear on the screen. You will not receive feedback on what is correct or incorrect. It will always be necessary to click on the stimulus in the center before clicking on the ones in the corners. Click 'Start' to begin the experiment.

If they met the 90% criterion, they advanced to Phase 6. If not, they received additional training and testing, identical to Phases 2 and 3.

Phase 6: Follow-up 1: Sorting

This phase mirrored the sorting tasks from Phases 1 and 4. Participants completed the sorting and were then informed they were done for the day.

Phase 7: Follow-up 2: Test for Emergent Relations

Fourteen to seventeen days after Phase 6, participants returned for the second follow-up test, conducted identically to the first follow-up.

Phase 8: Follow-up 2: Sorting

Following the second follow-up test, participants completed another sorting task, with the same procedure as in previous phases. After completing this final task, participants were debriefed and could view their results if desired.

Procedural Failures

Due to a programming error, baseline relations were not tested for P18953 during the post-test. Additionally, P18955 did not receive extra training during Follow-up 1, despite not meeting the test criterion for equivalence relations. However, during Follow-up 2, the participant reached the test criterion without the additional training.

Interobserver Agreement (IOA)

The second author and a trained observer scored 30% of all sorting sessions. Interobserver agreement was calculated using the formula: $(\text{Agreements} / (\text{Agreements} + \text{Disagreements})) \times 100$ (e.g., Kazdin, 2010). IOA was 100% for pre-sorting, 90% for post-sorting, 97% for Follow-up 1 sorting, and 100% for Follow-up 2 sorting.

Results

Sorting and Tailoring Stimuli

Across all groups, the majority of food items were incorrectly sorted during the pre-sorting phase. The food items selected for training and testing were those incorrectly sorted by each individual participant. Thus, the stimuli used in training and testing were tailored for each participant. The number of times a specific stimulus was used in training and testing is listed in Table 1. Five participants in Group 2 sorted one or more stimuli in the "don't know" pile, while three of the participants in Group 3 did so.

Training Trials Blocks

In Group 1, participants required an average of 12.4 blocks (range: 8–18, $SD=3.9$) to meet the mastery criterion (see Figure 2). In Group 2, participants averaged 11 blocks (range: 8–27, $SD=6.1$). A t-test revealed no statistically significant difference in the number of training blocks ($p=0.56$). In Group 3, participants had an average of 22.6 blocks (range: 20–27, $SD=2.3$) to meet the mastery criterion. From Group 1, P18981 required additional training, completing 11 blocks. From Group 2, P18957 and P18980 also received additional training, completing 6 and 7 blocks, respectively. In Group 3, P18962 required additional training and required 18 blocks to meet the mastery criterion.

Equivalence Class Formation

Participants were tested for equivalence class formation after reaching mastery in conditional discriminations, followed by two maintenance tests in Follow-up 1 and Follow-up 2.

Immediate Test

Immediately after training, 89% (8/9) of the participants in Group 1, 100% (9/9) of the participants in Group 2, and 75% (6/8) of the participants in Group 3 formed equivalence classes (see the upper panels in Figures 3, 4, and 5 for individual data). A Mann-Whitney test did not reveal a statistically significant difference when comparing Groups 1 and 2 to Group 3 ($p = 0.17$).

Follow-up Tests

In the Follow-up 1 test (see the middle panels of Figures 3, 4, and 5), one additional participant formed equivalence classes after extra training; thus, 100% of participants formed equivalence classes in Group 1. In Group 2, seven participants maintained their equivalence classes, while two participants, after receiving additional training, responded in accordance with stimulus equivalence in the subsequent test; thus, 100% of the participants formed classes. In Group 3, five participants maintained their equivalence classes, and three participants who had not previously formed equivalence classes successfully did so. Additionally, one participant formed classes after receiving further training; thus, 100% of participants formed classes.

In the Follow-up 2 test (see the bottom panels of Figures 3, 4, and 6), 100% of participants in Groups 1 and 2 maintained equivalence classes. In Group 3, 88% (7/9) of the participants maintained their classes.

Sorting Tests

In the experiment's setup, with tailored or customized stimuli based on the pre-test sorting, participants across all groups scored 0% correct on categorizing the stimuli used in the conditional training and testing.

Immediate Test

During the post-test sorting, 78% (7/9) of participants in Group 1, 89% (8/9) of participants in Group 2, and 50% (4/8) of participants in Group 3 sorted the cards correctly into the experimenter-defined classes (see the upper panels in Figures 3, 4, and 5 for individual data).

Follow-up Tests

In the first follow-up test (see the middle panels in Figures 3, 4, and 5 for individual data), 89% (8/9) of the participants in Group 1, 78% (7/9) of the participants in Group 2, and 50% (4/8) of the participants in Group 3 sorted all items correctly. For the Follow-up 2 sorting (see the lower panels in Figures 3, 4, and 5 for individual data), 89% (8/9) of the participants in Groups 1 and 2, and 63% (5/8) of the participants in Group 3 sorted correctly. See Table 3 for an overview.

Sorting Not in Experimenter-Defined Classes

Sorting could either be scored as categorization in accordance with experimenter-defined classes or scored as not in accordance with these defined classes, but the scoring could also be based on participant-defined classes. Five of the participants (P18981 from Group 1; P18980 and P18957 from Group 2; P18962 and P18975 from Group 3) made one or more errors. P18980, P18957, and P18962 sorted one or more stimuli in "don't know" category. P18960 from Group 3 sorted the stimuli according to participant-defined categories in all sorting tests, while three participants (P18979 from Group 1; P18967 from Group 2; P18958 from Group 3) sorted partly according to participant-defined classes. P18958 also sorted one or more stimuli in the "don't know" category.

Discussion

This study aimed to establish knowledge about carbohydrates using EBI procedures and investigate the maintenance of emergent relations in two follow-up tests using MTS and sorting tests. The first MTS test after the conditional discrimination training showed that 88,9%, 100%, and 75% of the participants in Groups 1, 2, and 3, respectively, formed equivalence classes. In the first follow-up test, 100% of participants in all groups formed equivalence classes. Four of the participants required additional training to form classes. In the second follow-up test, all participants in Groups 1 and 2, and all but one participant in Group 3, formed equivalence classes. In experiments comparing results on the MTS and sorting tests, the sorting test includes the same stimuli as in the MTS training and test (e.g., Arntzen et al., 2021). However, in the present experiment, the sorting post-test included all 60 stimuli used in the pre-test. The results of the sorting test, conducted after the MTS test, showed that 78%, 90%, and 50% of the participants in Groups 1, 2, and 3, respectively, sorted the stimuli according to the experimenter-defined classes. Furthermore, 89%, 78%, and 50% of the participants in Groups 1, 2, and 3, respectively, sorted the stimuli in experimenter-defined classes in the first follow-up test. Finally, 89% of the participants in Groups 1 and

2, and 63% of the participants in Group 3, sorted stimuli into experimenter-defined classes in the second follow-up test.

Tailoring of Stimuli

An important aspect of the present study was the tailoring of stimuli based on the results of the pre-sorting test. Participants were presented with 60 stimuli, while the conditional discrimination training and testing for emergent relations involved only the stimuli that the participant had not correctly sorted in the pre-test. This tailoring of stimuli in the present experiment ensured that the stimuli used in the conditional discrimination training were not already part of the correct category of carbohydrates. Tailoring of stimuli could be an essential variable to consider in EBI experiments since the type of stimuli used in EBI research differs significantly from those used in basic research. In basic research within stimulus equivalence studies, abstract shapes are typically used, and the relations between stimuli are arbitrarily defined (e.g., Sidman & Tailby, 1982). Therefore, to ensure that the stimulus set used in conditional discrimination training is not already partitioned into different classes, it is essential to tailor the stimulus set for each participant. The tailoring procedure used in the present experiment aligns with other experiments exploring nutrition skill teaching (Arntzen & Eilertsen, 2020; Nastally et al., 2010; Oenema et al., 2001).

Replication

MTS Training Trials Blocks

Participants in Groups 1 and 2 had a similar number of blocks to reach mastery, while those in Group 3 required more than twice as many blocks compared to the other groups. This difference is likely due to participants in Groups 1 and 2 training with three 3-member classes, whereas those in Group 3 trained with three 5-member classes. When considering this variation, all groups approximately utilized the same number of blocks to establish relations. The number of blocks in this study replicates the findings from Arntzen and Eilertsen (2020).

MTS Test

The results of the emergent relations test, conducted after the conditional discrimination training,

replicated those of Arntzen and Eilertsen (2020). In Arntzen and Eilertsen (2020), 100 % of participants in Groups 1 and 2, and 83.3% in Group 3 demonstrated the emergence of equivalence relations, while in the current study, 88.9% of participants in Group 1, 100% in Group 2, and 75% in Group 3 formed equivalence classes. The small difference between the two experiments in the number of participants who responded in accordance with equivalence (yields) could be due to the difference in training structures (OTM vs MTO). Research has shown that there are slight differences in outcome on MTS tests between OTM and MTO training structures (For an overview, see Arntzen, 2012). Furthermore, the results on equivalence class formation align with other studies focusing on EBI, such as Trucil et al. (2015) and Vladescu et al. (2021), in which all participants formed equivalence classes. Additionally, the present experiment corresponds with results on equivalence class formation from experiments using abstract shapes using the MTO training structure (e.g., Ayres-Pereira & Arntzen, 2021). Finally, the yields for small classes versus larger classes replicates experiments with abstract shapes, which are in accordance with previous research (e.g., Arntzen & Hansen, 2011; Fields et al., 1999).

Sorting Test

The sorting results following the MTS test support the findings of Arntzen and Eilertsen (2020) and those of Nastally et al. (2010). In all three studies, most participants demonstrated improvement from pre-sorting to post-sorting. In the present experiment, the average increase was 92%.

In Arntzen and Eilertsen (2020) reported a 100% correct sorting for Groups 1 and 2, and an 83.3% correct sorting for Group 3. In the current study, 77.8% of participants in Group 1, 88.9% in Group 2, and 50% in Group 3 sorted correctly in the first sorting test. Although the percentage of correct sorting was slightly lower in this study, it may be attributed to the difference in the complexity (number of discriminations) in the arrangement of the sorting tests. In both experiments, participants were asked to sort the stimuli used in the equivalence test. However, in the present experiment, all 60 stimuli were presented during the

sorting tests, while in Arntzen and Eilertsen, the post-sorting test only included stimuli that had been part of the equivalence test. In the present experiment, the participants had to sort six stimuli (Groups 1 and 2) and 12 stimuli (Group 3) according to the three categories ('below 20', '20–40', and more than '40') and to discriminate these stimuli from the other 54 stimuli (for Groups 1 and 2) and 48 stimuli (for Group 3).

Maintenance and Follow-Up Tests

Maintenance has been listed as one of the dimensions of applied behavior analysis: "A behavior change may be said to have generality if it proves durable over time, if it appears in a wide variety of possible environments, or if it spreads to a wide variety of related behaviors" (Baer et al., 1968, p. 96). Review articles on maintenance behavior analysis have been done in different areas as for example, functional analysis of problem behavior (Hanley et al., 2003), maintenance of preschool children's social skills (Chandler et al., 1992), maintenance of school-based functional behavior assessments and behavior intervention plans (Pennington et al., 2019), maintenance of behavior treatment across settings (Podlesnik et al., 2017), and long-term response maintenance in large-scale behavioral interventions for industrial safety (Boyce & Geller, 2001). Relevant to the present paper, Regaço et al. (2023) showed that even stimulus equivalence has been an active research area for more than 50 years, relatively few studies have included measures of maintenance of equivalence classes. They found in their review, six published articles that compared the effect of different class sizes in maintenance. For example, Camargo and Haydu (2015) used an MTO training structure and familiar stimuli with undergraduate students as participants. They found that 6-member classes were maintained more often than three-member classes. There are a couple of differences between Camargo and Haydu and the present experiment that could be related to the different outcomes. Camargo and Haydu employed a simple-to-complex training and testing protocol, which means that participants trained to form 6-member classes were exposed to a significantly higher number of symmetry trials than those trained to form 3-member classes, as observed in the present experiment.

Two studies, by Hausman et al. (2014) and Trucil et al. (2015), with a focus on nutrition, investigated maintenance after one and two weeks, respectively. In both studies, responses improved or remained stable compared to baseline. The current study's results are consistent with these findings, as the number of correct trials during the MTS test remained similar or improved across the follow-up tests. In the second follow-up test, there was only a minor difference between the groups (1 and 2), with three members compared to Group 3, which had five members. Three of the four participants, all from Groups 1 and 2 who did not meet the test criterion during follow-up 1, received extra training and showed maintenance of equivalence classes in the second follow-up test. The primary difference between this study and previous ones lies in the timing of the emergent relation tests, which were conducted at three distinct points: post-test, follow-up 1, and follow-up 2. Conversely, other studies mainly concentrated on post-tests, with only Trucil et al. (2015) including maintenance, but limited to portion size estimates.

Correspondence Between MTS and Sorting Tests

Another essential topic to emphasize is the correspondence between the MTS test and the sorting test. The arrangement of the sorting test in the present experiment, by including 60 stimuli and not only the stimuli used in the conditional discrimination, could influence the correspondence between MTS and the sorting test. For the participants who responded in accordance with stimulus equivalence in the MTS test (see the upper panels in Figures 3, 4, and 5), 88% (7/8) of the participants in Group 1, 89% (8/9) of the participants in Group 2, and 40% (2/5) of the participants in Group 3, respectively, sorted the stimuli into experimenter-defined classes. For participants who did not respond in accordance with stimulus equivalence on the MTS test, only two participants in Group 3 (66% or 2/3) sorted stimuli into experimenter-defined classes. In the first follow-up test (see the middle panels in Figures 3, 4, and 5), for the participants who responded in accordance with stimulus equivalence in the MTS test, 88% (7/8) of the participants in Group 1, 86%

(6/7) of the participants in Group 2, and 54% (4/7) of the participants in Group 3, respectively, sorted the stimuli into experimenter-defined classes. In the second follow-up test (see the lower panels in Figures 3, 4, and 5), for the participants who responded in accordance with stimulus equivalence in the MTS test, 89% (8/9) of the participants in Groups 1 and 2, and 71% (5/7) of the participants in Group 3 sorted the stimuli into experimenter-defined classes. For participants who did not respond in accordance with stimulus equivalence on the MTS test, none of them sorted stimuli into experimenter-defined classes.

The present experiment demonstrated a relatively high correspondence between MTS and sorting tests for both the immediate test and the follow-up tests. The correspondence was lower for Group 3 compared to Groups 1 and 2, a difference most likely due to the number of stimuli in each class. The correspondence was higher in Group 3 in the last follow-up test compared to the first test. The present experiment supports the argument that sorting is an alternative method for assessing the emergence of equivalence classes (e.g., Arntzen et al., 2021; Rustad Bevolden & Arntzen, 2018).

Limitations and Future Research

Several potential limitations should be considered. The first limitation is the variation in time intervals between maintenance and follow-up phases, which ranged from 7 to 10 days for the first test and 14 to 17 days for the second test. Further studies should impose stricter time intervals to reduce variability. The first limitation relates to the repetitive nature of MTS testing, which could lead to participant fatigue. Future studies might consider alternating between MTS and sorting tests to introduce variation. However, none of the participants in the present study reported experiencing such issues.

Another limitation is that we did not provide extra training for participants who did not meet the test criterion for emergent relations in the first MTS test. P18981 did not form equivalence classes in the first test and did not do so in the follow-up test 1. The number of correct responses in the first MTS test was relatively low; therefore, it is not surprising that the number of correct responses remained

low in the first follow-up MTS test, despite no additional training. Likewise, one of the participants in Group 3 (P18962) had a relatively low score of correct responses in the first sorting test. However, the two other participants in Group 3 (P19868 and P19855) had a number of correct responses in the first MTS test that were very close to the test criterion and showed the formation of equivalence classes in the follow-up test 1.

Finally, we did not specifically test in advance whether the participants knew the relations between the names of foods on the stimulus cards and the actual foods. However, we assume that in the present experiment, when adults served as participants all names of the food items presented on the stimulus cards were known.

Understanding carbohydrates holds significant educational and social value for populations across various age groups and cultural backgrounds. Carbohydrates, being a primary source of energy, play a central role in diet and health. Future studies could also examine the distinctions between simple and complex carbohydrates, as complex carbohydrates offer greater health benefits. Investigating participants who have a heightened focus on nutritional content, such as those following specific diets or managing lifestyle diseases like diabetes, may reveal differences in knowledge acquisition and retention.

Marin and Fienup (2024) described a hypothetical experiment with different levels of lower mastery criteria of baseline relation before testing for emergent relations. Such an arrangement could be helpful in studying the effects of retraining and retesting participants who did not form equivalence classes. Projects similar to the present experiment could, by such an arrangement, find more effective methods to form classes of food items categorized by carbohydrate levels.

Summary

This study replicated and extended Arntzen and Eilertsen (2020), incorporating additional food items during sorting. Emergent relations were consistently observed across all groups in the MTS post-tests and maintained over a three-week period. Sorting tests yielded similar results for Groups 1 and 2, with responses remaining stable

across the follow up tests. Although sorting results for Group 3 were slightly lower than those for the MTS tests, they also remained consistent over time.

References

- Albright, L., Reeve, K. F., Reeve, S. A., & Kisamore, A. N. (2015). Teaching statistical variability with equivalence-based instruction. *Journal of Applied Behavior Analysis, 48*(4), 883–894. <https://doi.org/10.1002/jaba.249>
- Arntzen, E. (2012). Training and testing parameters in formation of stimulus equivalence: Methodological issues. *European Journal of Behavior Analysis, 13*(1), 123–135. <https://doi.org/10.1080/15021149.2012.11434412>
- Arntzen, E., Dechsling, A., & Fields, L. (2021). Expansion of arbitrary stimulus classes and function-transfer measured by sorting performances. *Journal of the Experimental Analysis of Behavior, 115*(1), 326–339. <https://doi.org/10.1002/jeab.665>
- Arntzen, E., & Eilertsen, J. M. (2020). Using stimulus-equivalence technology to teach skills about nutritional content. *Perspectives on Behavior Science, 43*(3), 469–485. <https://doi.org/10.1007/s40614-020-00250-2>
- Arntzen, E., & Hansen, S. (2011). Training structures and the formation of equivalence classes. *European Journal of Behavior Analysis, 12*(2), 483–503. <https://doi.org/10.1080/15021149.2011.11434397>
- Ayres-Pereira, V., & Arntzen, E. (2021). A descriptive analysis of baseline and equivalence-class performances under many-to-one and one-to-many structures. *Journal of the Experimental Analysis of Behavior, 115*(2), 540–560. <https://doi.org/10.1002/jeab.678>
- Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1*, 91–97. <https://doi.org/10.1901/jaba.1968.1-91>
- Bolanos, J. E., Reeve, K. F., Reeve, S. A., Sidener, T. M., Jennings, A. M., & Ostrosky, B. D. (2020). Using stimulus equivalence-based instruction to teach young children to sort recycling, trash, and compost items. *Behavior and Social Issues, 29*, 78–99. <https://doi.org/10.1007/s42822-020-00028-w>
- Boyce, T. E., & Geller, E. S. (2001). Applied behavior analysis and occupational safety: The challenge of response maintenance. *Journal of Organizational Behavior Management, 21*, 31–56. https://doi.org/10.1300/J075v21n01_03
- Camargo, J., & Haydu, V. B. (2015). Class size effects on the establishing, maintenance, and recovery of equivalence relations. *Behavior Analysis: Research and Practice, 15*(1), 36–49. <https://doi.org/10.1037/h0101068>
- Chandler, L. K., Lubeck, R. C., & Fowler, S. A. (1992). Generalization and maintenance of preschool children's social skills: A critical review and analysis. *Journal of Applied Behavior Analysis, 25*(2), 415–428. <https://doi.org/10.1901/jaba.1992.25-249>
- Critchfield, T. S., & Fienup, D. M. (2010). Using stimulus equivalence technology to teach statistical inference in a group setting. *Journal of Applied Behavior Analysis, 43*(4), 763–768. <https://doi.org/10.1901/jaba.2010.43-763>
- Díaz, M., Garcia, Y., & Mahoney, A. (2023). A review on equivalence-based procedures to teach food-portion estimation skills. *Behavioral Interventions, 38*(3), 477–493. <https://doi.org/10.1002/bin.1934>
- Fields, L., Hobbie-Reeve, S. A., Adams, B. J., & Reeve, K. F. (1999). Effects of training directionality and class size on equivalence class formation by adults. *The Psychological Record, 49*, 703–724. <https://doi.org/10.1007/BF033>
- Fienup, D. M., Covey, D. P., & Critchfield, T. S. (2010). Teaching brain-behavior relations economically with stimulus equivalence technology. *Journal of Applied Behavior Analysis, 43*(1), 19–33. <https://doi.org/10.1901/jaba.2010.43-19>
- Green, G., & Saunders, R. R. (1998). Stimulus equivalence. In K. A. Lattal & M. Perone (Eds.), *Handbook of research methods in human operant behavior* (pp. 229–262). Plenum Press.
- Hanley, G. P., Iwata, B. A., & McCord, B. E. (2003). Functional analysis of problem behavior: A review. *Journal of Applied Behavior Analysis, 36*(2), 147–185. <https://doi.org/10.1901/jaba.2003.36-147>
- Hausman, N. L., Borrero, J. C., Fisher, A., & Kahng,

- S. (2014). Improving accuracy of portion-size estimations through a stimulus equivalence paradigm. *Journal of Applied Behavior Analysis*, 47(3), 485–499. <https://doi.org/10.1002/jaba.139>
- Hausman, N. L., Borrero, J. C., Fisher, A., & Kahng, S. (2017). Teaching young children to make accurate portion size estimations using a stimulus equivalence paradigm. *Behavioral Interventions*, 32(2), 121–132. <https://doi.org/10.1002/bin.1466>
- Hill, K. E., Griffith, K. R., & Miguel, C. F. (2020). Using equivalence-based instruction to teach piano skills to children. *Journal of Applied Behavior Analysis*, 53(1), 188–208. <https://doi.org/10.1002/jaba.547>
- Kazdin, A. E. (2010). *Singe-Case Research Designs* (2nd ed.). Oxford University Press.
- LaFond, T. R., Reeve, K. F., Day-Watkins, J., Reeve, S. A., Vladescu, J. C., & Jennings, A. M. (2021). Using stimulus equivalence-based instruction to teach young children their caregivers' contact information. *Behavioral Interventions*, 36(1), 105–125. <https://doi.org/10.1002/bin.1742>
- Marin, R., & Fienup, D. M. (2024). Relating in the Wild: Toward an Analysis of Equivalence Relations Under More Naturalistic Conditions. *Perspectives on Behavior Science*, 47(3), 603–626. <https://doi.org/10.1007/s40614-024-00420-6>
- Nastally, B., Dixon, M. R., McKeel, A., & Fleming, M. (2010). Teaching healthy food choices through stimulus equivalence. *Journal of Behavioral Health and Medicine*, 1(2), 119–126. <https://doi.org/10.1037/h0100545>
- Ong, T., Normand, M. P., & Schenk, M. J. (2018). Using equivalence-based instruction to teach college students to identify logical fallacies. *Behavioral Interventions*, 33(2), 122–135. <https://doi.org/10.1002/bin.1512>
- Pennington, B., Simacek, J., McComas, J., McMaster, K., & Elmquist, M. (2019). Maintenance and Generalization in Functional Behavior Assessment/Behavior Intervention Plan Literature. *Journal of Behavioral Education*, 28(1), 27–53. <https://doi.org/10.1007/s10864-018-9299-6>
- Podlesnik, C. A., Kelley, M. E., Jimenez-Gomez, C., & Bouton, M. E. (2017). Renewed behavior produced by context change and its implications for treatment maintenance: A review. *Journal of Applied Behavior Analysis*, 50(3), 675–697. <https://doi.org/10.1002/jaba.400>
- Regaço, A., Zapparoli, H. R., Aggio, N. M., Silveira, M. V., & Arntzen, E. (2023). Maintenance of Stimulus Equivalence Classes: A Bibliographic Review. *The Psychological Record*, 73(1). <https://doi.org/10.1007/s40732-023-00535-3>
- Regan, B., Vladescu, J. C., Reeve, K. F., & DeBar, R. M. (2018). Equivalence-based instruction with non-food items to increase portion-size estimation accuracy. *The Psychological Record*, 68(4), 537–547. <https://doi.org/10.1007/s40732-018-0299-7>
- Rustad Bevolden, K., & Arntzen, E. (2018). How robust is the correspondence between responding on matching-to-sample tests and sorting tests? *Norwegian Journal of Behavior Analysis*, 45(2), 125–142. <https://nta.atferd.no/journalissue.aspx?IdDocument=729>
- Sidman, M. (1992). Equivalence relations: Some basic considerations. In S. C. Hayes & L. J. Hayes (Eds.), *Understanding verbal relations* (pp. 15–27). Context Press.
- 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. <https://doi.org/10.1901/jeab.1982.37-5>
- Trucil, L. M., Vladescu, J. C., Reeve, K. F., DeBar, R. M., & Schnell, L. K. (2015). Improving portion-size estimation using Equivalence-Based Instruction. *The Psychological Record*, 65(4), 761–770. <https://doi.org/10.1007/s40732-015-0146-z>
- Vladescu, J. C., Marano, K. E., & Reeve, K. F. (2021). Equivalence-based instruction to improve portion-size estimation with different vessels. *The Psychological Record*, 71(1), 179–188. <https://doi.org/10.1007/s40732-020-00388-0>

Histórico do Artigo

Submetido em: 11/02/2025

Aceito em: 13/08/2025

Editor Associado: Ramon Marin

Table 1. Overview of the Stimuli and Number of times stimuli were used for conditional discrimination training and testing

1	No. of times	2	No. of times	3	No. of times
Less than 20		20-40		More than 40	
Thousand Island max 3% fat	10	Crispy fried onions	9	Egg noodles	11
Grilled sausage	8	Chili explosion	6	Raw quinoa	8
Peanut butter	6	Ice crème strawberry	6	Raisins	7
Fishing sticks	5	Pizza Grandiosa Original	6	Coarse rolls	5
Banana	5	Boiled basmati rice	5	Polar bread	5
Liver pâté	5	Chilli nuts	5	Wasa smallholder crispbread	5
Mustard	5	Brown cheese G35	4	Cornflakes	4
Cottage cheese	4	French fries (frozen)	4	Rice cakes	4
Whipped cream (37% fat)	4	Cooked pasta	4	Sweet chili sauce	4
Blueberry	3	Taco spice mix	4	Special K Breakfast Cereals	3
Canned chickpeas	3	Dark chocolate (70%)	3	White pepper	3
Fresh garlic	3	Cooked couscous	3	AXA fruit muesli	2
Grape	2	Tomato paste	3	Garlic powder	2
Nestea iced tea	1	BigOne classic	2	Freia milk chocolate	2
Carrot	1	Oat bread	2	Oatmeal	2
Potato	1	Ketchup	1	Taco shells	2
Semi-skimmed milk	1	Cashews	1	Popcorn	1
Almonds	0	Pancakes	1	Tortilla	1
Brie cheese	0	Toro Waffles	0	After Eight	0
Raspberry	0	Baking powder	0	Salt sticks	0

Note. A “Don’t Know” stimulus was available for participants in Conditions 2 and 3.

Table 2. Overview of Parameters for Trained and Tested Relations Groups 1 and 2

	Relations and trials per block	Mastery and test criterion (%)	Probability of programmed consequences (%)
First relation	15 AC trials	95	100
Second relation	15 BC trials	95	100
Mix	30 AC and BC trials	95	100, 75, 50, 0
Testing	30 baseline, 30 symmetry, and 30 equivalence trials AC, BC, CA, CB, AB, BA	Minimum 100 of each relationship type	0

Table 3. Overview of Parameters for Trained and Tested Relations Group 3

First relation	15 AE trials	95	100
Second relation	15 BE trials	95	100
Mix 1	30 AE and BE trials	95	100, 75, 50, 0
Third relation	15 CE trials	95	100
Mix 2	45 AE, BE, and CE trials	95	100, 75, 50, 0
Fourth relation	15 DE trials	95	100
Mix 3	60 AE, BE, CE, and DE trials	95	100, 75, 50, 0
Testing	60 baseline, 60 symmetry, and 180 equivalence trials AE, BE, CE, DE, EA, EB, EC, ED, AB, AC, AD, BA, BC, BD, CA, CB, CD, DA, DB, DC	Minimum 100 of each relationship type	0

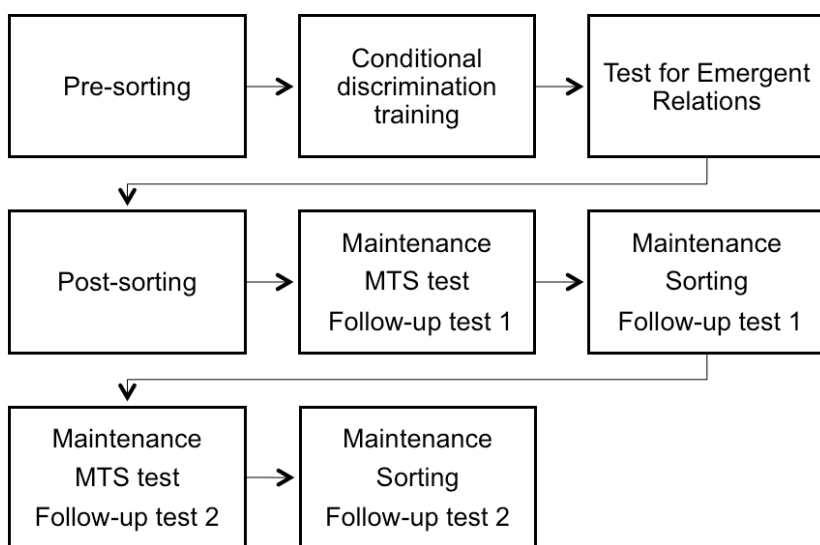


Figure 1. Flowchart of the Training and Test Phases

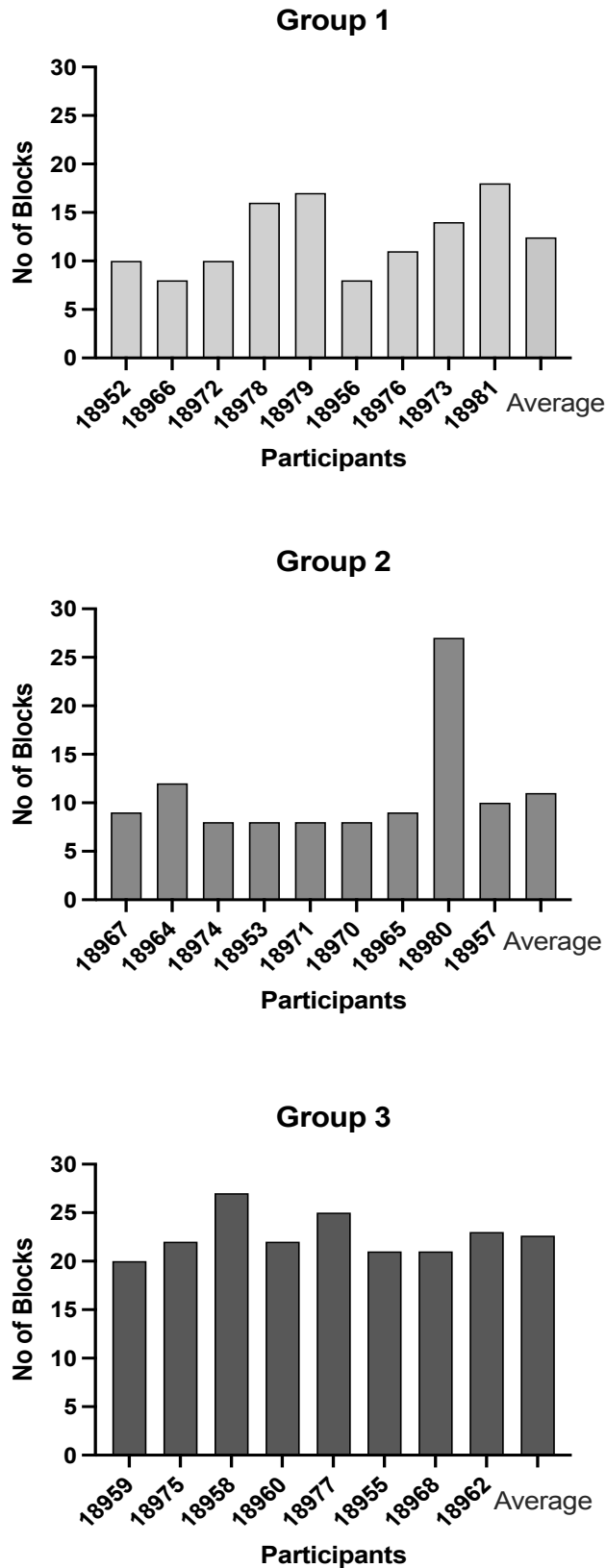
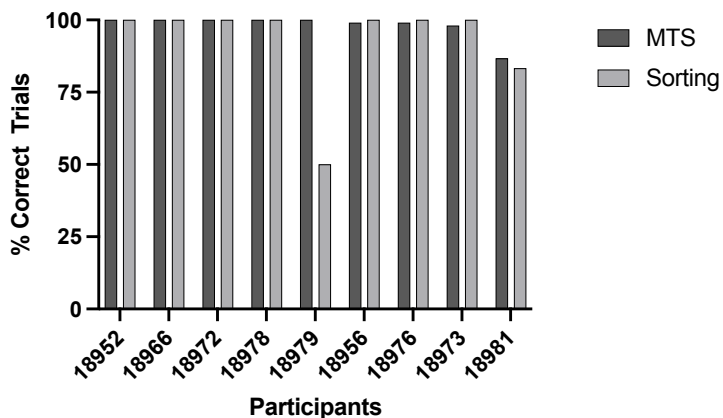


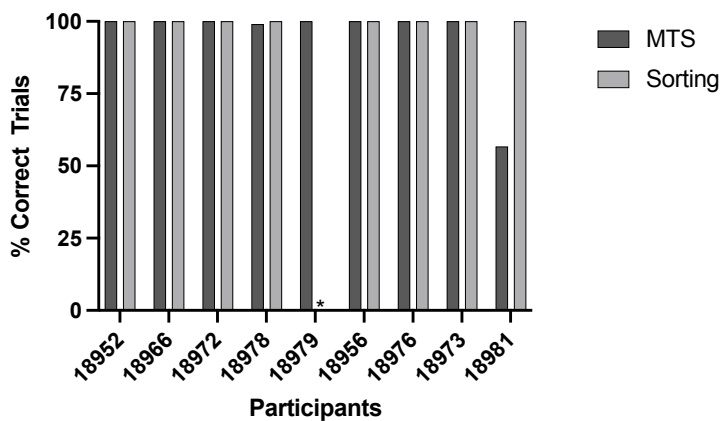
Figure 2

Note. The right most bar in all panels is the average score of training blocks

Tests After Conditional Discrimination Training



Follow-up 1



Follow-up 2

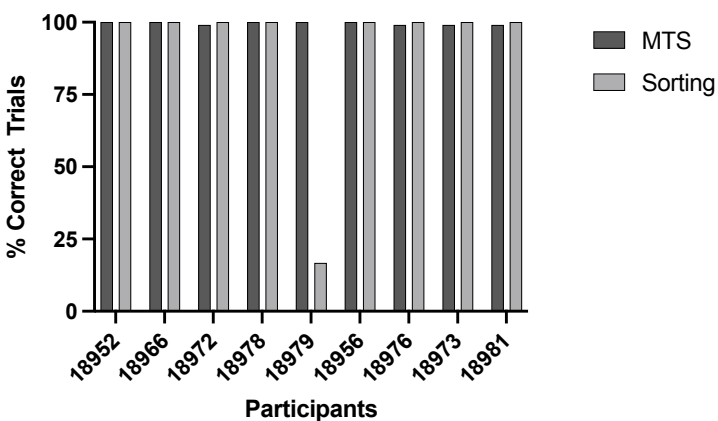
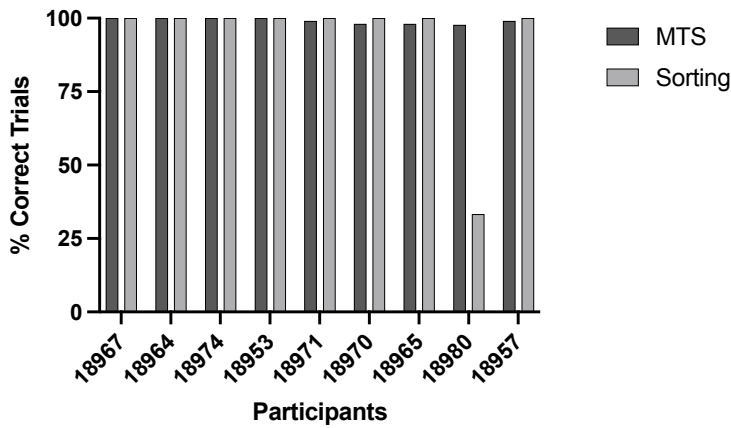


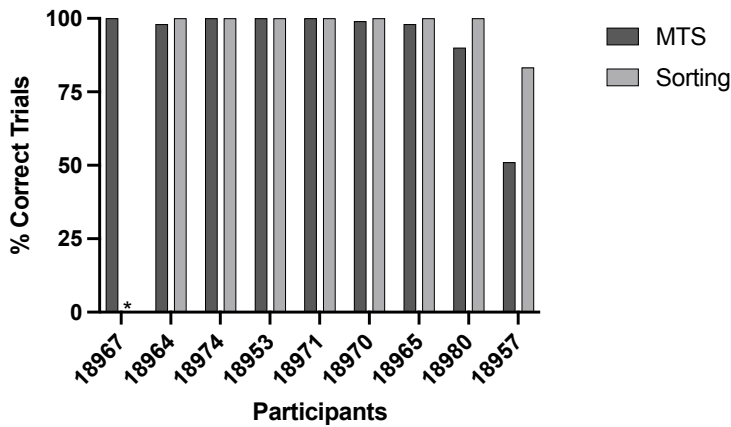
Figure 3. MTS and Sorting Scores for Group 1

Note. The figure illustrates the percentage of correct trials in the MTS and sorting tests. * indicates zero correct trials. #18981 received remedial training and passed the first follow-up test.

Tests After Conditional Discrimination Training



Follow-up 1



Follow-up 2

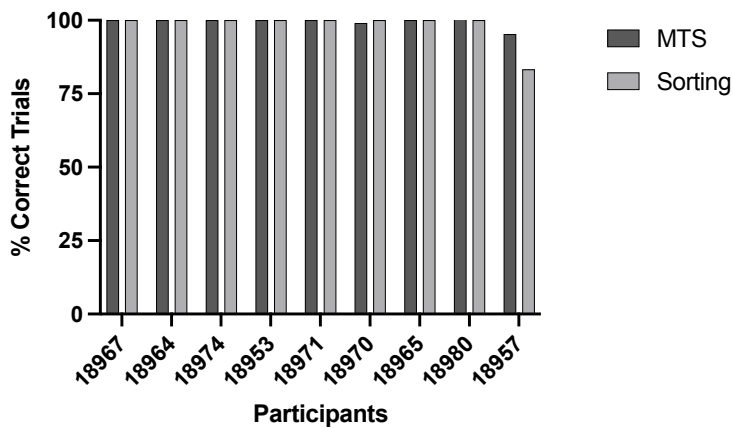
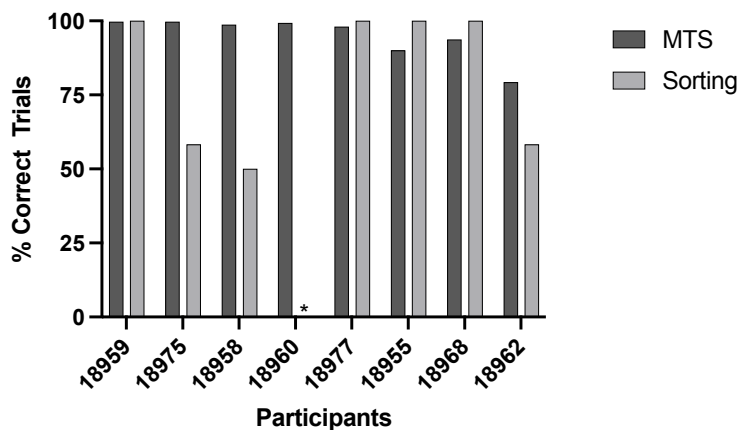


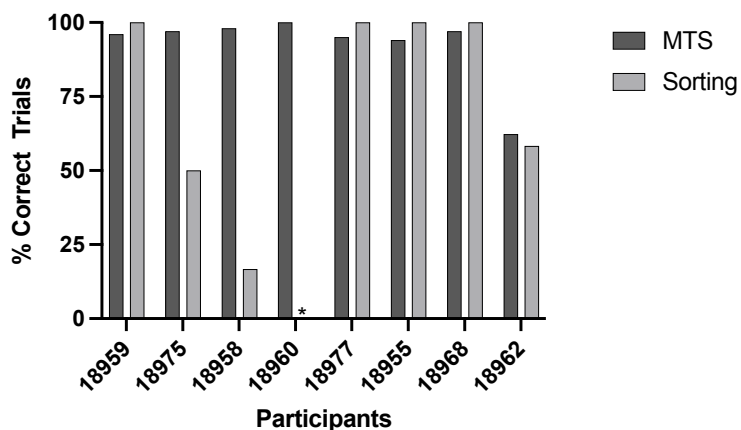
Figure 4. MTS and Sorting Scores for Group 2

Note. The figure illustrates the percentage of correct trials in the MTS and sorting tests. * indicates zero correct trials. #18957 and 18980 received remedial training and passed the first follow-up test.

Tests After Conditional Discrimination Training



Follow-up 1



Follow-up 2

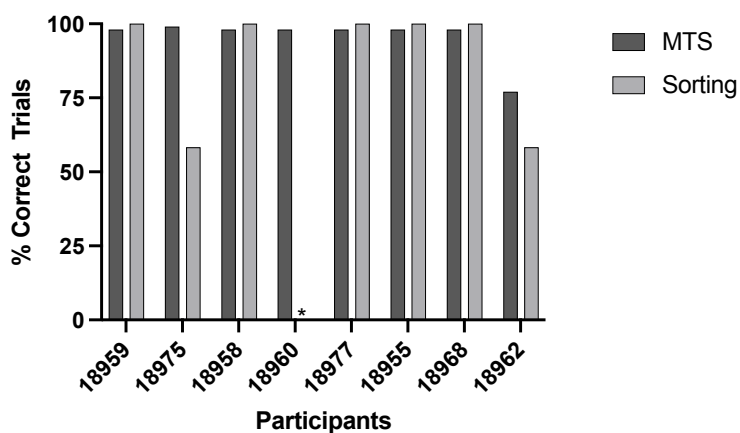


Figure 5. MTS and Sorting Scores for Group 3

Note. The figure illustrates the percentage of correct trials in the MTS and sorting tests. * indicates zero correct trials. #18962 received remedial training and passed the first follow-up test.