Preprints.org

Article

The Role of Equivalence-Based Procedures in Establishing Knowledge of Nutritional Content

Erik Arntzen [*] and Ina Charlotte Spilling

Article

The Role of Equivalence-Based Procedures in Establishing Knowledge of Nutritional Content

Erik Arntzen and Ina Charlotte Spilling *

Oslo Metropolitan University, Norway

*   Correspondence: erik.arntzen@equivalence.net

Abstract: The purpose of the study was to train participants on the nutritional values (carbohydrates) of various food items and evaluate the effect over three weeks. Twenty-seven participants were assigned to three different conditions. Initially, participants pre-sorted cards labeled with food items and categories, and any incorrectly sorted items were identified and customized for each participant. Tailored stimuli were then used for many-to-one conditional discrimination training. Conditions 1 and 2 aimed to establish three-member classes, while condition 3 aimed for five-member classes. Participants in all conditions sorted food items into categories of "more than 40," "20–40," and "less than 20" grams of carbohydrates. Additionally, participants in conditions 2 and 3 could sort items into a "don't know" category. Maintenance of the learned categories was tested after one week. Participants who did not meet the mastery criterion received additional training and testing. A follow-up test was conducted two weeks after the maintenance test. During both the maintenance and follow-up phases, tests for emergent relations and sorting were conducted. The results showed that most participants responded correctly in both the emergent relations and sorting tests. The majority of participants did not require additional training or testing, and the equivalence classes were maintained across all conditions over the three-week period.

Keywords: carbohydrates; follow-up; nutritional values; sorting; stimulus equivalence; tailoring of stimuli

1. The Role of Equivalence-Based Procedures in Establishing Knowledge of Nutritional Content

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. There are several training structures used to establish conditional discriminations: (a) Linear Series (LS), (b) One-to-Many (OTM), and (c) Many-to-One (MTO). These structures differ in how stimuli are connected to the nodes. A node is as stimulus that connect at least two other stimuli. LS includes at least one node and the nodes always serve as both comparisons and samples. In OTM the node is always the sample stimulus connected to several comparisons. In MTO the comparison stimuli serve as nodes connected several samples. Various studies have reported differing results regarding the most effective training structure, but LS appears to be the least effective (Arntzen, 2012).

Equivalence-Based Instructions (EBI) are grounded in the principles of stimulus equivalence and previously referred to as Stimulus Equivalence Technology (Critchfield & Fienup, 2010). 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). Díaz et al. (2023) conducted a review on the use of EBI to learn portion sizes and nutritional content in food.

One topic that was emphasized in the review was the need for more research focusing on retention and long-term follow-up. Only two studies included in the review (Hausman et al., 2014; Trucil et al., 2015) examined the maintenance of results—Hausman et al. (2014) after one week and Trucil et al. (2015) after one and two weeks.

Hausman et al. (2014) examined EBI's effectiveness in estimating portion sizes, assessing whether the training generalized to new, untrained food items and if the results were maintained after one week. The study involved nine students with an average BMI of 25.8, divided into baseline, post-training, maintenance, and generalization phases. Participants were trained on connections between dry food items, measuring cups, and portion size measurement aids (PSMA). Post-training, eight out of nine participants showed improvement, five maintained accurate estimations after one week, and four generalized their skills to new food items. The authors suggested checking maintenance over longer intervals.

Trucil et al. (2015) replicated Hausman et al. (2014)'s study and extended it by examining the effects of EBI on different portion sizes (1/4 cup, 1/2 cup, and 1 cup), along with generalization to novel food items. Three participants aged 22–23 were involved, and training included 48 trials per session with a mastery criterion of 90% correct responses. After training, the participants were tested for symmetry relations and generalization. The results showed improvement, with two participants accurately estimating exercise food items, and all participants showing more accurate responses to generalization food items compared to baseline.

Nastally et al. (2010) used an LS training structure to explore nutritional content and healthier food choices using a computerized stimulus equivalence procedure. Six students aged 20–28 participated, with stimuli including laminated images of fast food and logos. A pre- and post-preference test was conducted, where the food items were ranked based on how much the participant wanted to eat them. Then, a calorie discrimination pre- and post-test was carried out, where the participant was asked to place pictures of the food items into the category they believed they belonged to. A conditional discrimination training was performed. Here, two relations were trained, then trained in a mix, before they were tested for equivalence. The stimuli used during training were food items that participants had previously sorted incorrectly. The results showed that participants, on average, improved their sorting into categories by 45%. Four out of six participants self-reported an increase in healthier food choices, and it was found that BMI had no impact on correct responding. The authors suggest that future studies could focus on other nutritional content besides calories, to see if this produces a similar effect.

Arntzen and Eilertsen (2020) wanted to expand the Nastally et al. (2010) study 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, often found in white rice, white pasta, and sugary drinks, lack fiber, vitamins, and minerals, while complex carbohydrates—present in whole grains, vegetables, and fruits—offer additional health benefits (Campos et al., 2022). In the Arntzen and Eilertsen (2020) study, twenty-two participants aged 19–54 were divided into three groups, and the experiment involved sorting food items into categories based on carbohydrate content.

The experiment was divided into different phases. The first phase consisted of pre-sorting food items, where incorrectly sorted items were used as stimuli in the following phases. Group 1 sorted into the categories 'less than 20,' '20-40,' and 'more than 40.' Groups 2 and 3 also sorted into the category 'don't know.' The second phase consisted of an OTM training structure. Groups 1 and 2 trained on 3 stimulus classes with 3 members in each class. Group 3 trained on 3 stimulus classes with 5 members in each class. 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 mastery criterion for Groups 1 and 2 was 17/18 correct trials, while for Group 3 it was a minimum of 34/36 correct baseline and symmetry trials, and 102/180 equivalence trials. The fourth phase consisted of post-sorting. The results showed that six out of eight participants in Group 1 demonstrated equivalence. All participants in Group 2 demonstrated equivalence, and five out of six participants

in Group 3 demonstrated equivalence. In the post-sorting phase, all participants in Groups 1 and 2 correctly sorted the stimuli into the appropriate categories. In Group 3, all participants sorted correctly, except for the participant who did not demonstrate equivalence. Arntzen and Eilertsen suggest that further studies should include follow-up tests to examine the effect over time.

The present study systematically replicates and extends Arntzen and Eilertsen (2020) in which participants were trained and tested for the formation of small and large equivalence classes. Stimuli presented in the pre-test were expanded to 60 stimuli from materials were expanded to include more food items. Maintenance checks were conducted one- and two-weeks post-training.

2. Method

2.1. Participants

Twenty-seven participants were recruited from a student pool and through personal contacts. The group consisted of nine men and eighteen women, aged 20 to 54 years, with an average age of 31. Participants were randomly assigned to one of three conditions: ten in Condition 1, nine in Condition 2, and eight in Condition 3. 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.

2.2. Apparatus and Stimuli

The experimental sessions took place in a university laboratory within small cubicles designed for testing. Each cubicle, approximately 1.5 square meters, 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 stimuli and recorded responses.

The stimuli used in the experiment are listed in Table 1. In the MTS program, each word and number were 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 stimuli. The remaining stimuli were arranged as B and C stimuli for Conditions 1 and 2, and as B, C, D, and E stimuli for Condition 3, depending on the tailored selection (see Pre-sorting below). Participants used a Microsoft mouse to click on the stimuli.

Additionally, 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." The category cards were always positioned in a fixed order, with "Less than 20" at the top of the pile, followed by "20–40," "More than 40," and, for Conditions 2 and 3, "Don't Know."

2.3. Design

We used a group design with Conditions 1 and 2 which involved training with six food items, while Condition 3 included twelve. Participants in Condition 1 sorted items into three categories: "More than 40," "20–40," and "Less than 20." In Conditions 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, 2 (Follow-up 1) and 4 (Follow-up 2) weeks after training and testing.

The dependent variables measured were the number of incorrectly sorted stimuli in the pre-sorting phase, the number of trials required to reach the mastery criterion during conditional discrimination training, the percentage of correct responses during tests for equivalence relations, and the number of correctly sorted stimuli during the post-sorting phase.

2.4. Procedure

The procedure included eight phases (see Figure 1): Pre-sorting and tailoring of stimuli, conditional discrimination training, tests for emergent relations, post-sorting, two follow-up tests for

4

emergent relations, and two follow-up sorting tests. If participants did not meet the test criterion during the follow-up tests, they received additional training and testing.

2.4.1. Phase 1: Pre-sorting and Tailoring of Stimuli

Participants were provided with a randomized stack of laminated cards representing food items, while the category cards remained fixed at the top. In Conditions 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 card positions were documented through photographs taken by the second author. Incorrectly sorted cards with the names of the food items were selected for further training and testing—six items for Conditions 1 and 2, and twelve for Condition 3. For Conditions 1 and 2, two cards from each category and for Condition four cards from each category. If all items within a category were correctly sorted, two items were randomly selected for additional training. This was the case for only P18969, in which two sorted correctly from the category less than 20.

2.4.2. 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. 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, text such as 'awesome,' 'correct,' 'super,' etc., appeared on the screen, while for incorrect choices, the text 'incorrect' 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 wrong. 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 Condition 1, AC and BC trials were presented separately, then mixed. For Conditions 2 and 3, AE, BE, CE, and DE trials were trained in separate blocks, progressively mixed. The mastery criterion was set at 95% correct responses before proceeding to the next block. If participants did not meet the mastery criterion, the block was repeated. The probability of programmed consequences 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.

2.4.3. 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, Conditions 1 and 2 had 90 total trials (30 each for baseline, symmetry, and equivalence relations), while Condition 3 had 300 total trials (60 for baseline, symmetry, and equivalence). The test criterion was 90% correct responses.

2.4.4. 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.

2.4.5. 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 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. 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.

2.4.6. 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.

2.4.7. 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.

2.4.8. 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.

2.5. 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.

2.6. Interobserver Agreement (IOA)

The second author and a trained observer scored 30% of all sorting sessions. Interobserver agreement was calculated using the formula: (Agreements/Agreements + Disagreements) x 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.

3. Results

3.1. Sorting and Tailoring Stimuli

Figure 2 shows the distribution of the top three most frequently incorrect sorted food items: white pepper, Toro waffles, mustard, BigOne Classic, cooked couscous, cooked pasta, Crown ice cream, strawberries, pancakes, Pizza Grandiosa, and raw quinoa. White pepper was incorrectly sorted by all participants, Toro waffles were incorrect sorted by 25 participants, while the remaining food items were incorrect sorted 24 times. See Figure 2 for a detailed overview of how these food items were typically sorted.

Across all conditions, the majority of food items were incorrect sorted during the pre-sorting phase. However, the food item that was most often sorted correctly was milk chocolate, which was sorted correctly by 13 participants. The food items selected for training and testing were those incorrectly sorted by all participants except P18969, who sorted all the food items in the "Less than 20" category correctly. For this participant, two correctly sorted stimuli were randomly selected. See Figure 3 for an overview of the food items used in training and testing.

3.2. Training Trials

In Condition 1, participants required an average of 258 trials (range: 180–375) to meet the mastery criterion. In Condition 2, participants averaged 228 trials (range: 180–495), while in Condition 3, participants used an average of 713 trials (range: 660–795) to meet the mastery criterion (see Table 3 for individual data). From Condition 1, P18981 required additional training, using 270 trials. From Condition 2, P18957 and P18980 also received additional training, using 165 and 150 trials, respectively. In Condition 3, P18962 needed extra training, requiring 630 trials to reach the mastery criterion. P18955 from Condition 3 should have received extra training but did not due to a programming error.

3.3. 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.

3.3.1. Immediate Test

Immediately after training, nine out of 10 participants in Condition 1, all 10 participants in Condition 2, and five out of eight participants in Condition 3 formed equivalence classes (see Table 3 for individual data), and Figure 4 for the percentage of participants forming equivalence classes.

3.3.2. Follow-up Tests

In the Follow-up 1 test, all nine participants in Condition 1 maintained their equivalence classes. In Condition 2, seven participants maintained their equivalence classes, while two participants, after receiving additional training, responded in accordance with stimulus equivalence in the subsequent test. In Condition 3, five participants maintained their equivalence classes, and one participant who had not previously formed equivalence classes successfully did so.

In the Follow-up 2 test, all 10 participants in Condition 1 maintained their equivalence classes. In Condition 2, all nine participants also maintained their equivalence classes. In Condition 3, six participants maintained their classes, and one participant who had not formed classes in the previous test successfully did so.

3.4. Sorting Tests

3.4.1. Immediate Test

During the post-training sorting, eight out of ten participants in Condition 1, eight out of nine in Condition 2, and four out of eight in Condition 3 sorted the cards correctly into experimenter-defined classes.

3.4.2. Follow-up Tests

In the first test for maintenance of sorting, nine out of 10 participants in Condition 1, seven out of nine in Condition 2, and four out of eight in Condition 3 sorted all items correctly. For the Follow-up 2 sorting, nine out of 10 participants in Condition 1, all participants in Condition 2, and five out of eight in Condition 3 sorted correctly. See Table 3 for an overview.

Participants P18960, P18962, and P18975, all in Condition 3, were the only ones who sorted incorrectly in all sorting tests. P18952 sorted incorrectly during post-training and Follo sorting, while P18957 and P18965 made sorting errors in only one test.

3.5. Correspondence Between Test Formats

Among the participants who sorted incorrectly, only P18962 failed to respond according to equivalence during any of the tests, except for the test conducted immediately after extra training.

This participant correctly sorted seven out of 12 stimuli across all sorting tests, corresponding to 58% correct sorts.

4. Discussion

The purpose of this study was to establish knowledge about carbohydrates using EBI procedures and to investigate the maintenance of emergent relations over three weeks, using both MTS and sorting tests. In the MTS tests, 74% of participants responded according to equivalence during the post-test, 74% correctly during maintenance, and 96% correctly during follow-up. Sorting results showed that 74% of participants sorted correctly during post-sorting, 74% during maintenance sorting, and 85% during follow-up. Only four out of 27 participants received extra training and testing, and only one participant did not meet the equivalence criterion during follow-up.

4.1. Training Trials

Participants in Conditions 1 and 2 used a similar number of trials to achieve mastery, while those in Condition 3 required more than twice the number of trials compared to the other conditions. This discrepancy is likely due to the fact that participants in Conditions 1 and 2 trained on three 3-member classes, while those in Condition 3 trained on three 5-member classes. When accounting for this difference, all conditions used roughly the same number of trials to establish relations. Most participants used a similar number of trials compared to the average for their condition, except for P18980 in Condition 2, who used 495 trials—more than twice the condition average of 228. There were no recorded disturbances during the conditional discrimination training, though external factors like the participant's fatigue from a full day's work may have played a role.

Four participants (P18981, P18980, P18957, and P18962) received additional training. During this extra training, participants used fewer trials (270, 150, 165, and 630, respectively) to reach the mastery criterion compared to their initial training (375, 495, 210, and 735 trials). All participants, except P18962, showed equivalence class formation during follow-up. P18955 should have received extra training, but due to a calculation error, it was not conducted. Nevertheless, this participant still reached the mastery criterion during follow-up.

The number of trials used in this study (averages of 258, 228, and 713 trials) aligns with previous studies by Arntzen et al. (2021), and Arntzen and Mensah (2020). Arntzen et al. (2021) used an LS training structure and reported averages of 176–206 trials for three 3-member classes and 600–800 trials for three 5-member classes.

4.2. MTS Tests and Emergence of Equivalence Classes

The results of the emergent relations tests are consistent with those from Arntzen and Eilertsen (2020), Trucil et al. (2015), and Vladescu et al. (2021). In Arntzen and Eilertsen (2020), 87.5% of participants in Condition 1, 100% in Condition 2, and 66.7% in Condition 3 showed emergence of equivalence relations. Trucil et al. (2015) and Vladescu et al. (2021) found that all participants formed equivalence classes. In the current study, 93.3% of participants in Condition 1, 92.6% in Condition 2, and 75% in Condition 3 met the test criterion for equivalence class formation across all MTS tests. The primary distinction between this study and previous studies is the timing of emergent relation tests, which were conducted at three different time points (post-test, maintenance, and follow-up). In contrast, other studies primarily focused on post-tests, with only Trucil et al. (2015) incorporating maintenance but limited to portion size estimates.

Participant P18962 did not meet the test criterion on any of the MTS tests, except after additional training. During debriefing, P18962 explained he/she had assumed the pre-sorting method was the correct solution, even after conditional discrimination training. After extra training, the participant changed the strategy but considered the new solution illogical, reverting to the original strategy in follow-up.

For P18953, baseline relations were not tested during the post-test but were tested during maintenance and follow-up, where the participant responded with 100% accuracy across all phases,

including sorting tests. It is likely this participant would have met the test criterion for baseline relations during the post-test.

4.3. Sorting Tests

The sorting results align with findings from Nastally et al. (2010) and Arntzen and Eilertsen (2020). In all three studies, most participants showed improvement from pre-sorting to post-sorting. Arntzen and Eilertsen (2020) reported 100% correct sorting for Conditions 1 and 2, and 83.3% correct sorting for Condition 3. In the current study, 86.7% of participants in Condition 1, 88.9% in Condition 2, and 54% in Condition 3 sorted correctly in all combined sorting tests. Although the percentage of correct sorting was slightly lower in this study, it may be due to the additional tests for post-sorting, maintenance, and follow-up sorting. One clear reason for the difference in outcomes in the sorting tests is that, in the present study, all 60 stimuli were presented during the tests, whereas in Arntzen and Eilertsen, the post-sorting test only included stimuli that had been part of the MTS training and testing. The use of the whole stimulus set as in the present experiment strengthen the results of post-sorting test.

The use of pre-sorting in studies involving meaningful stimuli ensures that stimuli are not already part of a participant's learned stimulus classes, which could impact emergent relation testing (Arntzen & Mensah, 2020). Sorting tests serve as an alternative method for testing the emergence of equivalence classes, with results showing a strong correspondence between sorting and MTS tests in many cases (e.g., Arntzen et al., 2021; Bevolden & Arntzen, 2018).

4.4. Maintenance and Follow-Up

Two studies, Hausman et al. (2014) and Trucil et al. (2015), tested for maintenance after one week 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, with participant responses remaining similar or improving over three weeks. Extra training was provided for five participants who did not meet the mastery criterion during follow-up.

4.5. 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, and the conditional discrimination training and testing for emergent relations involved only the stimuli that the individual participant had not correctly sorted in the pre-test. This tailoring ensured that the stimuli were not already part of existing classes, which can pose a challenge in EBI experiments or when using familiar stimuli in training prerequisites for tests of emergent relations. The type of stimuli used in EBI research differs significantly from those used in basic research. In stimulus equivalence studies, abstract shapes are typically used, and the relations between stimuli are arbitrarily defined (Sidman & Tailby, 1982). To ensure that the stimulus set used in conditional discrimination training is not already partitioned into different classes, it is essential to tailoring the stimulus set used for each participant. The tailoring procedure used in the present experiment is in accordance with other experiments exploring teaching nutrition skills (Arntzen & Eilertsen, 2020; Nastally et al., 2010; Oenema et al., 2001).

4.6. 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 17 days. Further studies should impose stricter time intervals to reduce variability. Additionally, the use of correctly sorted stimuli for training, as in P18969's case, may have influenced the outcome and should be further investigated. The first 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 such issues.

One limitation of the study was the frequent use of custom categories by participants, particularly in Condition 1, where they did not have the option to sort into a 'Don't Know' category. This was less common in Conditions 2 and 3. The differences in sorting instructions between this study and those used by Arntzen and Eilertsen (2020) may have influenced participants' sorting strategies, and this should be examined in future research.

Future studies could also explore distinctions between simple and complex carbohydrates, as complex carbohydrates provide greater health benefits. Investigating participants with a heightened focus on nutritional content, such as those on specific diets or managing lifestyle diseases like diabetes, could reveal differences in knowledge acquisition and retention.

4.7. Summary

This study replicated and extended Arntzen and Eilertsen (2020), incorporating additional food items during sorting. Emergent relations were consistently observed across all conditions in the MTS post-tests and were maintained over a three-week period. Sorting tests yielded similar results for Conditions 1 and 2, with responses remaining stable. Although sorting results for Condition 3 were slightly lower than MTS tests, they remained consistent over time.

Author Contributions: Investigation, Ina Charlotte Spilling; Writing – review & editing, Erik Arntzen.

Institutional Review Board Statement:

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. 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
2. 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
3. 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
4. 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
5. Campos, V., Tappy, L., Bally, L., Sievenpiper, J. L., & Le, K. A. (2022). Importance of carbohydrate quality: What does it mean and how to measure it? J Nutr, 152(5), 1200-1206. https://doi.org/10.1093/jn/nxac039
6. 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
7. Díaz, M., Garcia, Y., & Mahoney, A. (2023). A review on equivalence-based procedures to teach food-portion estimation skills. Behavioral Interventions. https://doi.org/10.1002/bin.1934
8. 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, 19–33. https://doi.org/10.1901/jaba.2010.43-19
9. 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.
10. 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
11. 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

12. 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
13. Kazdin, A. E. (2010). Singe-Case Research Designs (2nd ed.). Oxford University Press.
14. 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, 105–125. https://doi.org/10.1002/bin.1742
15. Nastally, B., Dixon, M. R., McKeel, A., & Fleming, M. (2010). Teaching healthy food choices through stimulus equivalence. Journal of Behavioral Health and Medicine, 1, 119–126.
16. Oenema, A., Brug, J., & Lechner, L. (2001). Web-based tailored nutrition education: results of a randomized controlled trial. Health Education Research, 16, 647–660.
17. Ong, T., Normand, M. P., & Schenk, M. J. (2018). Using equivalence-based instruction to teach college students to identify logical fallacies. Behavioral Interventions, 1–14. https://doi.org/10.1002/bin.1512
18. Regan, B., Vladescu, J. C., Reeve, K. F., & DeBar, R. M. (2018). Equivalence-based instruction with nonfood items to increase portion-size estimation accuracy. The Psychological Record. https://doi.org/10.1007/s40732-018-0299-7
19. Sidman, M. (1992). Equivalence relations: Some basic considerations. In S. C. Hayes & L. J. Hayes (Eds.), Understanding verbal relations (pp. 15–27). Context Press.
20. 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
21. 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
22. 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