UNIVERSIDADE FEDERAL DE SÃO CARLOS – UFSCAR

CENTRO DE ESTUDOS DE CIÊNCIAS HUMANAS – CECH

PROGRAMA DE PÓS-GRADUAÇÃO EM PSICOLOGIA

LABORATÓRIO DE ESTUDO DO COMPORTAMENTO HUMANO – LECH

DELAYED MTS(2s) AND TRAINING-IRAP PROCEDURES: EXPLORING EFFICACY, EFFICIENCY AND THEIR TRAINING FORMAT IN THE STUDY OF DERIVED RELATIONS

Marcello Henrique Silvestre

São Carlos - SP

Março/2024

UNIVERSIDADE FEDERAL DE SÃO CARLOS – UFSCAR

CENTRO DE ESTUDOS DE CIÊNCIAS HUMANAS – CECH

PROGRAMA DE PÓS-GRADUAÇÃO EM PSICOLOGIA

LABORATÓRIO DE ESTUDO DO COMPORTAMENTO HUMANO – LECH

DELAYED MTS(2s) AND TRAINING-IRAP PROCEDURES: EXPLORING EFFICACY, EFFICIENCY AND THEIR TRAINING FORMAT IN THE STUDY OF DERIVED RELATIONS[1]

Marcello Henrique Silvestre

Dissertação de Mestrado submetida ao Programa de Pós-Graduação em Psicologia da Universidade Federal de São Carlos, sob orientação do Prof. Dr. Júlio César de Rose.

[1] O presente trabalho foi realizado com o apoio da bolsa de mestrado da Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP (Processo #2022/02297-0). Os resultados experimentais incluídos nesta dissertação fazem parte do programa de pesquisas do Instituto Nacional de Ciência e Tecnologia sobre Comportamento, Cognição e Ensino (INCT-ECCE).

Agradecimentos

Todo o caminho trilhado até então não foi feito sozinho...

Agradeço aos meus pais João e Núbia por sempre acreditarem no meu potencial (pessoal e profissional), e melhor do que acreditar: oferecer as oportunidades para o meu desenvolvimento. Agradeço também ao meu irmão Marcos, por estar presente em minhas conquistas.

Agradeço a minha namorada Barbara, pelo apoio, torcida, paciência, compreensão, parceria e leveza ao longo de todo esse tempo. A passagem do tempo só tem sentido com você, ao seu lado. Agradeço também pela ampliação do meu olhar enquanto pesquisador para a aplicação. Temos muitos trabalhos futuros a fazer!

Agradeço a Denise, Colin e Djenane, pela amizade e oportunidades. A vida acadêmica não é fácil, mas vejo através de vocês a alegria por ela.

Agradeço imensamente ao meu orientador Júlio, pela confiança e liberdade aqui me dada. Minha experiência e humildade enquanto pesquisador é reflexo seu.

Agradeço aos meus amigos que fiz em São Carlos pelos ótimos momentos. Mudar de cidade não é fácil, mas vocês tornaram essa experiência melhor. Agradeço especialmente ao Léo, conviver com você foi simples e divertido.

Agradeço à turma da aula de RFT optativa. A experiência de “dar aula” para vocês está presente comigo até hoje, foi transformadora. Espero vocês enquanto futuros amigos de profissão.

The only future that is known is the past that has been experienced
(Hayes et al., 2001)

Silvestre, M. H. (2024). *Delayed MTS(2s) and Training-IRAP Procedures: Exploring Efficacy, Efficiency and Their Training Format in the Study of Derived Relations* [Dissertação de Mestrado, Universidade Federal de São Carlos].

Abstract

A common method for the study of derived relations is the Matching-to-Sample (MTS) preparation. However, certain aspects of the procedure arguably hinder the emergence of new relations, such as different sources of stimulus control. Consequently, the investigation of other procedures seems justified. The Training-IRAP may be a plausible alternative. The primary objective was to compare the effectiveness of Delayed MTS(2s) and Training-IRAP procedures in terms of participant derived relations yield. The secondary objective was to compare the average number of training blocks needed per procedure. Due to additional components in the Standard Training-IRAP not found in MTS, some edits were made to the former procedure yielding two new versions: the Modified Training-IRAP and Delayed Modified Training-IRAP(2s). Sixty-eight students participated, 17 per group. Nine participants met yield criteria in the Delayed MTS(2s) group, three in the Standard Training-IRAP group, 11 in the Modified Training-IRAP group, and 12 in the Delayed Modified Training-IRAP(2s) group. The average number of blocks to complete training phases was lowest for the Delayed MTS(2s) and Delayed Modified Training-IRAP(2s) groups. The formats of the procedures and implications of the findings toward an overarching goal of greater precision, scope and depth in conceptual, experimental and applied settings are discussed.

Keywords: derived relations, yield, Training-IRAP, Delayed MTS(2s), training format

List of Figures

List of Tables

Sumário

Introduction

Language and cognition as symbolic behavior have been studied by behavior analysts largely through research on emergent or derived relations (Barnes-Holmes et al., 2018; Hayes et al., 2001; Perez et al., 2023; Sidman, 1994). The most common procedure employed in this regard (i.e., establishing relations between stimuli and testing for new relations that were never directly taught) is Matching-to-Sample (MTS; Fienup & Brodsky, 2020; Pilgrim, 2020; Sidman, 1971). In a typical MTS preparation, a sample stimulus (e.g., A1) is presented, and a response to this stimulus (e.g., clicking on that stimulus with a computer mouse) leads to the appearance of two or more comparison stimuli, one of which is experimentally correct (S+; e.g., B1) and the others incorrect (S-; e.g., B2). In the presence of A1, choosing B1 is reinforced, whereas choosing B2 results in no programmed reinforcement. The response to B2 will only be reinforced if presented with an alternative stimulus, A2. Given the training of AB and BC relations (two trained relations), the learner typically demonstrates, without direct training, the derived relations BA, CB, AC, and CA (four untrained, derived relations). Although various facilitating parameters for the emergence of derived relations have been studied within the MTS procedure (Arntzen, 2012), some aspects of MTS have been shown to hinder the emergence of new relations and transfers of function between stimuli (i.e., a phenomenon whereby a stimulus acquires the functions of another stimulus with which it has been related; e.g., Dougher et al., 1994; Martins et al., 2023; Perez, de Almeida, et al., 2020). These aspects include, for example, different sources of stimulus control during baseline training (i.e., selection and rejection control; e.g., Carrigan & Sidman, 1992; Johnson & Sidman, 1993; Perez et al., 2015, 2017, 2019; Perez, Huziwara, et al., 2020).

In the example above, choosing B1 can be controlled by both the relation between the sample stimulus (i.e., A1) and the comparison stimulus defined as S+ (i.e., B1), and the relation between the sample stimulus and the comparison stimulus defined as S- (i.e., B2). In the latter case, B2 is argued to control the selection of the other available comparison stimulus (i.e., B1). In other words, the individual would reject B2 and select B1. As such, the choice response may be controlled by both the sample-S+ relation (selection control), where the individual responds to S+, and/or the sample-S- relation (rejection control), where the individual responds to S- (de Rose, 1996). However, the controlling relation may not be evident in any given case without additional tests (Perez & Tomanari, 2013).

Despite the initial hypothesis that rejection control would hinder tests of derived relations (Carrigan & Sidman, 1992), some studies have demonstrated better performance on derived relations tests when conjoint S+ and S- control is established for all baseline relations (e.g., Arantes & de Rose, 2015; de Rose et al., 2013; Plazas, 2019). In addition, other studies have indicated the importance of S- control for producing derived relations (Harrison & Green, 1990; Plazas & Peña, 2016; Plazas & Villamil, 2016a, b). However, to induce control by both S+ and S- in an MTS task, additional procedures are necessary and often their effectiveness depends on combinations of procedures (see Perez & Tomanari, 2014, for a methodological review of these procedures; see also Perez et al., 2019, for an empirical analysis of procedures used to establish S- control). In light of this, investigation of other effective procedures for the study of derived relations seems justified.

The Training-IRAP, a procedure originating from Relational Frame Theory (RFT; Barnes-Holmes & Harte, 2022a; Hayes et al., 2001), has been employed recently to establish relations between stimuli and testing for derived relations (Harte et al., 2018, 2020, 2021; Harte, Barnes-Holmes, Moreira, et al., 2021; Leech & Barnes-Holmes, 2020; Leech et al., 2018). This methodology is a version of the Implicit Relational Assessment Procedure (IRAP; Barnes-Holmes et al., 2010), a procedure originally developed to assess the probability of particular relational responses. However, while the IRAP requires that participants alternate between responses consistent and contrary to their individual learning histories (i.e., toward the aim of assessing response probability), the training version (i.e., Training-IRAP) does not require such contrary responding (i.e., it is simply used to train particular relations; see Barnes-Holmes & Harte, 2022b, for more details on the difference between the Training-IRAP and IRAP format). The Training-IRAP procedure involves presenting a pair of stimuli (e.g., A1 and B1, or A1 and B2) and two response options (e.g., "Same" and "Different"); the task requires the participant to respond both quickly and accurately to the stimulus pair with one of the two response options. In other words, the Training-IRAP specifies the relations between the stimuli (i.e., A1/B1-Same, A1/B2-Different) in a way that MTS does not.

The primary objective of the present study was to compare the effectiveness of MTS and Training-IRAP procedures with regards to participants' performance on subsequent tests for derived relations. The measure used was *yield* (Fields et al., 2020), defined as the percentage of participants who reached the required criterion for class formation (i.e., a set of stimuli that have come to function interchangeably). The secondary objective was to compare the average number of training blocks needed per procedure. For current purposes, we used the Delayed Matching-to-Sample 2s [DMTS(2s)] format rather than other MTS alternatives, which has proven to be an effective MTS preparation (e.g., Bortoloti & de Rose, 2009, 2012). However, due to the fact that the Standard Training-IRAP presents additional components not usually found in MTS, a new version, excluding these components, was created and also employed: the Modified Training-IRAP. In this way, a more similar methodological comparison could arguably be made between the procedures (e.g., removing the Training-IRAP requirement to respond under time pressure; see Method section below for a full explication of these changes between procedures). However, given that additional DMTS(2s) components are known to positively affect yield, but was not an element involved in the Standard Training- or Modified Training-IRAPs, two final procedural manipulations were made. These components are the observational response to the stimuli and the 2-second delay between the removal of the sample stimulus and the presentation of the comparison stimuli. Consequently, in order to investigate the impact of these variables and whether their introduction would be effective in the Training-IRAP (as it is with MTS procedure), this final alteration was made to the Modified Training-IRAP, making it a Delayed Modified Training-IRAP(2s). Overall, therefore, the study consisted of four procedures that were compared to each other in terms of yield and the average number of blocks needed to establish baseline relations: (i) DMTS(2s), (ii) Standard Training-IRAP, (iii) Modified Training-IRAP, and (iv) Delayed Modified Training-IRAP(2s).

Method

Participants

Sixty-eight students, from [REMOVED FOR BLINDING PURPOSES], participated (*M*= 36, *F*= 32), with an age range of 18-30 (*M*= 21, *SD*= 2.6). They were native speakers of Brazilian Portuguese, without any mental health diagnosis, without knowledge of Stimulus Equivalence and RFT, and without prior experience with the procedures of this study. The participants were recruited through convenience sampling. Before the experiment, they read and signed the Informed Consent Form (approved by the Brazilian platform for ethical committees, Plataforma Brasil, CAAE: [REMOVED FOR BLINDING PURPOSES]). At the end of the experiment, participants were fully debriefed. None received any compensation for participation.

Location and Equipment

The study was conducted in a room at the [REMOVED FOR BLINDING PURPOSES] using a computer (LG, Intel® Core™ i5-5200U CPU @ 2.20GHz, 4G RAM, 1TB HD) equipped with the PsychoPy software (Peirce et al., 2019). Procedures were programmed with this software, and it automatically recorded and stored data in an Excel spreadsheet.

Stimuli

Ten abstract shape stimuli were used in the study (see Figure 1).

INSERT FIGURE 1 HERE

Experimental Design

We employed a between-groups comparison design. Participants were equally and randomly divided into four experimental groups, each with *N*= 17: DMTS(2s), Standard Training-IRAP, Modified Training-IRAP and Delayed Modified Training-IRAP(2s).

Dependent Variables

The main comparison between groups was the number of participants who reached the pre-experimentally established criteria (see below) for class formation (yield). The groups were also compared based on the average number of blocks needed to complete the baseline relation training.

Procedures

The experimental session took place in a single day. The experiment involved relational training to form two five-member stimulus classes: A1B1C1D1E1 and A2B2C2D2E2. A Linear Series (LS) training structure was chosen to avoid a ceiling effect (Arntzen, 2012). After baseline training, cumulative baseline blocks were presented, first with differential consequences, and then without. Following the attainment of all criteria, probes for derived relations were conducted, from the largest to the smallest nodal distance (i.e., EA, DA, EB, CA, and DB, respectively) divided into blocks and interspersed within cumulative baseline blocks without differential consequences. Criteria for class formation was a score of more than 90% correct responses on each probed relation and a minimum of 87.5% correct responses in the cumulative baseline blocks interspersed between the probes. In three of the four training and testing procedures [DMTS(2s), Modified Training-IRAP, Delayed Modified Training-IRAP(2s)], participants selected stimuli by clicking on them with the mouse. The training and testing procedures involved in the Standard Training-IRAP required pressing specific keys on the computer keyboard to respond to stimuli.

DMTS(2s)

Each trial started with the presentation of a sample stimulus (e.g., A1) at the top of the screen. An observing selection response to the sample produced its withdrawal and, after 2 seconds, two comparison stimuli (e.g., B1 and B2) were displayed at the bottom of the screen, side by side (see Figure 2). Selecting one of them resulted in their withdrawal, and then one of the differential consequences was presented for 1 second: a correct response (e.g., selecting B1 given sample A1) displayed the word "Correct" in green; an incorrect response (selecting B2 given sample A1) displayed the word "Wrong" in red. The screen was darkened in the inter-trial interval (ITI) of 40ms, which separated the end of the consequence from the beginning of the next trial. To illustrate, in an AB training block, this training format comprised two types of trials: if sample A1, then choose B1; and if sample A2, then choose B2.

INSERT FIGURE 2 HERE

Standard Training-IRAP

Each trial began with the presentation of a sample stimulus at the top of the screen (e.g., A1), a comparison stimulus in the center of the screen (e.g., B1), and two response options at the bottom of the screen (i.e., "Same" and "Different"), one in each corner of the screen, with their positions randomized for each trial[2] (see Figure 3). Participants pressed the 'd' key on the keyboard to select the response option on the left, and the ‘k’ key to select the option on the right. If the chosen option was correct (e.g., if the response option “Same” was

[2] In the RFT literature, the stimulus presented at the top of the screen within an IRAP is referred to as the label stimulus, and the stimulus in the center of the screen is referred to as the target stimulus. However, in order to facilitate the comparison of trial structure between MTS and Training-IRAP procedures, we substituted the terms “label” and “target” for those more commonly employed in the MTS literature: sample stimulus and comparison stimulus, respectively.

selected in upon presentation of sample A1 and comparison B1), the word "Correct" was displayed in green for 1 second, positioned on the screen between the response options. If the chosen response was incorrect (e.g., if in response to A1 and B1, "Different" was selected), a red "X" appeared between the response options and remained until the correct response option (i.e., "Same") was selected. After delivering the feedback "Correct" or following the correction of the incorrect response, there was an ITI of 40ms before the start of a new trial. At the end of each block of trials, the participant was presented with a screen displaying their percentage of correct responses and response latency for that block. Additionally, the scores (accuracy and latency) required to advance to the next block were shown. An average latency ≤ 3000 ms was required. If necessary, a performance-specific tip was provided (e.g., "Respond with greater accuracy," "Respond more quickly"). Unlike MTS, Training-IRAP format training blocks comprised four types of trials based on the sample-comparison stimulus combinations. For example, in the AB training block, the training trial-format was (Sample/Comparison-Correct response): A1/B1-Same, A1/B2-Different, A2/B1-Different, A2/B2-Same.

INSERT FIGURE 3 HERE

Modified Training-IRAP

The trial structure of this procedure remained the same as the Standard Training-IRAP described above. However, some components were changed or removed with the aim of making it more similar to MTS: 1. The selection of one of the response options was now made using the mouse (i.e., clicking with the mouse on one of the response options) instead of pressing the 'd' and 'k' keys on the keyboard; 2. The correction requirement following an incorrect response was removed. As such, after emitting an incorrect response, the feedback "Wrong" was displayed in red for 1 second and the trial ended; and 3. The latency criterion for making a response was removed, as was 4. the post-block screen displaying that block's accuracy and latency.

Delayed Modified Training-IRAP(2s)

This procedure was identical to the Modified Training-IRAP, except that the 2-second delay component present in DMTS(2s) was also added. Thus, a trial began with the presentation of the sample (e.g., A1). After an observing response to the sample was emitted, this stimulus was removed and, after 2 seconds, one of the comparison stimuli (e.g., B1 or B2) appeared. An observing response to the comparison stimulus was also required, followed immediately by the presentation of the two response options: "Same" and "Different" (see Figure 4).

INSERT FIGURE 4 HERE

Instructions

Before the start of the experiment, instructions were given in writing (in Portuguese), and if there were no questions, the experiment began. The following instructions were provided for all groups: *"Please, store your cellphone (put it on silent) and avoid using it during the experimental session."* and *"There will come a point in the experiment where the computer will no longer indicate if choices are correct or incorrect, but there will always be a correct response, and the computer will continue to record whether the choice was correct or not. If you have any questions, ask the experimenter"*. Specific instructions for the different groups were as follows:

DMTS(2s). *"Initially, an image will be presented at the top of the screen. Observe the image and click on it with the left mouse button. This image will disappear, and two other images will appear shortly afterward at the bottom of the screen. Your task is to discover the combinations/relations of these images with the one that appeared at the top of the screen. To select one of the images, move the mouse cursor to it and click the left mouse button. If your* *choice is correct, the word 'Correct' (in green) will appear on the screen, and if your choice is incorrect, the word 'Wrong' (in red) will appear in the center of the screen."*

Standard Training-IRAP. *"You will need to select one of the words 'Same' and 'Different' that will be presented at the bottom of the computer screen. An image will be shown at the top of the screen, followed by another one presented at the center. Your task is to discover the relationship between the images by selecting one of the options, 'Same' or 'Different,' whose positions will be randomized in each trial. To choose the response option on the left side of the screen, press the 'd' key, and to select the option on the right, press the 'k' key on your computer. If your choice is correct, the word 'Correct' (in green) will appear on the screen, and you will move on to the next trial. If your choice is incorrect, an 'X' (in red) will appear on the screen. To proceed to the next trial, select the correct response option. At the end of each block of trials, a screen will show your percentage of correct responses and speed in the block, along with the scores you need to achieve to advance in the task. If necessary, a hint (in red) will also appear. Respond with the utmost accuracy. When you learn to be precise, you will naturally respond faster as well."*

Modified Training-IRAP. As noted above, in the modified versions of the Training-IRAP, the typical score screen (accuracy and latency) was not presented at the end of each block of trials in an attempt to reduce additional extraneous variables not present in MTS studies. In addition, although instructions for the Modified Training-IRAP group were similar to the previous group, the following two modifications were made regarding (i) making a response and (ii) feedback, respectively: *"To select one of the response options ('Same' or 'Different'), move the mouse cursor over the desired option and click the left mouse button.";* and *"If your choice is correct, the word 'Correct' (in green) will appear on the screen, and if your choice is incorrect, the word 'Wrong' (in red) will appear in the center of the screen."*

Delayed Modified Training-IRAP(2s). The instructions involved for the Delayed Modified Training-IRAP(2s) were similar to the previous group with one exception. Specifically, to reflect the additional observational response and 2-second delay components involved, the initial part of the instructions was modified: *"Initially, an image will be presented at the top of the screen. Observe the image and click the left mouse button on it. This image will disappear, and another image will appear shortly afterward, in the center of the screen. Observe the image and click the left mouse button on it. Right after, two words will appear at the bottom of the screen: Same and Different."*. The rest of the instructions were identical to those for the group with Modified Training-IRAP.

Experimental Phases

Phase 1. Baseline Training. The training blocks were divided and respectively trained in the following order: AB, BC, CD, and DE; with each block consisting of 24 trials. The DMTS (2s) group had 12 trials with each sample in these blocks (e.g., 12 with sample A1 and 12 with sample A2 in the AB training block). The Standard Training-IRAP, Modified Training-IRAP, and Delayed Modified Training-IRAP(2s) groups had 6 trials for each trial-type (e.g., 6 for A1B1-Same, 6 for A1B2-Different, 6 for A2B1-Different, 6 for A2B2-Same, in the AB training block). The differential consequences for correct and incorrect responses were the words "Correct" and "Wrong" presented for 1 second. Participants were required to make 22 correct responses per block (> 90% accuracy), and in the Standard Training-IRAP, Modified Training-IRAP, and Delayed Modified Training-IRAP(2s) groups, an additional criterion of no more than one error per trial-type was employed to ensure a minimum of 83.3% correct responses for each trial-type. Only in the Standard Training-IRAP group was an additional requirement for average response latencies to be no longer than 3,000 milliseconds (≤ 3,000 ms) employed.

Phase 2. Cumulative Baseline. This block consisted of the randomized presentation of all previously trained baseline relations, with each block comprising 64 trials; both the ITI and the presentation of differential consequences remained the same as in Phase 1. In the DMTS(2s) group there were 8 trials with each sample (i.e., 8 for A1B1, 8 for A2B2, 8 for B1C1, 8 for B2C2, and so on). In the Standard Training-IRAP, Modified Training-IRAP, and Delayed Modified Training-IRAP(2s) groups there were 4 trials for each trial-type (i.e., 4 for A1B1-Same, 4 for A1B2-Different, 4 for A2B1-Different, 4 for A2B2-Same, 4 for B1C1-Same, 4 for B1C2-Different, and so on). The required performance was 58 correct responses in the block (> 90% accuracy), and in the Standard Training-IRAP, Modified Training-IRAP, and Delayed Modified Training-IRAP(2s) groups, an additional criterion of no more than one error per trial-type was set. Once again, for the Standard Training-IRAP group only, average response latencies were required to be equal to or less ≤ 3000 ms. If these criteria were not met but at least 75% of the responses were correct (i.e., 48 in total), the block could be repeated up to more two times. However, if any of the required criteria were not reached after three presentations of this block, or if the total correct responses in any presentation fell below 75%, the participant returned to the previous phase.

Phase 3. Cumulative Baseline Block Without Differential Consequences. This block was identical to the previous phase, except that differential consequences were removed. If any of the same criteria as above were not achieved in the first presentation of the block, the participant returned to Phase 2.

Phase 4. Probe for Derived Relations. The probes for derived relations were divided into EA, DA, EB, CA, and DB blocks, in that order, with each block consisting of 24 trials. These were interspersed with cumulative baseline blocks, each comprising 16 trials. That is, after each derived relation probe block, a cumulative baseline block was presented only if it was followed by the next derived relation probe block (i.e., after the DB probe block there was not a cumulative baseline block). None of these blocks had differential consequences. In the DMTS(2s) group, each test block had 12 trials with each sample (e.g., 12 with E1 and 12 with E2 in the EA test block); in the Standard Training-IRAP, Modified Training-IRAP, and Delayed Modified Training-IRAP(2s) groups there were 6 trials for each trial-type in each testing block (i.e., 6 for E1A1-Same, 6 for E1A2-Different, 6 for E2A1-Different, 6 for E2A2-Same; and so on). In the cumulative baseline blocks, interspersed between testing blocks, there were two trials with each sample in the DMTS(2s); in the Standard Training-IRAP, Modified Training-IRAP, and Delayed Modified Training-IRAP(2s) groups there was one trial of each trial-type. The required criteria for class formation were a minimum of 14 correct responses in the cumulative baseline blocks (87.5% correct) and a minimum of 22 correct responses in the derived relations probe blocks (> 90% correct).

Results

Yield

Figure 5 presents, for each group, the number of participants that attained criteria for class formation. Nine participants met criteria in the DMTS(2s) group, three in the Standard Training-IRAP group, 11 in the Modified Training-IRAP group, and 12 in the Delayed Modified Training-IRAP(2s) group. Given the small sample sizes of the groups, a Fisher's exact test was conducted to assess the differences in yield between the groups (Fields et al., 2020). This test revealed a significant difference between the groups (two-sided $p < .01$). A post-hoc Dunn test with Bonferroni-adjusted contrasts was employed (p-value * 6), along with calculating effect size using the Odds Ratio (OR). OR is a measure of association between an exposure (i.e., different procedures) and an outcome (i.e., achieved or not the criteria during probe of derived relations), compared to the odds of the outcome in a different exposure. An OR= 1.0 suggests that the exposure does not affect odds of outcome, OR> 1.0 suggests that the exposure is associated with higher odds of outcome, and OR< 1.0 suggests that the exposure is associated with lower odds of outcome (Szumilas, 2010). With the p-value adjusted for Bonferroni correction, a significant difference was observed between the Delayed Modified Training-IRAP(2s) and Standard Training-IRAP groups (*p*= .029; *OR* 10.26, 95% *CI* 1.8-81.9). Other group comparisons did not show statistically significant differences: DMTS(2s) x Standard Training-IRAP (*p*= .42; *OR* 4.98, 95% *CI* 0.9-37.2), Modified Training-IRAP x DMTS(2s) (*p*= 1; *OR* 1.6, 95% *CI* 0.3-8.1), Delayed Modified Training-IRAP(2s) x DMTS(2s) (*p*= .1; *OR* 2.1, 95% *CI* 0.4-11.2), Modified Training-IRAP x Standard Training-IRAP (*p*= .06; *OR* 7.95, 95% *CI* 1.4-61.3), and Delayed Modified Training-IRAP(2s) x Modified Training-IRAP (*p*= 1; *OR* 1.3, 95% *CI* 0.3-7.1).

INSERT FIGURE 5 HERE

Average Number of Trial Blocks to Complete Training Phases

The average number of blocks to complete the training phases (i.e., baseline, cumulative baseline, and cumulative baseline without differential consequences) is presented in Table 1 (see Appendix for complete individual data per block and procedure).

INSERT TABLE 1 HERE

Baseline

With respect to the baseline blocks, a Kruskal-Wallis test indicated a significant difference between the groups (H(3) = 32.78; *p*< .001). A post-hoc Dunn test with p-value adjusted for Bonferroni corrections indicated statistically significant differences between some groups remained: DMTS(2s) x Standard Training-IRAP (*p*< .001), DMTS(2s) x Modified Training-IRAP (*p*< .006), Standard Training-IRAP x Delayed Modified Training-IRAP(2s) (*p*< .006). Overall, therefore, the results indicated that participants required significantly less blocks to reach criterion on the DMTS(2s) than on the Standard Training-IRAP and Modified Training-IRAP, and significantly less on the Delayed Modified Training-IRAP(2s) relative to the Standard Training-IRAP.

Average Number of AB, BC, CD, and DE Baseline Blocks. The baseline block consists of the sequential presentation of the AB, BC, CD, and DE blocks. Figure 6 presents the average number of AB, BC, CD, and DE baseline blocks, along with their respective standard deviations for each procedure.

INSERT FIGURE 6 HERE

AB Block. A Kruskal-Wallis test indicated a significant difference between the groups (H(3) = 23.91; $p < .001$). With the p-value adjusted for Bonferroni correction, statistically significant differences between some groups remained (via a post-hoc Dunn test): DMTS(2s) x Standard Training-IRAP ($p < .001$), Standard Training-IRAP x Delayed Modified Training-IRAP(2s) ($p = .012$).

BC Block. A Kruskal-Wallis test indicated a significant difference between the groups (H(3) = 25.22; $p < .001$). With the p-value adjusted for Bonferroni correction statistically significant differences between some groups remained (via a post-hoc Dunn test): DMTS(2s) x Standard Training-IRAP ($p < .001$), DMTS(2s) x Modified Training-IRAP ($p = .006$), Standard Training-IRAP x Delayed Modified Training-IRAP(2s) ($p = .024$).

CD Block. A Kruskal-Wallis test did not indicate a significant difference between the groups (H(3) = 7.46; $p = .06$).

DE Block. A Kruskal-Wallis test indicated a significant difference between the groups (H(3) = 8.95; $p = .03$), but after the post-hoc Dunn test and adjusting the p-value for Bonferroni correction no statistically significant differences between groups remained ($p < .049$).

Overall, therefore, the results indicated that the differences between the average number of baseline blocks per procedures were significant only for the initial two blocks (i.e., AB and BC); no significant difference was evident for the final two blocks (i.e., CD and DE). In AB blocks, participants needed less blocks in the DMTS(2s) relative to the Standard

Training-IRAP, and less block in Delayed Modified Training-IRAP(2s) relative to the Standard Training-IRAP. In BC blocks, participants needed less blocks in the DMTS(2s) relative to the Standard and Modified Training-IRAP, and less blocks in the Delayed Modified Training-IRAP(2s) relative to the Standard Training-IRAP.

Cumulative Baseline

With respect to the cumulative baseline blocks, a Kruskal-Wallis test indicated a significant difference between the groups (H(3) = 14.95; $p < .01$). A post-hoc Dunn test with Bonferroni correction indicated differences between two groups: DMTS(2s) x Standard Training-IRAP ($p < .006$). In general, therefore, the results indicated that participants required significantly less blocks to reach criterion only on the DMTS(2s) only relative to the Standard Training-IRAP.

Cumulative Baseline Without Differential Consequences

With respect to the cumulative baseline without differential consequences blocks, a Kruskal-Wallis test indicated differences between the groups (H(3) = 12.39; $p < .01$). With the p-value adjusted for Bonferroni correction, a post-hoc Dunn test indicated a significant difference between: DMTS(2s) x Standard Training-IRAP ($p = .018$), and Standard Training-IRAP x Modified Training-IRAP ($p = .018$). Overall, therefore, the results indicated that participants required significantly less blocks to reach criterion only on the DMTS(2s) relative to the Standard Training-IRAP, and on the Modified Training-IRAP relative to the Standard Training-IRAP.

Response latency during probes for derived relations in Standard Training-IRAP and Modified Training-IRAP

Although no latency criterion was in place during probe blocks across any of the groups, the Standard Training-IRAP involved a latency criterion for all other stages of the training procedure. Given how poor the yield was for the Standard Training-IRAP group relative to the other groups, one possibility may be that participants continued to try to respond quickly during these probe trials even though fast responding was not required. In order to explore this possibility, a post-hoc analysis was conducted between participant overall median response latencies during probe trials for the Standard Training-IRAP and Modified Training-IRAP. This exploratory comparison was made between these two procedures and not the others given the close structural similarities involved (i.e., no delayed stimulus presentation). To this end, a Mann-Whitney U-test was conducted between Standard Training-IRAP (*M*= 1986ms; *Mdn*= 1808ms) and Modified Training-IRAP (*M*= 2672ms; *Mdn*= 2400ms). The result showed a statistically significant difference between the median response latencies of these groups ($U = 6912$; $p < .001$). That is, participants responded significantly faster during the derived relations probes within the Standard Training-IRAP procedure than they did when responding to the derived relations probes within the Modified Training-IRAP procedure, despite neither procedure requiring fast responding during this phase.

Discussion

The main objective of the current study was to compare yield between DMTS(2s) and methodological variations of the Training-IRAP. The results did not show a statistically significant difference in yield between DMTS(2s) and the three Training-IRAP procedures. However, statistically significant differences in yield were observed between the Standard Training-IRAP and Delayed Modified Training-IRAP(2s). Despite the lack of a statistically significant difference between the Standard Training-IRAP and DMTS(2s), a visual inspection of Figure 5 above indicates that the Standard Training-IRAP produced lower yield than all other procedures, DMTS(2s) and Training-IRAP variations alike. Thus, the odds of achieving the pre-established criteria in the probes of derived relations were 4.98 (95% *CI* 0.9-37.2) higher given DMTS(2s) compared to the Standard Training-IRAP (despite $p = .42$).

Results of the post-hoc analysis suggested that this difference may have been partly attributable to participants attempting to respond quickly during the probe blocks, perhaps influenced by the previous training context that demanded fast responding during this particular procedure. Therefore, it could be useful for future studies to systematically investigate the potential impact of this variable. For example, would adding an instruction before probe blocks stating that there is no longer need for fast responding result in increased yield? In any case, the current findings indicate that the Standard Training-IRAP procedure and its modified versions[3] are statistically as effective as DMTS(2s), but both modified Training-IRAP versions seem to be the most effective, visually speaking. Of course, these findings would need to be replicated in future studies.

Interestingly, and perhaps in contrast to the current findings, a number of other studies in the literature have reported the utility of the Standard Training-IRAP for training and testing derived relations (e.g., Harte et al., 2018, 2020, 2021; Leech & Barnes-Holmes, 2020; Leech et al., 2018). It should be noted, however, that the quantity of stimuli involved in these studies were notably fewer than the number employed currently. For example, Harte et al. (2020) trained A1B1C1 and A2B2C2 before testing for derived AC relations, while the current study involved a larger number of relations trained and tested, as well as employing more stringent performance requirements. It should also be noted that the primary focus of these other studies was not to methodologically test the procedure but rather to produce

[3] The reader should note that the modifications yielding the Modified Training-IRAP resulted in a procedure that is somewhat similar to what has been termed the Relational Evaluation Procedure (REP, a method designed for participants to report or evaluate stimulus relations; see, for example, Barnes-Holmes et al., 2001; Hayes et al., 2016). This may not be surprising given that the IRAP itself was developed, in part, from the REP (see, for example, Barnes-Holmes et al., 2010, and Barnes-Holmes & Harte, 2022b, for further details). Given that the current study began with the Standard Training-IRAP and involved making modifications to that format specifically, we have found it more accurate and appropriate to refer to the resulting procedures as modified versions of the Training-IRAP rather than REPs. Furthermore, the REP refers to a wide range of performances and procedures (e.g., Hayes et al., 2016, compared to Stewart et al., 2004) whereas the Training-IRAP generally refers to a more uniform training preparation (e.g., Harte et al., 2021, compared to Leech et al., 2018). Nonetheless, it is important to recognize that the modified versions of the Training-IRAP involved in the current study may be seen as REP-like, and thus the resulting discussions and implications may also be relevant for those who use or want to use the REP.

derived relations toward other ends (e.g., investigating their role in persistent derived rule-following or fear and avoidance responding). In any case, the visually poorer yield observed from the Standard Training-IRAP in the current study relative to the others in the literature which employed fewer stimuli may suggest potential boundary conditions for use of the Standard Training-IRAP in this way. Future work could systematically explore this suggestion.

One suggestion made in the Introduction was that the different sources of stimulus control potentially at play within MTS versus Training-IRAP formats may impact the yield produced by these procedures (e.g., S+/S- control; see Carrigan & Sidman, 1992). As suggested in the Introduction, the four trial-type structure of the Training-IRAP (i.e., A1/B1-Same, A1/B2-Different, A2/B1-Different, A2/B2-Same), in contrast to MTS (i.e., if A1 then B1, if A2 then B2), may circumvent the issue of stimulus control by specifying each stimulus relation and thereby may generate a higher yield (see Harte, Barnes-Holmes, Moreira, et al., 2021, for a relevant discussion). However, despite a visually higher yield produced by the modified versions of Training-IRAP, the lack of statistically significant results relative to DMTS(2s) raise doubts about whether increased experimental control of specific stimulus relations alone is sufficient to ensure higher yield. If a training format, such as the Training-IRAP, controls and ensures both Stimulus Control Topography (SCTs; Carrigan & Sidman, 1992) for all baseline relations, by specifying the stimulus relations between Sample/Comparison S+ or S- (e.g., A1/B1-Same, A1/B2-Different), but does not show a statistically significant impact on yield relative to the MTS training format, then the current results do not support the hypothesis that controlling the stimulus relations and ensuring both SCTs for all baseline relations are determinants of derived relations (e.g., Arantes & de Rose, 2015; de Rose et al., 2013; Grisante et al., 2014). One possible alternative explanation for why these studies reported high yield could be that they all included a “pre-training” session before the experimental phases. In this session, participants underwent multiple-exemplar training (MET) with stimuli different from those employed experimentally but presented in the same task structure. In principle, this “pre-training” session could have led to generalized responding and improved learning in the experimental task (e.g., Nedelcu et al., 2015; Saunders & Spradlin, 1990, 1993). Therefore, it could be useful for future work to explore the impact of MET on yield using different methodological procedures, such as those employed in the current study.

The secondary objective of the current study was to investigate the average number of training blocks needed for participants to demonstrate baseline criterion using DMTS(2s) and methodological variations of the Standard Training-IRAP. The results showed that participants needed significantly fewer baseline relation training blocks using the DMTS(2s) procedure compared to both the Standard Training-IRAP and Modified Training-IRAP. However, no significant difference was found between DMTS(2s) and Delayed Modified Training-IRAP(2s). As such, it appears that although participants in the Standard Training-IRAP and Modified Training-IRAP blocks tended to require more baseline relation training blocks compared to DMTS(2s), the addition of the observational response and 2-second delay components to the Modified Training-IRAP resulted in increased efficiency in baseline relation training to near DMTS(2s) levels. With respect to the Standard Training-IRAP and its modified versions, there was a reduction in the average number of baseline relation training blocks required by participants from the Standard Training-IRAP (in which participants required the most baseline relation blocks) to the Delayed Modified Training-IRAP (2s; in which participants required the least baseline relation blocks). This finding indicates that the Delayed Modified Training-IRAP(2s) contributed to baseline relation criterion being reached with the fewest number of baseline training blocks of any of the Training-IRAP variations, and statistically significantly so relative to the Standard Training-IRAP. As such, it seems that the use of an observational response and 2-second delay component may have been important variables with respect to establishing baseline relation training, but not necessarily final yield. Interestingly, this latter result seems to be at odds with some findings from the wider MTS literature in which yield reportedly improves as a function of increasing such presentation delays during training (Arntzen, 2006). However, the current findings need to be replicated, and the impact of different delays on yield in the context of the Training-IRAP format could be explored (e.g., 4, 6, and 12-second delays; e.g., Arntzen, 2006; Lian & Arntzen, 2013).

In addition, a further practical (and indeed conceptual) consideration seems to arise with respect to what is arguably established when employing these different methodologies. That is, both procedures can be used to establish equivalence or coordination relations between stimuli, but given the differences in training formats between these procedures, it seems that MTS does not establish responding in accordance with a particular contextual cue (e.g., sameness). The Training-IRAP on the other hand, more explicitly establishes such cues within its format. Awareness of such differences seems important for researchers to consider depending on their procedural and analytic goals. Indeed, it may be useful for future research to explore these issues.

Before concluding, it may be useful to briefly consider other areas within the study of derived relations that may connect with the current work, beyond a focus on yield alone. First, research to date has explored the impact of different MTS procedures on subsequent transfer of stimulus functions (e.g., Simultaneous Matching-to-Sample [SMTS] versus DMTS; Bortoloti & de Rose, 2009, 2012). This work has generally found that a higher level of functional transfer is usually observed with DMTS (Bortoloti & de Rose, 2009, 2012). It may be useful, therefore, to explore any potential differences between establishing derived relations via MTS and the Standard Training-IRAP (and indeed its variations) and their impact on subsequent transformations of stimulus functions (a broader term than *transfer*; see

Hayes & Barnes, 1997; Hayes et al., 2001). This exploration aims to understand how subtleties in learning history may (or may not) differentially impact this effect. Indeed, most research on transformations of function to date seems to rely largely upon the MTS training format (e.g., Bortoloti et al., 2019; Gomes et al., 2019; Perez et al., 2019; Schmidt et al., 2021), and as far as we know, only two studies have thus far employed the Standard Training-IRAP format in this regard (i.e., Leech & Barnes-Holmes, 2020; Leech et al., 2018). Systematic explorations of the impact of MTS versus the Training-IRAP and the variations described currently may be worthy of further experimental inquiry.

Finally, the current findings may be useful in applied settings. That is, the Standard Training-IRAP format has been successfully employed to teach relational responding in autistic children (Murphy & Barnes-Holmes, 2017; Murphy et al., 2019). Given the efficiency of the Delayed Modified Training-IRAP(2s) demonstrated currently, it may be useful to explore whether this increased efficacy (relative to the Standard Training-IRAP) will also hold in this population and indeed in teaching contexts involving neurotypical and neurodivergent children more generally. Moreover, being able to respond to stimuli as both coordinate and different has been recently highlighted as a key element relevant to a range of language and academic behavioral repertoires (see Ming & Stewart, 2017, for a review). Given that the structure of the Delayed Modified Training-IRAP (2s; and indeed, the other Training-IRAP variations) allows for the teaching of both of these relations simultaneously, this general preparation may further lend itself to teaching contexts. But of course, such efficacy remains an empirical matter and should be subject to future systematic investigation.

Conclusion

This current study represents a first systematic exploration of the utility of the Standard Training-IRAP (and its modified versions) on yield and baseline relation training. While important similarities and differences between the procedures and potentially important variables have emerged, as well as further questions, it seems important to not only focus on which procedure is “better,” but to also consider a larger picture. That is, it seems important to strive to demonstrate and observe the phenomenon of interest (in this case derived relational responding) across different procedures, so that it may be liberated from any specific methodology toward an overarching goal of greater precision, scope and depth (see, for example, Clayton & Hayes, 2004; Hayes & Barnes, 1997; Hayes et al., 2012). We believe that the current work contributes, at least in part, to this ongoing effort.

References

Arantes, A., & de Rose, J. C. (2015). High probability of equivalence class formation with both Sample-S+ and Sample-S-controlling relations in baseline. *The Psychological Record*, *65*, 743-748. https://doi.org/10.1007/s40732-015-0143-2

Arntzen, E. (2006). Delayed matching-to-sample: Probability of stimulus equivalence as a function of delays between sample and comparison stimuli during training. *The Psychological Record*, *56*(13), 135-167. https://doi.org/10.1007/BF03395541

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

Barnes-Holmes, D., & Harte, C. (2022a). Relational frame theory 20 years on: The Odysseus voyage and beyond. *Journal of the Experimental Analysis of Behavior*, *117*(2), 240-266. https://doi.org/10.1002/jeab.733

Barnes-Holmes, D., & Harte, C. (2022b). The IRAP as a measure of implicit cognition: A case of Frankenstein's Monster. *Perspectives on Behavior Science*, *45*(3), 559-578. https://doi.org/10.1007/s40614-022-00352-z

Barnes-Holmes, D., Barnes-Holmes, Y., Stewart, I., & Boles, S. (2010). A sketch of the Implicit Relational Assessment Procedure (IRAP) and the Relational Elaboration and Coherence (REC) model. *The Psychological Record, 60,* 527-542. https://doi.org/10.1007/BF03395726

Barnes-Holmes, D., Finn, M., McEnteggart, C., & Barnes-Holmes, Y. (2018). Derived stimulus relations and their role in a behavior-analytic account of human language and cognition. *Perspectives on Behavior Science, 41,* 155-173. https://doi.org/10.1007/s40614-017-0124-7

Barnes-Holmes, D., Hayes, S. C., Dymond, S., & O'Hora, D. (2001). Multiple stimulus relations and the transformation of stimulus functions. In S. C. Hayes, D. Barnes-Holmes, & B. Roche (Eds.), *Relational Frame Theory: A post-Skinnerian account of human language and cognition* (pp. 51-71). Kluwer Academic/ Plenum Publishers

Bortoloti, R., & de Rose, J. C. (2009). Assessment of the relatedness of equivalent stimuli through a semantic differential. *The Psychological Record*, *59*(4), 563-590. https://doi.org/10.1007/BF03395682

Bortoloti, R., & de Rose, J. C. (2012). Equivalent stimuli are more strongly related after training with delayed matching than after simultaneous matching: A study using the Implicit Relational Assessment Procedure (IRAP). *The Psychological Record*, *62*(1), 41-54. https://doi.org/10.1007/BF03395785

Bortoloti, R., de Almeida, R. V., de Almeida, J. H., & de Rose, J. C. (2019). Emotional faces in symbolic relations: A happiness superiority effect involving the equivalence paradigm. *Frontiers in Psychology*, *10*, 1-12. https://doi.org/10.3389/fpsyg.2019.00954

Carrigan, P. F., & Sidman, M. (1992). Conditional discrimination and equivalence relations: A theoretical analysis of control by negative stimuli. *Journal of the Experimental Analysis of Behavior*, *58*(1), 183-204. https://doi.org/10.1901/jeab.1992.58-183

Clayton, M. C., & Hayes, L. J. (2004). A comparison of match-to-sample and respondent-type training of equivalence classes. *The Psychological Record*, *54*, 579-602. https://doi.org/10.1007/BF03395493

de Rose, J. C. (1996). Controlling factors in conditional discriminations and tests of equivalence. In T. R. Zentall & P. M. Smeets (Eds.), *Advances in Psychology* (Vol. 117, pp. 253-277). North-Holland. https://doi.org/10.1016/S0166-4115(06)80112-3 de Rose, J. C., Hidalgo, M., & Vasconcellos, M. (2013). Controlling relations in baseline conditional discriminations as determinants of stimulus equivalence. *The Psychological Record*, *63*(1), 85-98. https://doi.org/10.11133/j.tpr.2013.63.1.007

Dougher, M. J., Augustson, E., Markham, M. R., Greenway, D. E., & Wulfert, E. (1994). The transfer of respondent eliciting and extinction functions through stimulus equivalence classes. *Journal of the experimental analysis of behavior*, *62*(3), 331-351. https://doi.org/10.1901/jeab.1994.62-331

Fields, L., Arntzen, E., & Doran, E. (2020). Yield as an essential measure of equivalence class formation, other measures, and new determinants. *The Psychological Record*, *70*(2), 175-186. https://doi.org/10.1007/s40732-020-00377-3

Fienup, D. M., & Brodsky, J. (2020). Equivalence-based instruction: Designing instruction using stimulus equivalence. In M. J. Fryling, R. A. Rehfeldt, J. Tarbox, & L. J. Hayes (Eds.), *Applied behavior analysis of language and cognition: Core concepts and principles for practitioners* (1st ed., Vol 1., pp. 157-173). Context Press.

Gomes, C. T., Perez, W. F., de Almeida, J. H., Ribeiro, A., de Rose, J. C., & Barnes-Holmes, D. (2019). Assessing a derived transformation of functions using the implicit relational assessment procedure under three motivative conditions. *The Psychological Record*, *69*(4), 487-497. https://doi.org/10.1007/s40732-019-00353-6

Grisante, P. C., de Rose, J. C., & McIlvane, W. J. (2014). Controlling relations in stimulus equivalence classes of preschool children and individuals with Down syndrome. *The Psychological Record*, *64*, 195-208. https://doi.org/10.1007/s40732-014-0021-3

Harrison, R. J., & Green, G. (1990). Development of conditional and equivalence relations without differential consequences. *Journal of the Experimental Analysis of Behavior*, *54*(3), 225-237. https://doi.org/10.1901/jeab.1990.54-225

Harte, C., Barnes-Holmes, D., Barnes-Holmes, Y., & McEnteggart, C. (2018). The impact of high versus low levels of derivation for mutually and combinatorially entailed relations on persistent rule-following. *Behavioural Processes*, *157*, 36-46. https://doi.org/10.1016/j.beproc.2018.08.005

Harte, C., Barnes-Holmes, D., Barnes-Holmes, Y., & McEnteggart, C. (2021). Exploring the impact of coherence (through the presence versus absence of feedback) and levels of derivation on persistent rule-following. *Learning and Behavior*, *49*, 222-239. https://doi.org/10.3758/s13420-020-00438-1

Harte, C., Barnes-Holmes, D., Barnes-Holmes, Y., McEnteggart, C., Gys, J., & Hasler, C. (2020). Exploring the potential impact of relational coherence on persistent rule-following: The first study. *Learning & Behavior*, *48*, 373-391. https://doi.org/10.3758/s13420-019-00399-0

Harte, C., Barnes-Holmes, D., Moreira, M., de Almeida, J. H., Passarelli, D., & de Rose, J. C. (2021). Exploring a Training IRAP as a single participant context for analyzing reversed derived relations and persistent rule-following. *Journal of the Experimental Analysis of Behavior*, *115*(2), 460-480. https://doi.org/10.1002/jeab.671

Hayes, J., Stewart, I., & McElwee, J. (2016). Assessing and training young children in same and different relations using the relational evaluation procedure (REP). *The Psychological Record*, 66, 547-561. https://doi.org/10.1007/s40732-016-0191-2

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.). (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition*. Kluwer Academic/ Plenum Publishers.

Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (2012). *Acceptance and commitment therapy: The process and practice of mindful change* (2nd ed.). Guilford Press.

Hayes, S., & Barnes, D. (1997). Analyzing derived stimulus relations requires more than the concept of stimulus class. *Journal of the Experimental Analysis of Behavior*, *68*(2), 235-244. https://doi.org/10.1901/jeab.1997.68-235

Johnson, C., & Sidman, M. (1993). Conditional discrimination and equivalence relations: Control by negative stimuli. *Journal of the Experimental Analysis of Behavior*, *59*(2), 333-347. https://doi.org/10.1901/jeab.1993.59-333

Leech, A., & Barnes-Holmes, D. (2020). Training and testing for a transformation of fear and avoidance functions via combinatorial entailment using the Implicit Relational Assessment Procedure (IRAP): Further exploratory analyses. *Behavioural processes*, *172*, 104027. https://doi.org/10.1016/j.beproc.2019.104027

Leech, A., Bouyrden, J., Bruijsten, N., Barnes-Holmes, D., & McEnteggart, C. (2018). Training and testing for a transformation of fear and avoidance functions using the Implicit Relational Assessment Procedure: The first study. *Behavioural processes*, *157*, 24-35. https://doi.org/10.1016/j.beproc.2018.08.012

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

Martins, T. H. S., Rodrigues, R. M., Araújo, F. C. O., Cedro, A. M., Bortoloti, R., Varella, A. A. B., & Huziwara, E. M. (2023). Transfer of functions based on equivalence class formation using musical stimuli. *Journal of the Experimental Analysis of Behavior*, *120*(3), 394-405. https://doi.org/10.1002/jeab.881

Ming, S., & Stewart, I. (2017). When things are not the same: A review of research into relations of difference. *Journal of applied behavior analysis*, *50*(2), 429-455. https://doi.org/10.1002/jaba.367

Murphy, C., & Barnes-Holmes, D. (2017). Teaching important relational skills for children with autism spectrum disorder and intellectual disability using freely available (GO-IRAP) software. *Austin journal of autism & related disabilities*, *3*(2), 1-6.

Murphy, C., Lyons, K., Kelly, M., Barnes-Holmes, Y., & Barnes-Holmes, D. (2019). Using the Teacher IRAP (T-IRAP) interactive computerized programme to teach complex flexible relational responding with children with diagnosed autism spectrum disorder. *Behavior Analysis in Practice*, *12*, 52-65. https://doi.org/10.1007/s40617-018-00302-9

Nedelcu, R. I., Fields, L., & Arntzen, E. (2015). Arbitrary conditional discriminative functions of meaningful stimuli and enhanced equivalence class formation. *Journal of the Experimental Analysis of Behavior*, *103*(2), 349-360. https://doi.org/10.1002/jeab.141

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior research methods*, *51*, 195-203. https://doi.org/10.3758/s13428-018-01193-y

Perez, W. F., & Tomanari, G. Y. (2013). Inferring the occurrence of select and reject controls in matching-to-sample tasks: a methodological review. *Acta Comportamentalia*, *21*(2), 211-225.

Perez, W. F., & Tomanari, G. Y. (2014). Indução de controles por seleção e por rejeição em tarefas de emparelhamento com o modelo: Uma revisão metodológica. *Acta Comportamentalia: Revista Latina de Análisis de Comportamiento*, *22*(2), 227-242.

Perez, W. F., de Almeida, J. H., de Rose, J. C., Dorigon, A. H., de Vasconcellos, E. L., da Silva, M. A., Lima, N. D. P., de Almeida, R. B. M., Montan, R. N. M., & Barnes-Holmes, D. (2019). Implicit and explicit measures of transformation of function from facial expressions of fear and of happiness via equivalence relations. *The Psychological Record*, *69*(1), 13-24. https://doi.org/10.1007/s40732-018-0304-1

Perez, W. F., de Almeida, J. H., Soares, L. C., Wang, T. F., de Morais, T. E., Mascarenhas, A. V., & de Rose, J. C. (2020). Fearful faces and the derived transfer of aversive functions. *The Psychological Record*, *70*(3), 387-396. https://doi.org/10.1007/s40732-020-00390-6

Perez, W. F., Huziwara, E. M., Rodrigues, R. M., Vilela, E. C., Tomanari, G. Y., & Vaidya, M. (2020). Effect of observing response requirements to sample and comparison stimuli on the establishment of reject control (sample/S-relations). *Journal of the experimental analysis of behavior*, *114*(1), 60-71. https://doi.org/10.1002/jeab.602

Perez, W. F., Tomanari, G. Y., & Vaidya, M. (2015). Effects of select and reject control on equivalence class formation and transfer of function. *Journal of the experimental analysis of behavior*, *104*(2), 146-166. https://doi.org/10.1002/jeab.164

Perez, W. F., Tomanari, G. Y., & Vaidya, M. (2017). Effects of heterogeneous controlling relations on tests of transitivity and equivalence: An exploratory study. *European Journal of Behavior Analysis*, *18*(2), 251-263. https://doi.org/10.1080/15021149.2017.1309957

Perez, W. F., Vaidya, M., Huziwara, E. M., & Tomanari, G. Y. (2019). Empirical analysis of biasing procedures to establish reject-control relations. *European Journal of Behavior Analysis*, *20*(2), 244-260. https://doi.org/10.1080/15021149.2019.1677079

Perez, W. F., Zuppani, T. S., Dorigon, A. H., de Vasconcellos, E. L., da Silva, M. A., Lima, N. D. P., de Almeida, R. B. M., Montan, R. N. M., de Almeida, J. H., & de Rose, J. C. (2023). The Transfer of Meaning via Contextually Controlled Equivalence Relations. *The Psychological Record*, *73*, 587-592. https://doi.org/10.1007/s40732-023-00568-8

Pilgrim, C. (2020). Equivalence-based instruction. In J. O. Cooper, T. E. Heron, & W. L. Heward, *Applied behavior analysis* (3rd ed. Global Edition, pp. 442–496). Pearson UK.

Plazas, E. A. (2019). Transfer of baseline reject control to transitivity trials and its effect on equivalence class formation. *Journal of the Experimental Analysis of Behavior, 111*(3), 465-478. https://doi.org/10.1002/jeab.519

Plazas, E. A., & Peña, T. E. (2016). Effects of procedural variations in the training of negative relations for the emergence of equivalence relations. *The Psychological Record*, *66*, 109-125. https://doi.org/10.1007/s40732-015-0157-9

Plazas, E. A., & Villamil, C. W. (2016a). Effects of between-classes negative relations training on equivalence class formation across training structures. *The Psychological Record*, *66*, 489-501. https://doi.org/10.1007/s40732-016-0189-9

Plazas, E. A., & Villamil, C. W. (2016b). Efecto del entrenamiento de relaciones negativas entre-clases y estructuras de entrenamiento en la formación de relaciones de equivalencia. *International Journal of Psychology and Psychological Therapy*, *16*(3), 265-314.

Saunders, K. J., & Spradlin, J. E. (1990). Conditional discrimination in mentally retarded adults: The development of generalized skills. *Journal of the Experimental Analysis of Behavior*, *54*(3), 239-250. https://doi.org/10.1901/jeab.1990.54-239

Saunders, K. J., & Spradlin, J. E. (1993). Conditional discrimination in mentally retarded subjects: Programming acquisition and learning set. *Journal of the Experimental Analysis of Behavior*, *60*(3), 571-585. https://doi.org/10.1901/jeab.1993.60-571

Schmidt, M., de Rose, J. C., & Bortoloti, R. (2021). Relating, orienting and evoking functions in an IRAP study involving emotional pictographs (emojis) used in electronic messages. *Journal of Contextual Behavioral Science, 21,* 80-87. https://doi.org/10.1016/j.jcbs.2021.06.005

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

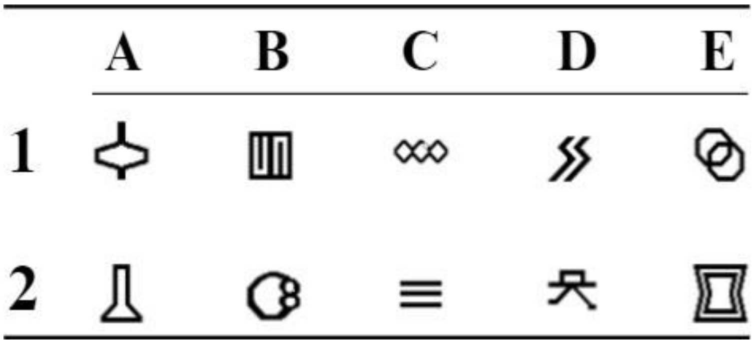
Sidman, M. (1994). *Equivalence relations and behavior: A research story*. Authors Cooperative.

Stewart, I., Barnes-Holmes, D., & Roche, B. (2004). A functional-analytic model of analogy using the relational evaluation procedure. *The Psychological Record*, *54*(4), 531-552. https://doi.org/10.1007/BF03395491

Szumilas, M. (2010). Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, *19*(3), 227-229.

Figure 1

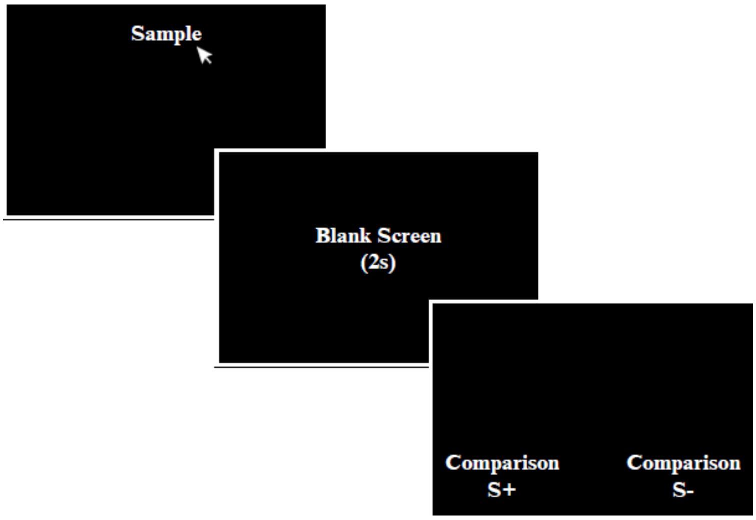
Stimuli used in the experiment



Note: each stimulus is labeled with a number and a letter. Numbers indicate the class (i.e., class 1 and class 2), and letters indicate the exemplar within that class.

Figure 2

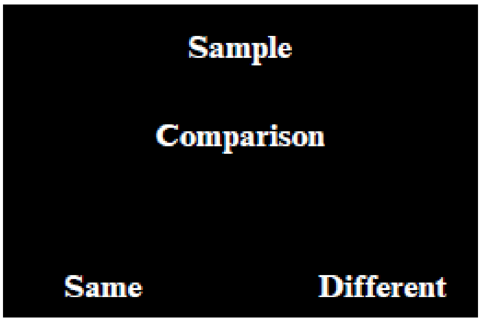
Training Structure of Delayed Matching-to-Sample(2s)



Note: In each trial, one of the sample stimuli were presented. An observational response to the sample was conducted by clicking on that stimulus with a computer mouse. The blank screen was represented by a completely black screen and lasted two seconds. The positions of the stimulus comparisons (i.e., the experimentally correct - S+, or the incorrect - S-) were randomized across trials. The selection of one of the comparisons was made by clicking on the stimulus with a computer mouse.

Figure 3

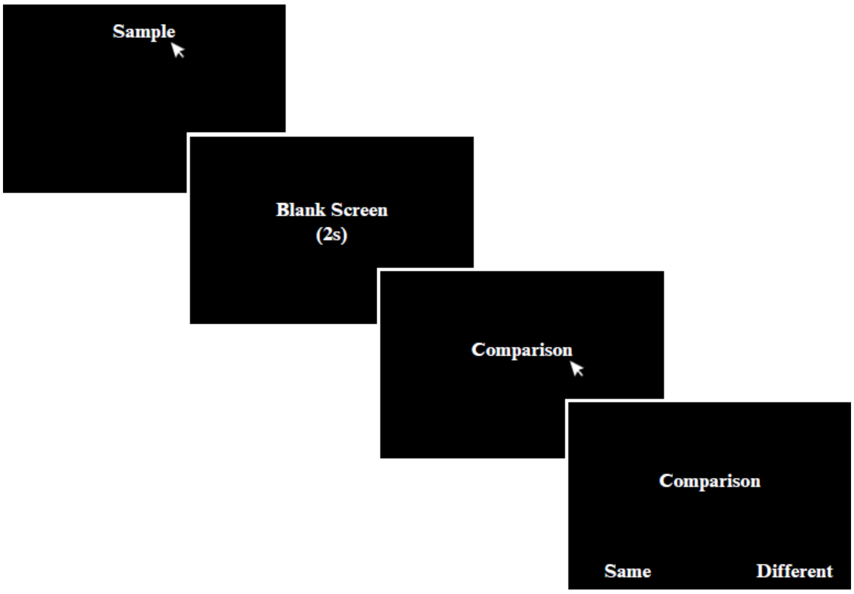
Training Structure of Standard Training-IRAP



Note: In each trial, one of the sample stimuli were presented along with one of the comparison stimuli, and two response options (i.e., “Same” and “Different”) with their positions randomized between trials. The selection of one of the response options was made using the keyboard. The 'd' key on the keyboard was used to select the response option on the left, and the ‘k’ key was used to select the option on the right.

Figure 4

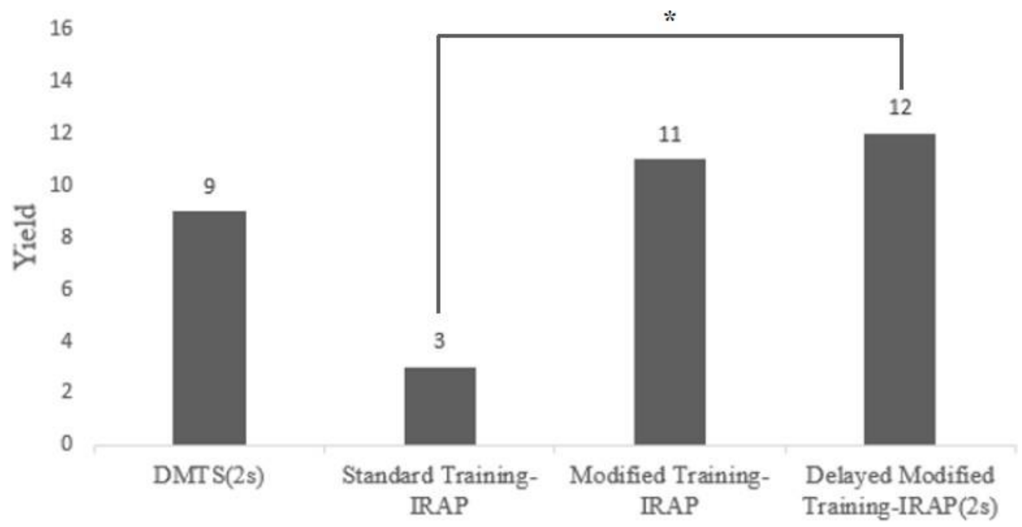
Training Structure of Delayed Modified Training-IRAP(2s)



Note: In each trial, one of the sample stimuli were presented. An observational response to the sample was conducted by clicking on that stimulus with a computer mouse. The blank screen was represented by a completely black screen and lasted two seconds. One of the comparison stimuli were then presented, an observational response to which was required to proceed. After the observational response, the comparison stimulus remained on the screen and two response options (i.e., “Same” and “Different”) with their positions randomized across trials was immediately presented. The selection of one of the response options was made by clicking on with a computer mouse.

Figure 5

Yield per group



Note. Each bar represents the number of participants per group that successfully formed the class in accordance with experimental criteria.

*Statistically significant difference

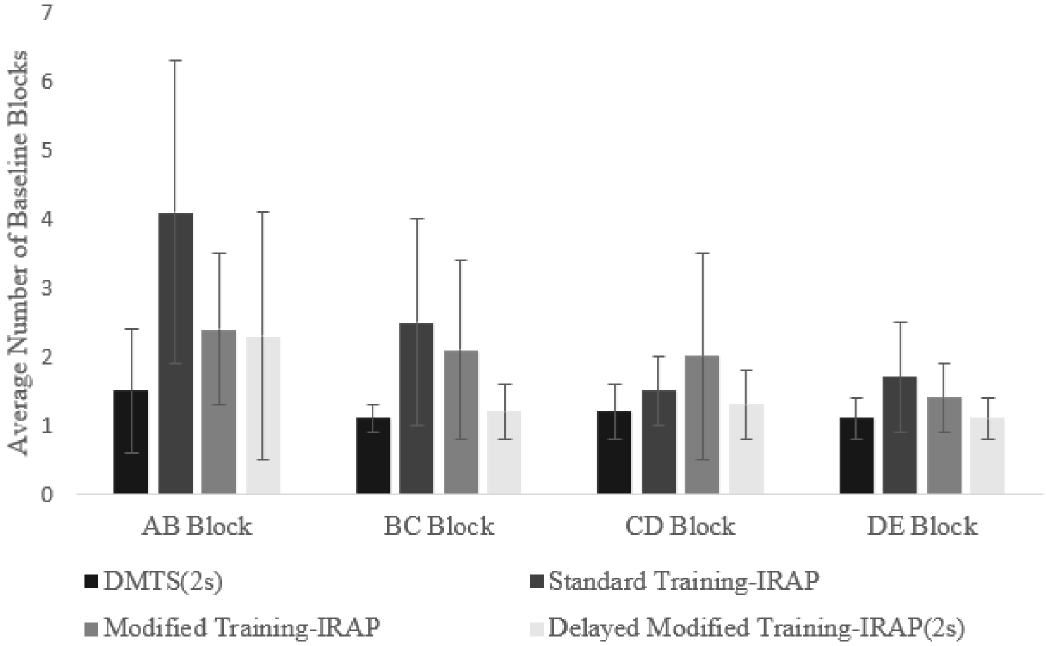
Table 1

Average Number of Blocks to Complete the Training Phases

Procedure	Baseline	Cumulative Baseline	Cumulative Baseline Without Differential Consequences
DMTS(2s)	4.8	1.2	1.0
Standard Training-IRAP	11.2	2.7	1.5
Modified Training-IRAP	8.9	1.8	1.0
Delayed Modified Training-IRAP(2s)	6.2	1.8	1.1

Figure 6

Average Number of AB, BC, CD, and DE Baseline Blocks



Note: each bar represents the average number of baseline blocks needed per procedure in AB, BC, CD and DE, along with their respective standard deviations.

Appendix A

Number of Training Blocks

The tables present individual participant data regarding the number of training blocks in DMTS(2s; Table 1), Standard Training-IRAP (Table 2), Modified Training-IRAP (Table 3), and Delayed Modified Training-IRAP(2s; Table 4).

Table 1

Number of Training Blocks in DMTS(2s) for Each Individual Participant

	Baseline				Cumulative Baseline	Cumulative Baseline Without Differential Consequences
	AB	BC	CD	DE		
P01	4	1	1	1	2	1
P02*	1	1	2	1	1	1
P03	1	1	1	1	1	1
P04	2	1	1	1	2	1
P05*	1	1	1	2	1	1
P06*	1	1	2	1	1	1
P07*	1	1	1	1	1	1
P08*	1	1	2	1	1	1
P09*	1	1	1	1	1	1
P10	1	1	1	1	1	1
P11	3	1	1	1	2	1
P12	2	2	1	2	1	1
P13	1	1	1	1	1	1
P14*	2	1	1	1	1	1
P15*	1	1	1	1	2	1
P16*	1	1	1	1	1	1
P17	1	1	1	1	1	1

*Participants who pass the derivation tests.

Table 2

Number of Training Blocks in Standard Training-IRAP for Each Individual Participant

	Baseline				Cumulative Baseline	Cumulative Baseline Without Differential Consequences
	AB	BC	CD	DE		
P01	2	3	2	3	2	2
P02	3	1	2	2	2	1
P03	6	4	3	4	4	1
P04	4	2	1	1	3	1
P05	3	9	3	3	4	2

P06*	4	2	1	1	3	4
P07	10	4	3	3	5	1
P08	4	2	2	1	1	1
P09	5	2	1	1	2	1
P10	4	2	1	3	1	1
P11*	3	8	2	2	4	2
P12	5	3	2	2	5	1
P13	2	3	2	2	3	1
P14	3	1	2	1	2	1
P15	7	2	1	2	2	1
P16*	2	1	1	1	2	1
P17	8	2	2	1	1	3

*Participants who pass the derivation tests.

Table 3

Number of Training Blocks in Modified Training-IRAP for Each Individual Participant

	Baseline				**Cumulative Baseline**	**Cumulative Baseline Without Differential Consequences**
	AB	**BC**	**CD**	**DE**		
P01	5	3	1	2	2	1
P02*	4	6	7	2	1	1
P03*	4	1	2	1	1	1
P04*	3	1	2	2	1	1
P05	3	4	4	3	2	1
P06*	3	1	1	1	1	1
P07*	3	2	2	2	2	1
P08	4	3	4	2	3	1
P09	2	3	1	1	3	1
P10	2	1	4	1	2	1
P11	3	5	2	3	4	1
P12*	2	2	1	1	1	1
P13*	2	2	1	1	1	1
P14*	1	1	1	1	1	1
P15*	2	2	3	2	4	1
P16*	1	1	1	1	1	1
P17*	2	2	2	1	1	1

*Participants who pass the derivation tests.

Table 4

Number of Training Blocks in Delayed Modified Training-IRAP(2s) for Each Individual Participant

	Baseline				**Cumulative Baseline**	**Cumulative Baseline Without Differential Consequences**
	AB	**BC**	**CD**	**DE**		
P01	8	1	1	1	2	1
P02*	2	1	1	2	1	1
P03*	2	1	1	1	1	1
P04	5	2	2	2	2	1
P05*	2	1	1	1	2	1
P06*	3	2	2	2	5	1
P07*	1	2	1	1	1	1
P08*	1	1	1	1	1	1
P09	3	1	1	1	2	1
P10	2	1	2	1	3	1
P11*	1	1	2	1	1	1
P12	2	1	1	1	1	1
P13*	3	2	2	1	1	2
P14*	1	1	1	1	1	1
P15*	1	1	2	1	1	1
P16*	2	1	1	1	2	1
P17*	1	2	1	1	3	1

*Participants who pass the derivation tests.

Appendix B

These programs were written in PsychoPy v2020.2.9 using Python language. They are free software: you can redistribute them and/or modify them. Remember to mention the modifications.

You can access them by GitHub.com.

Modified Training-IRAP

https://github.com/marcellosilvestre/Modified-Training-IRAP

Delayed Modified Training-IRAP(2s)

https://github.com/marcellosilvestre/Delayed-Modified-Training-IRAP-2s-