ESTABLISHING AND EVALUATING TOKENS AS CONDITIONED REINFORCERS

By

TRACY ARGUETA

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2022
To my parents who sacrificed so much so that I could be anything and to Alex who left us too soon.
ACKNOWLEDGMENTS

First, I want to express my deepest gratitude to my parents for their unconditional love and support. Without them, I would not be who and where I am today. I am also deeply thankful to my sister for her love, kindness, and encouragement throughout this journey. She is my rock. I would also like to thank my friends for their love and support, and for allowing me to lean on them time and time again.

I also want to thank my advisor, Dr. Iser G. DeLeon for his invaluable guidance and mentorship. It has been a privilege to learn from him to be a better scientist, teacher, and student. Thank you for sharing your love and enthusiasm for science with me and thank you for your confidence in my abilities. I would also like to thank my committee members, Drs. Tim Vollmer, Brian Reichow, and Darragh Devine for their thoughtful feedback and guidance on this dissertation.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
</tr>
<tr>
<td>ABSTRACT</td>
</tr>
</tbody>
</table>

## CHAPTER

1. INTRODUCTION | 11 |
2. STUDY 1 | 19 |
   - Procedures | 19 |
     - Eligibility Criteria | 19 |
     - Search Methods and Identification of Studies | 20 |
     - Data Extraction | 21 |
     - Data Analysis | 21 |
   - Results | 23 |
   - Discussion | 25 |
3. STUDY 2 | 36 |
   - Method | 36 |
     - Participants, Setting, and Materials | 36 |
     - Response Measurement | 37 |
     - Pre-Experimental Procedures | 38 |
       - Preference and reinforcer assessments | 38 |
       - Verbal screener | 41 |
     - General Procedure | 41 |
       - Baseline | 42 |
       - Token baseline | 43 |
       - S-S pairing | 43 |
       - R-S pairing | 43 |
       - S-E-S pairing | 44 |
       - R-S-E pairing | 44 |
       - Thinned token-exchange test | 44 |
   - Results | 46 |
   - Discussion | 49 |
4. STUDY 3 | 56 |
   - Method | 56 |
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>List of included studies with success estimates (SEs) per study and procedure. ( ^{U} ) indicates the study is an unpublished dissertation or thesis.</td>
<td>31</td>
</tr>
<tr>
<td>2-2</td>
<td>Types of target stimuli by study. ( ^{U} ) indicates studies that were unpublished dissertations or theses.</td>
<td>34</td>
</tr>
<tr>
<td>2-3</td>
<td>Types of existing reinforcers by study. ( ^{U} ) indicates studies that were unpublished dissertations or theses.</td>
<td>35</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>------</td>
</tr>
<tr>
<td>2-1</td>
<td>PRISMA flowchart for article screening.</td>
<td>30</td>
</tr>
<tr>
<td>2-2</td>
<td>Success estimates in percentages for each procedure (left panel) and the frequency of successes to implementations per procedure (right panel).</td>
<td>32</td>
</tr>
<tr>
<td>2-3</td>
<td>Subgroup analyses for the type and number of target stimuli (upper panel) and existing reinforcer(s) (lower panel).</td>
<td>33</td>
</tr>
<tr>
<td>3-1</td>
<td>Results for S-S, R-S, S-E-S, and R-S-E pairing for Doug (top panel), Arnold (middle panel), and Desmond (bottom panel).</td>
<td>54</td>
</tr>
<tr>
<td>3-2</td>
<td>Results for S-S, R-S, S-E-S, and R-S-E pairing for Miguel.</td>
<td>55</td>
</tr>
<tr>
<td>4-1</td>
<td>Cumulative correct responses for FR-VR comparison for Ignacio (top panel) and Zola (bottom panel).</td>
<td>69</td>
</tr>
<tr>
<td>4-2</td>
<td>Percent correct responses across sessions for the FR-VR comparison for Ignacio (top panel) and Zola (bottom panel).</td>
<td>70</td>
</tr>
</tbody>
</table>
ESTABLISHING AND EVALUATING TOKENS AS CONDITIONED REINFORCERS

By

Tracy Argueta

August 2022

Token economies are ubiquitous in behavior analytic practice; yet, there is a lack of applied research evaluating methods for establishing tokens as conditioned reinforcers for individuals with limited verbal and imitative repertoires. However, applied researchers have investigated methods for establishing other stimuli as conditioned reinforcers. Therefore, in Study 1, we conducted a systematic review of studies that evaluated stimulus-stimulus (S-S) pairing, response-stimulus (R-S) pairing, and operant discrimination training (ODT) for establishing stimuli as conditioned reinforcers for individuals with intellectual and developmental disabilities. Of the 31 included studies, none attempted to establish tokens as conditioned reinforcers and the majority evaluated R-S pairing. Results revealed that, across the three procedures, pairing is effective in only half of implementations with R-S and S-S pairing being the most and second most effective. Therefore, in Study 2, we evaluated S-S and R-S pairing as well as stimulus-exchange-stimulus (S-E-S) pairing and response-stimulus plus exchange (R-S-E) pairing to establish tokens as conditioned reinforcers in a single-case design with children with autism spectrum disorder. The latter two procedures were included because they have been recommended or described in the literature for establishing
tokens as conditioned reinforcers. Results showed that only R-S and R-S-E pairing were effective at establishing tokens as conditioned reinforcers. However, results under a modified testing condition for R-S-E pairing suggest that the traditional method of testing for conditioned reinforcement effects under extinction conditions might impede detection of the effects of pairing procedures on tokens. Regardless of the pairing procedure used, once tokens are successfully established as conditioned reinforcers, they are often used to target skill acquisition. However, many variables can influence skill acquisition, including schedules of reinforcement. Token economies involve three different schedules of reinforcement each of which might differentially influence skill acquisition. One of these schedules is the exchange-production schedule. Study 3 compared the effects of fixed- vs. variable-ratio exchange-production schedules in a token economy on the rate of skill acquisition for individuals with autism. Results suggest that FR and VR exchange-production schedules produce different rates of skill acquisition. However, the more efficient schedule varies by learner.
Tokens are previously neutral stimuli that increase the future probability of responses they follow and are exchangeable for other reinforcers (i.e., backup reinforcers; Hackenberg, 2018; Hine et al., 2017). Tokens are the second most common consequence delivered by professionals, including behavior analysts who provide services to individuals with neurodevelopmental disorders (Graff & Karsten, 2012). Presumably, tokens function as conditioned reinforcers established as such via pairings with the backup reinforcers for which they are exchangeable (Hackenberg, 2018; Hine et al., 2017). Procedural recommendations vary for pairing tokens with backup reinforcers. For example, Hine et al. (2017) recommended repeated noncontingent token delivery immediately followed by noncontingent backup reinforcer delivery (i.e., stimulus-stimulus [S-S] pairing). By contrast, Kazdin (1977) suggested noncontingent token delivery immediately followed by a token exchange for backup reinforcers. Although these and other recommendations vary, few human studies have directly evaluated token pairing procedures (Hackenberg, 2018), precluding conclusions about the overall and relative efficacy of variations.

Alternatively, researchers have evaluated and compared pairing procedures for establishing stimuli other than tokens as conditioned reinforcers for individuals with neurodevelopmental disorders. For example, Rodriguez and Gutierrez (2017) evaluated response-stimulus (R-S) pairing and a discriminative stimulus (SD) procedure (i.e., operant discrimination training [ODT]) for establishing social stimuli (e.g., hugs, smiles) as conditioned reinforcers for children with autism spectrum disorder (ASD). In R-S pairing, contingent on responding, the experimenters delivered the targeted social
stimulus immediately followed by an existing nonsocial reinforcer. In the SD procedure, the experimenters delivered the existing reinforcer contingent only on responses that occurred following the presentation of the targeted social stimulus. Responses that occurred in the absence of the social stimulus did not produce the reinforcer. Rodriguez and Gutierrez found that R-S pairing was effective at establishing the targeted social stimulus as a conditioned reinforcer for two participants, while the SD procedure was not effective for any participants. In contrast, Taylor-Santa et al. (2014) reported that ODT (similar to Rodriguez and Gutierrez’s SD procedure) was effective at temporarily establishing pictures paired with edibles as conditioned reinforcers for three children with ASD.

Dozier et al. (2012) evaluated S-S and R-S pairing for establishing praise statements as conditioned reinforcers for individuals with intellectual and developmental disabilities (IDDs). During S-S pairing, praise statements were presented and immediately followed by edible reinforcers on a time-based schedule. During R-S pairing, contingent on target responses, the experimenters delivered praise immediately followed by established edible reinforcers. The authors found that S-S pairing was not effective in establishing the praise statements as conditioned reinforcers for any participants, while R-S pairing was effective for four out of eight participants.

Researchers have also evaluated and recommended other procedures, such as observational learning (e.g., Leaf et al., 2012) and providing a verbal description of the contingencies (e.g., Kazdin, 1977) for establishing stimuli as conditioned reinforcers. However, the former requires learners to have a generalized imitative repertoire and appropriate attending skills (MacDonald & Ahearn, 2015), while the latter requires
learners who are responsive to verbal instruction. As such, both procedures lack utility for individuals with neurodevelopmental disorders who have more limited skill sets. In contrast, S-S pairing, R-S pairing, and ODT are less constrained to certain repertoires and, therefore, have utility for a greater proportion of individuals with neurodevelopmental disorders. However, mixed findings, methodological variations across studies, and the lack of direct comparisons within studies limit conclusions about, and necessitate further research on, the relative efficacy of S-S pairing, R-S pairing, and ODT.

The research cited above provides important guidance on establishing conditioned reinforcers. However, tokens differ from other conditioned reinforcers in important ways that call into question the relevance of methods to establish other sorts of stimuli as conditioned reinforcers. For example, unlike conditioned social reinforcers, tokens are directly exchangeable for other reinforcers and might be valuable only insofar as they are exchangeable (Shahan, 2010). Therefore, it is possible that pairing procedures might operate on tokens differently than they do on other stimuli, given the unique nature of tokens. Furthermore, the exchangeability of tokens introduces a potential additional component (i.e., exchange responses) to pairing procedures that is not relevant for other stimuli. For example, Kazdin’s (1977) recommended pairing procedure requires learners to exchange noncontingently delivered tokens to access backup reinforcers. Similarly, Argueta et al. (2019) implemented a multi-phase procedure that, at one point, combined R-S pairing with the method suggested by Kazdin by requiring a response to earn tokens and subsequently requiring a token-exchange to access backup reinforcers. However, Argueta et al. (2019) did not directly
evaluate this procedure nor report corresponding data. As such, further research
evaluating the effectiveness of pairing procedures specific to tokens is also needed.

Once tokens have been established as conditioned reinforcers and token training
is complete, tokens can be incorporated into treatment, in which they are typically used
to target acquisition responses. In fact, 87% of Board Certified Behavior Analysts and
Board Certified Assistant Behavior Analysts® reported using tokens in acquisition
training with individuals with IDD in clinical and educational settings (Fernandez et al.,
2021). Generally, acquisition interventions are intended to be as effective and efficient
as possible (Frank-Crawford et al., 2019). However, many variables impact acquisition,
including but not limited to schedules of reinforcement.

In the context of a token economy, there are several schedules of reinforcement
to consider that might impact response acquisition. Two of these schedules are the
token-production and exchange-production schedules. In acquisition training, the former
specifies how and when tokens will be delivered for acquisition responses, whereas the
latter specifies how and when opportunities to exchange earned tokens for backup
reinforcers will be presented (Hackenberg, 2009). During the initial stages of acquisition,
authors typically recommend using a fixed-ratio (FR) 1 schedule of reinforcement to
establish the target response (e.g., Cooper et al., 2020). In a token economy, this
recommendation involves setting the token-production schedule to an FR 1 such that
each acquisition response meeting a specified criterion results in a token delivery.

Authors generally recommend starting with a dense exchange-production
schedule that is gradually increased over time (e.g., Miltenberger, 2012). In fact,
recommendations usually are to increase the schedule during initial training of the token
economy (e.g., Leaf et al., 2012). As such, it is unsurprising that few clinicians report setting exchange-production schedules at an FR 1 and that most require learners to meet a more stringent response-based schedule (Fernandez et al., 2021). In response-based exchange-production schedules, learners access opportunities to exchange tokens after they have earned a given number of tokens specified by the schedule (Ivy et al., 2017). In contrast, in time-based exchange-production schedules, exchange opportunities are contingent on the passage of time (Ivy et al., 2017).

Response-based exchange-production schedules can be fixed- or variable-ratio. In an FR exchange-production schedule, the learner must earn the same number of tokens to produce each exchange opportunity. In a variable-ratio (VR) exchange-production schedule, the number of tokens required to earn an exchange opportunity varies from one opportunity to the next with the values centered around an average. Basic researchers have documented advantages of VR and similar schedules relative to FR schedules both within and outside the context of token economies. For example, Madden et al. (2005) compared pigeons’ response rates and reinforcer consumption on equivalent FR and random-ratio (RR) food delivery schedules with the same unit price (i.e., ratio of responses to one unit of reinforcer). RR schedules are similar to VR schedules, except that each changing response requirement is not pre-determined. Instead, a random number generator or similar tool determines which responses produce a reinforcer as responding is occurring. When comparing equivalent FR and RR schedules, Madden et al. found that, at lower ratios (i.e., FR vs. RR 48 and below), pigeons’ response rates and reinforcer consumption were similar. However, at higher ratios (i.e., FR vs. RR 192 and above), RR schedules consistently produced greater
rates of responding and reinforcer consumption. These results are interesting because they run counter to predictions that reinforcer efficacy and response rates should be roughly equivalent across schedule types with equivalent unit prices (Hursh, 1988). Similarly, in a token economy, Foster et al. (2001) found that, at higher ratios, pigeons response rates were higher and post-reinforcement pauses (PRPs) were lower under VR exchange-production schedules than under equivalent FR exchange-production schedules. However, Argueta et al. (2019) did not find systematic differences in PRPs or response rates of a mastered task completed by a child with autism under equivalent VR and FR exchange-production schedules. In similar studies with children with autism, Greaves (2008) and Moskovitz (2011) also failed to replicate the rate differences between FR and VR exchange-production schedules found by basic researchers.

Although applied researchers have not found differences in response rates between FR and VR exchange-production schedules, it is possible that differences might emerge in an acquisition context. Previous studies have focused only on mastered responses and have typically measured response rates to detect potential differences across schedules. However, response rate is sometimes an insensitive measure of motivation and preference (Roscoe et al., 1999). Instead, motivation is better measured by choice in concurrent-choice arrangements (Roscoe et al., 1999). In fact, in a concurrent-choice assessment, Argueta et al. (2019) found that their participant exclusively preferred working under the VR exchange-production schedule. To the extent that one arrangement is preferred and engenders more motivation, it might improve acquisition relative to the alternative. Such results are especially possible considering that authors typically recommend differentially reinforcing unprompted vs.
prompted acquisition responses (e.g., Karsten & Carr, 2013; Lerman et al., 2016). Thus, although overall rates of responding might be similar, learners likely earn fewer reinforcers under acquisition contexts than when engaging exclusively in mastered responses (Frank-Crawford et al., 2019). Under acquisition conditions, VR exchange-production schedules might be more preferred and might improve acquisition rates because the learner occasionally accesses backup reinforcers after just one or two unprompted correct response. As such, to the extent that individuals prefer VR-exchange-production schedules, they might acquire new skills more efficiently under VR-exchange-production schedules than under FR exchange-production schedules.

Another reason that FR and VR exchange-production schedules might differentially impact acquisition is related to tokens’ discriminative properties. Bullock and Hackenberg (2015) found that tokens have discriminative functions, with tokens delivered early on in an FR exchange-production schedule suppressing responding relative to tokens delivered toward the end of the response requirement. As such, tokens delivered early on in an FR exchange-production schedule might be less reinforcing than tokens delivered toward the end of the requirement due to their association with a greater delay and response requirement to back-up reinforcer delivery. Thus, acquisition responses reinforced with tokens early on in an FR exchange-production requirement might be subject to decreased motivation and might, therefore, be acquired more slowly than responses reinforced with later tokens. Clinically, such differences in reinforcer potency are problematic, because ideally acquisition programming should be optimal for all targets. One possibility for remediating differential token effectiveness is to eliminate the discriminative function of
tokens delivered at different points in the exchange-production schedule by using a VR rather than an FR exchange-production schedule. With the former, the temporal proximity and response requirement to backup reinforcers is unpredictable and indiscriminable upon any given token delivery. Therefore, VR exchange-production schedules might optimize the rate of acquisition across responses compared to FR schedules. As such, research comparing the effects of FR and VR exchange-production schedules on acquisition rates is needed.

The overall purpose of the following studies is to address the discussed gaps in applied research on establishing and evaluating tokens as conditioned reinforcers for learners with neurodevelopmental disorders and to advance our knowledge of best practices for token training. Study 1 is a systematic review to evaluate the overall and relative effectiveness of S-S pairing, R-S pairing, and ODT in establishing conditioned reinforcers. Study 2 extends the findings of Dozier et al. (2012) by evaluating and comparing the effectiveness of four procedures in establishing tokens as conditioned reinforcers for children with IDDs. Lastly, the purpose of Study 3 was to compare the rate of acquisition of tasks reinforced with tokens in a token economy with a fixed-ratio vs. variable-ratio exchange-production schedule.
The purpose of Study I was to assess the general and relative effectiveness of S-S pairing, R-S pairing, and ODT at establishing conditioned reinforcers for individuals with IDDs by systematically reviewing the relevant literature. The secondary purpose of Study I was to identify intervention components associated with differential levels of effectiveness at establishing conditioned reinforcers.

**Procedures**

**Eligibility Criteria**

The review included studies that used within-subject single-case designs (SCDs) that included comparable pre- and post-pairing measures. Thus, studies that used group designs or did not report comparable pre- and post-pairing measures were excluded. To be included, study participants could be of any age but had to be diagnosed with one or more IDDs, such as ASD, developmental delay, or Down syndrome. If some participants in a study met inclusion criteria while others did not, participants that met the criteria were included in the review while those that did not were excluded. We included published and unpublished studies available in English that evaluated S-S pairing, R-S pairing, and/or ODT to establish conditioned reinforcers. To be included, studies had to examine the effects of pairing on responses that a) were already in the learner’s repertoire to avoid conflating the effects of the intervention with acquisition processes and b) contingently produced the target stimulus or stimuli (i.e., stimulus or stimuli being established as conditioned reinforcers). These criteria included studies that measured the effects of pairing procedures via changes in stimulus ranking in stimulus preference assessments (SPAs). Studies were excluded if they evaluated
pairing a) using acquisition responses or b) to increase vocalizations emitted by participants by establishing those vocalizations as automatic reinforcers. Regarding the latter, such experiments usually pair experimenter-emitted vocalizations with existing reinforcers. When testing for the effects of pairing, experimenter vocalizations are not typically delivered contingent on responding. Ostensibly, increases in participant’s emission of the same sounds might indicate that the sound itself has become a conditioned reinforcer. However, one cannot be sure the sound itself has become a conditioned reinforcer without evaluating whether the experimenter-emitted sound maintains responding. Currently, many studies of this type do not test for pairing effects by contingently delivering experimenter-produced sounds for a target response. As such, we included studies that sought to establish vocalizations as conditioned reinforcers if those vocalizations were not produced by participants and their delivery was controlled by the experimenter (e.g., delivered via a recording or by the experimenter).

Search Methods and Identification of Studies

We searched PsycINFO (EBSCO) and ERIC (EBSCO). To maximize the likelihood of locating all relevant studies, search parameters did not include any filters (e.g., language or peer-reviewed sources) nor any date restrictions. The electronic search strategy for PsycINFO and ERIC, including the search terms, is provided in the Appendix.

Two researchers independently screened the titles and abstracts of all articles generated during electronic searches to exclude clearly irrelevant articles. Prior to title and abstract screening, the researchers completed a calibration exercise to ensure 90% inter-rater agreement. Next, the two researchers independently screened the full text of
articles designated as potentially relevant during the title and abstract screening to exclude articles that did not meet inclusion criteria. The researchers resolved disagreements during both screenings through consensus. If consensus had not been reached, a third researcher would have made a final decision.

**Data Extraction**

Prior to data extraction, two researchers completed a calibration exercise to ensure 90% interrater agreement. Then, researchers independently extracted data from all included studies and resolved discrepancies in the same manner described above.

Data were extracted on the following variables:

1. Study characteristics: Including, but not limited to, publication year, published vs. unpublished, risk of bias, and experimental design.
2. Population characteristics: Including, but not limited to, diagnoses, age, number of participants meeting inclusion criteria, and total number of participants.
3. Intervention characteristics: Including, but not limited to, pairing procedure(s) used, number of implementations for each procedure, target stimulus and existing reinforcer characteristics (e.g., social vs. non-social, primary vs. secondary), number of pairings, method used to identify existing reinforcer(s), number of target stimuli and existing reinforcers, target responses, and whether the target stimuli were tokens or not. When extracting data on the type of procedure used, we classified studies according to the description of the procedures and not by the authors' categorization of the procedure. As such, studies described as involving S-S pairing that had a response requirement (e.g., engaging with materials for a specified duration) for pairings to be completed were classified as R-S pairing.
4. Study results and effects: Including, but not limited to, dimension measured (e.g., rate), outcome measure used (e.g., SPA or reinforcer assessment), and assessment of functional relation based on visual analysis.

**Data Analysis**

The data extracted from included studies were analyzed using success estimates (SEs). An SE is the ratio of the total number of successful pairing implementations to the total number of attempted pairing implementations (Reichow & Volkmar, 2010). An
implementation was defined as any completed stand-alone SCD with comparable pre- and post-pairing measures. As such, each reversal design was considered a single implementation and, therefore, studies with multiple reversals had multiple implementations. Additionally, each multiple baseline design (MBD) across any number of baselines or tiers was defined as a single implementation given that a single tier from a multiple baseline is not sufficient to demonstrate a functional relation based on the experimental logic of the design. Moreover, Lanovaz and Turgeon (2020) found that relying on an effect shown in a single tier increases Type I errors (i.e., false positives) to unacceptable rates. However, if a study included more than one multiple baseline, each was counted as a separate implementation. We calculated success estimates for 1) implementations across all pairing procedures (i.e., an overall success estimate), 2) implementations of S-S pairing, 3) implementations of R-S pairing, and 4) implementations of ODT.

Successful implementations were determined via visual analysis to identify functional relations. A successful implementation was defined as any instance in which a) responding for the target stimulus during post-pairing measures was clearly elevated relative to pre-pairing levels and b) elevated post-pairing responding generally sustained across sessions and did not decrease back to pre-pairing levels. For multiple baseline designs and any variation thereof (e.g., nonconcurrent), only demonstrations in which at least two-thirds of the individual baselines displayed successful pairing as defined above were considered a successful implementation. The rationale for this definition is based on prior research demonstrating that three-tiered multiple baselines in which only two baselines show an effect produce acceptable rates of Type I errors.
and acceptable power levels (Lanovaz & Turegon, 2020). Accepting an effect shown in only one tier increases Type I errors to .22 and requiring all three tiers to demonstrate an effect might be overly stringent as it reduces power to unacceptable levels (i.e., .542). Additionally, for multiple baselines that included some participants who met the inclusionary criteria, while others did not, we considered the data for all participants, including those who were excluded (i.e., did not have IDDs), when making judgements about whether implementations were successful. For studies that conducted multiple, independent A-B (i.e., only measure pre- and post-pairing responding for each participant in isolation) or reversal (e.g., A-B-A-B) designs, each A-B and reversal implementation was visually analyzed independently. Two researchers independently concluded whether each implementation was successful and then resolved any disagreements via consensus. If consensus had not been obtained, a third researcher would have made the final decision.

We opted to use an SE instead of an effect size measure, because most existing effect sizes for SCDs lack validity and reliability (Wolery et al., 2010). However, SEs as calculated here may be susceptible to subjective bias from those conducting the visual analysis.

**Results**

Figure 2-1 shows the PRISMA flowchart that depicts the results of the search and screening process. The initial database search resulted in 718 articles for the title and abstract screening. Of those articles, 35 entered full-text review and 20 met inclusion criteria. The snowball search identified 74 articles for abstract and title screening. Of those articles, 39 entered full-text review, of which 11 met inclusion
criteria. In total, 31 articles were included in the review. Table 2-1 lists each of the included studies, 12 of which were unpublished dissertations or theses.

Table 2-1 also lists the SEs for each implemented procedure for each study as well as the total SEs for each procedure across studies. Eight studies evaluated S-S pairing, 20 evaluated R-S pairing, and nine evaluated ODT. The SE percentages for S-S pairing varied from 0-100% with an overall SE of 5 of 14 or 36%. The SEs for R-S pairing also varied from 0-100% with an overall SE of 33 of 53 or 61%. The SEs for ODT varied from 0-80% with an overall SE of 7 of 23 or 30%. Figure 2-2 depicts each of the aforementioned SEs for each procedure as well as a total SE across procedures. Overall, the SE was 45 of 90 or 50% when all procedures were combined.

Figure 2-3 depicts subgroup analyses for studies that utilized different categories of target stimuli and existing reinforcers. The top panel shows the SEs by target stimulus type. All included studies only targeted stimuli that, if established as conditioned reinforcers, would be classified as secondary reinforcing stimuli. As such, the SE for studies that targeted such stimuli is the same as the combined SE across all procedures and studies, 45 of 90 or 50%. Fourteen studies targeted social stimuli (e.g., praise, smiles, hugs), 18 targeted non-social stimuli (e.g., pictures, books, edible reinforcers), and one targeted both. The SE was 50, 51, and 67% for social, non-social, and both types of stimuli, respectively. Twenty-two studies targeted multiple stimuli for conditioning while 10 targeted a single stimulus. The SE was 31 and 80% for studies that used multiple stimuli vs. a single stimulus, respectively.

The bottom panel of Figure 2-3 depicts the SEs for studies that used different types of existing reinforcers. Fourteen studies used primary existing reinforcers (i.e.,
edible reinforcers), five used secondary existing reinforcers (e.g., tablets, tickles, praise), and 12 used both types of existing reinforcers. The SE was 45% for studies that used primary existing reinforcers, 63% for those that used secondary existing reinforcers, and 63% for those that used both. Three studies used social existing reinforcers, 18 used non-social existing reinforcers, and nine used both. The SE was 66% for studies that used social existing reinforcers, 43% for those that used non-social, and 79% for those that used both. Ten studies used a single existing reinforcer and 21 used two or more existing reinforcers. The SE was 36% for studies that used a single existing reinforcer and 57% for those that used multiple. Similarly, tables 2-2 and 2-3 show which studies used which types of target stimuli and existing reinforcers (e.g., primary vs. secondary vs. both), respectively.

**Discussion**

After extensive searching and screening, 31 studies met inclusion criteria. Based on data extraction, combined S-S pairing, R-S pairing, and ODT were effective at establishing conditioned reinforcers for individuals with IDDs only in half of implementations. Moreover, of the three reviewed procedures, R-S pairing was the most effective with an SE of 61%. In fact, R-S pairing was over 20 and 30% more effective than S-S pairing and ODT, respectively. These findings suggest that, out of the three procedures evaluated, R-S pairing should be the default pairing procedure when attempting to establish conditioned reinforcers.

Subgroup analyses revealed that all studies targeted only stimuli that would generally be classified as secondary conditioned reinforcers if pairing was effective. This finding is unsurprising given that primary reinforcers by definition do not need to be established as reinforcers. However, in some instances, specific stimuli (e.g., raisins)
that fall under the umbrella of primary reinforcers might not function as reinforcers for a given individual. Therefore, barring feeding disorders that require specialized intervention, it is possible that one might attempt to establish specific edibles as reinforcers, for example. In such circumstances, the effect of pairing procedures on establishing such stimuli as reinforcers is unknown. Of note, none of the included studies evaluated pairing procedures for establishing tokens as conditioned reinforcers. Instead, non-social target stimuli included but were not limited to pictures (e.g., Tyler-Santa et al., 2014) and leisure items such as books (e.g., Longano & Greer, 2006) and plush animals (e.g., Schnerch, 2014).

Moreover, it appears that targeting social or non-social stimuli exclusively does not differentially impact the effectiveness of pairing procedures. Social stimuli were defined as stimuli that cannot be consumed without intervention from another human (e.g., hugs, tickles). Non-social stimuli were defined as stimuli that can be consumed without intervention from another individual (e.g., edibles). These results suggest that, despite a putative preference for non-social over social rewards by children with ASD (Ruta et al., 2017), both types of stimuli may be established as conditioned reinforcers with similar effectiveness. In contrast, data from the subgroup analysis comparing studies that targeted one vs. multiple stimuli for conditioning revealed that studies that targeted a single stimulus were clearly more effective. In fact, the SE was over 50% higher for studies that attempted to establish only a single stimulus relative to those that targeted multiple stimuli. These findings are not surprising in light of previous basic research that indicates that a greater number of pairings leads to greater effects. As such, if multiple stimuli are targeted simultaneously for pairing, each stimulus is paired a
fewer number of times relative to if it was the only stimulus being paired. Thus, these data suggest that, to optimize the effectiveness of pairing procedures, only one stimulus should be targeted for conditioning at a time.

With respect to the types of existing reinforcers used in pairing procedures, subgroup analyses suggest that using a) secondary reinforcers only or b) a combination of primary and secondary reinforcers is over 20% more effective than using primary reinforcers only. However, these data must be interpreted cautiously as only six studies used only existing secondary reinforcers. Nonetheless, it is possible that using a combination of existing primary and secondary reinforcers during pairing procedures might produce better results. As such, further research is needed to clarify how the distribution of primary vs. secondary existing reinforcers impacts the effectiveness of R-S pairing, S-S pairing, and ODT.

Similarly, the analysis regarding the use of social vs. non-social existing reinforcers during pairing suggests that using a combination of social and non-social reinforcers is clearly more effective than using only non-social reinforcers. The data also suggest that using existing social reinforcers only is clearly more effective than using non-social reinforcers only. However, only three studies used social reinforcers exclusively and only nine used both social and non-social existing reinforcers. As such, these results should be interpreted cautiously.

Additionally, the subgroup analysis indicates that using multiple existing reinforcers is more effective than using only one. These results are interesting in light of previous token research. Although none of the studies included in this review evaluated tokens, prior researchers have found that tokens maintain higher rates of responding
when they have been paired with multiple vs. one backup reinforcer (Moher et al., 2008). In the context of tokens, these results have been explained in terms of motivating operations. Changes in satiation associated with a single stimulus are less likely to impact tokens’ potency if they are also exchangeable for other currently motivating stimuli. However, unlike tokens, most stimuli targeted for conditioning are not directly exchangeable for the stimuli with which they have been paired. However, it is possible that similar processes underly pairing procedures with other non-token stimuli. Perhaps, during the pairing process, stimuli that are paired with multiple reinforcers are less likely to be paired with abolishing operations (AOs) and, thus, less likely to be paired with temporarily neutral or aversive stimuli. As such, the stimulus targeted for conditioning are perhaps paired with greater reinforcement when paired with multiple existing reinforcers.

Relatedly, the aforementioned increased effectiveness of using a combination of a) social and non-social and b) primary and secondary existing reinforcers during pairing might be explained by similar processes. Satiation is less likely to occur across pairings if different categories of reinforcers are combined (e.g., edibles and leisure items) throughout the procedure. Therefore, in such arrangements, stimuli targeted for conditioning might be less likely to be paired with temporarily ineffective reinforcers which might reduce the effectiveness of pairing. This is especially pertinent considering that all included studies that used existing primary reinforcers relied on edibles.

While interpreting the results of this review, one must consider its limitations. First, we only evaluated three pairing procedures. It is possible that including other procedures would have identified more effective procedures and raise overall SEs.
Future researchers should include more pairing procedures in their reviews. Additionally, our review only included individuals with intellectual and developmental disabilities. Consequently, many studies including other populations for whom pairing may be warranted, such as those diagnosed with schizophrenia (e.g., Lovaas et al., 1966), were not included. Relatedly, research with neurotypical individuals was also excluded, although it might provide useful insights regarding the variables influencing effective pairing.

In summary, the results of this review highlight that our applied pairing technology is underdeveloped. Currently, R-S pairing appears to be the best option of the three procedures evaluated in this review. However, with a 61% SE, R-S pairing is far from the optimal pairing procedure. As such, further research designed to improve existing pairing procedures and to identify novel pairing procedures for individuals with IDDs is warranted. Increasing the number of available effective reinforcers for this population is often a critical treatment goal. Therefore, improving our methods for establishing conditioned reinforcers is essential. Nonetheless, based on the results of this review, current best practices for establishing conditioned reinforcers are to use R-S pairing, a single target stimulus at a time, and multiple existing reinforcers with which to pair the single target stimulus.
Figure 2-1. PRISMA flowchart for article screening.
Table 2-1. List of included studies with success estimates (SEs) per study and procedure. U indicates the study is an unpublished dissertation or thesis.

<table>
<thead>
<tr>
<th>Study</th>
<th>S-S pairing</th>
<th>R-S pairing</th>
<th>ODT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SE</td>
<td>SE %</td>
<td>SE</td>
</tr>
<tr>
<td>Ardoin et al. (2004)</td>
<td>-</td>
<td>-</td>
<td>2 of 4</td>
</tr>
<tr>
<td>Axe &amp; LaPrime (2017)</td>
<td>-</td>
<td>-</td>
<td>1 of 3</td>
</tr>
<tr>
<td>Beus (2014)U</td>
<td>-</td>
<td>-</td>
<td>1 of 1</td>
</tr>
<tr>
<td>Birnbauer (1971)</td>
<td>-</td>
<td>-</td>
<td>2 of 2</td>
</tr>
<tr>
<td>Buttigieg (2015)U</td>
<td>-</td>
<td>-</td>
<td>4 of 4</td>
</tr>
<tr>
<td>Chance et al. (2021)</td>
<td>-</td>
<td>-</td>
<td>2 of 3</td>
</tr>
<tr>
<td>Dozier et al. (2012)</td>
<td>0 of 4</td>
<td>0</td>
<td>4 of 8</td>
</tr>
<tr>
<td>Dudley et al. (2019)</td>
<td>-</td>
<td>-</td>
<td>2 of 3</td>
</tr>
<tr>
<td>Freeman (2005)U</td>
<td>1 of 1</td>
<td>100</td>
<td>-</td>
</tr>
<tr>
<td>Greer et al. (2011)</td>
<td>-</td>
<td>-</td>
<td>1 of 1</td>
</tr>
<tr>
<td>Greer &amp; Han (2015)</td>
<td>-</td>
<td>-</td>
<td>1 of 1</td>
</tr>
<tr>
<td>Hanley et al. (2003)</td>
<td>1 of 1</td>
<td>100</td>
<td>-</td>
</tr>
<tr>
<td>Hanney (2015)U</td>
<td>1 of 2</td>
<td>50</td>
<td>1 of 2</td>
</tr>
<tr>
<td>Jerome &amp; Sturmey (2014)</td>
<td>1 of 1</td>
<td>100</td>
<td>-</td>
</tr>
<tr>
<td>Kohler (2014)U</td>
<td>-</td>
<td>-</td>
<td>2 of 2</td>
</tr>
<tr>
<td>Longano (2008)U</td>
<td>-</td>
<td>0 of 1</td>
<td>0</td>
</tr>
<tr>
<td>Longano &amp; Greer (2006)</td>
<td>-</td>
<td>-</td>
<td>1 of 1</td>
</tr>
<tr>
<td>Lugo et al. (2017)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Maffei-Lewis et al. (2014)</td>
<td>-</td>
<td>-</td>
<td>1 of 1</td>
</tr>
<tr>
<td>Moore (2017)U</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Nuzzolo-Gomez et al. (2002)</td>
<td>-</td>
<td>-</td>
<td>2 of 2</td>
</tr>
<tr>
<td>Pastrana (2018)U</td>
<td>-</td>
<td>-</td>
<td>4 of 5</td>
</tr>
<tr>
<td>Pereira Delgado et al. (2009)</td>
<td>-</td>
<td>-</td>
<td>1 of 1</td>
</tr>
<tr>
<td>Petursdottir et al. (2011)</td>
<td>0 of 2</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>Reinecke (2005)U</td>
<td>0 of 1</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>Rodriguez and Gutierrez (2017)</td>
<td>-</td>
<td>-</td>
<td>1 of 2</td>
</tr>
<tr>
<td>Schenk (2016)U</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Schnerch (2014)U</td>
<td>1 of 2</td>
<td>50</td>
<td>-</td>
</tr>
<tr>
<td>Steinman (1968)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Taylor-Santa et al. (2014)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Wathen (2017)U</td>
<td>-</td>
<td>-</td>
<td>0 of 6</td>
</tr>
<tr>
<td>Total</td>
<td>5 of 14</td>
<td>36</td>
<td>33 of 53</td>
</tr>
</tbody>
</table>

31
Figure 2-2. Success estimates in percentages for each procedure (left panel) and the frequency of successes to implementations per procedure (right panel).
Figure 2-3. Subgroup analyses for the type and number of target stimuli (upper panel) and existing reinforcer(s) (lower panel). The number of studies included in each category are denoted above each bar.
Table 2-2. Types of target stimuli by study. $^U$ indicates studies that were unpublished dissertations or theses.

<table>
<thead>
<tr>
<th>Study</th>
<th>Primary</th>
<th>Secondary</th>
<th>Social</th>
<th>Non-social</th>
<th>Single</th>
<th>Multiple</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ardoin et al. (2004)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Axe &amp; LaPrime (2017)</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Beus (2014)$^U$</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Birnbrauer (1971)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Buttigieg (2015)$^U$</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Chance et al. (2021)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Dozier et al. (2012)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Dudley et al. (2019)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Freeman (2005)$^U$</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Greer et al. (2011)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Greer &amp; Han (2015)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Hanley et al. (2003)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Hanney (2015)$^U$</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Jerome &amp; Sturmey (2014)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Kohler (2014)$^U$</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Longano (2008)$^U$</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Longano &amp; Greer (2006)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Lugo et al. (2017)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Maffei-Lewis et al. (2014)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Moore (2017)$^U$</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Nuzzolo-Gomez et al. (2002)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Pastrana (2018)$^U$</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Pereira Delgado et al. (2009)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Petursdottir et al. (2011)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Reinecke (2005)$^U$</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Rodriguez and Gutierrez (2017)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Schenk (2016)$^U$</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Schnarch (2014)$^U$</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Steinman (1968)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Taylor-Santa et al. (2014)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Wathen (2017)$^U$</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>31</td>
<td>14</td>
<td>19</td>
<td>23</td>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>
Table 2-3. Types of existing reinforcers by study. U indicates studies that were unpublished dissertations or theses.

<table>
<thead>
<tr>
<th>Study</th>
<th>Primary</th>
<th>Secondary</th>
<th>Social</th>
<th>Non-social</th>
<th>Single</th>
<th>Multiple</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ardoin et al. (2004)</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Axe &amp; LaPrime (2017)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Beus (2014) U</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Birnbrauer (1971)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Buttigieg (2015) U</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chance et al. (2021)</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dozier et al. (2012)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dudley et al. (2019)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Freeman (2005) U</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greer et al. (2011)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greer &amp; Han (2015)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Hanley et al. (2003)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hanney (2015) U</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jerome &amp; Sturmey (2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kohler (2014) U</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Longano &amp; Greer (2006)</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lugo et al. (2017)</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maffei-Lewis et al. (2014)</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moore (2017) U</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nuzzolo-Gomez et al. (2002)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pastrana (2018) U</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pereira Delgado et al. (2009)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Petursdottir et al. (2011)</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reinecke (2005) U</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rodriguez and Gutierrez (2017)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schenk (2016) U</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schnerch (2014) U</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Steinman (1968)</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taylor-Santa et al. (2014)</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wathen (2017) U</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>26</td>
<td>16</td>
<td>12</td>
<td>27</td>
<td>12</td>
<td>21</td>
</tr>
</tbody>
</table>
CHAPTER 3
STUDY 2

Study 1 highlighted that there is a lack of research evaluating pairing procedures for establishing tokens as conditioned reinforcers. Thus, the purpose of Study 2 was to evaluate and compare the effectiveness of 1) S-S pairing, 2) R-S pairing, 3) stimulus-exchange-stimulus (S-E-S) pairing as suggested by Kazdin (1977), and 4) R-S plus exchange (R-S-E) pairing similar to Argueta et al.’s (2019) methods at establishing tokens as conditioned reinforcers for children with ASD and/or IDDs.

Method

Participants, Setting, and Materials

Participants included four children diagnosed with intellectual and developmental disabilities (IDDs) who attend a behavioral services clinic and did not have any known experience with tokens. Additionally, we only included participants who did not currently exhibit any severe problem behavior. Doug was an eight-year-old White non-Hispanic male diagnosed with ASD and Down Syndrome. Doug was able to follow one- to two-step instructions, mand using two to three words, and engage in simple tacts. Arnold was a three-year-old Multiracial non-Hispanic male diagnosed with ASD who could follow one-step instructions and communicated with gestures. Desmond was a 6-year-old White non-Hispanic male diagnosed with ASD and a glutamate ionotropic receptor NDMA type subunit 2B gene mutation. Desmond was able to follow one and two-step instructions, mand using one to two words, tact using one word, and select stimuli form an array of three stimuli. Miguel was a 3-year-old White non-Hispanic male diagnosed with ASD who could follow two- to three-step instructions, manded using three- to five-
word phrases (e.g., “I want food, please”), emitted multi-word tacts (e.g., “blue toy”), and could select stimuli from an array of four stimuli.

Sessions took place in the clinic’s therapy rooms containing a table and chairs. To increase discriminability, tokens were topographically distinct (e.g., round plastic coins vs. foam squares) for each procedure for all participants. Tasks materials and reinforcers were individualized for each participant. Doug’s task materials included a stack of laminated sheets with pre-printed circular outlines and dry erase markers. Arnold’s task materials were a wooden peg attached to a base, a wooden ring, and Legos™. Desmond’s task materials were crayons and a pencil bag. Miguel’s task materials included a stack of flashcards with five letters on each card.

**Response Measurement**

During the preference assessment, the experimenter collected paper and pencil data on item selections, defined as touching a presented stimulus. During reinforcer assessments, the experimenter used the Countee app© (Gavran & Hernandez, 2020) to collect live data on task completion, defined on an individual basis for each participant. During the verbal screener, data were collected on correct responses defined as independently and accurately describing the consequences for engaging in target responses and exchanging tokens in the token system. We also collected data on incorrect responses defined as incorrectly describing the consequences for target responses and exchanging tokens or emitting no response.

During the experiment proper, we again collected data using the Countee app© (Gavran & Hernandez, 2020). The primary dependent variable was task completion, generally defined as independently and correctly engaging in the target response. Doug’s target response was tracing which was defined as drawing three circles with a
marker on a single laminated sheet by following and staying within each of three pre-printed circular outlines. If any portion of any of the three circles fell outside the pre-printed outline, the response was considered incorrect. Arnold’s first target response was stacking a ring by using his hands and fingers to slide a wooden ring by an opening at its center onto a wooden peg such that the ring touched the base to which the peg was attached. Arnold’s second target response was stacking Legos™ which was defined as using the hands and fingers to place one Lego™ directly on top of another such that it interlocked with the bottom Lego™. Desmond’s target response was storing crayons which was defined as using the hands and fingers to insert one or more crayons past the opening of a pencil bag. The crayon(s) had to be a different crayon(s) than the one(s) inserted in the previous response. Miguel’s target response was reading letters which was defined as correctly stating each of five letters on a flashcard from left to right without repeating any letters except the first. If Miguel incorrectly read any of the letters on the flashcard, the response was not counted as meeting the operational definition. Researchers also collected data on token exchanges, defined as releasing a single token into the experimenter’s hand.

Pre-Experimental Procedures

Preference and reinforcer assessments

All participants except for Doug completed a paired-stimulus preference assessment (PSPA; Fisher et al., 1992), containing eight stimuli identified via informal caregiver and therapist interviews (RAISD; Fisher et al., 1996). Prior to PSPAs, participants sampled leisure items for 30-s or a small piece of each edible. During the PSPA, the experimenter presented stimuli in pairs ensuring that each stimulus was presented once with all other stimuli. The experimenter blocked attempts to select both
stimuli and represented the trial. Participants completed the PSPA twice and participants for whom we evaluated both edible and leisure items completed separate PSPAs for edibles and leisure items (i.e., four total). The three stimuli with the highest selection percentages across both PSPAs were included in a reinforcer assessment. For Doug, two single-stimulus engagement (SSE) preference assessments (DeLeon et al., 1999) were completed after he refused to make choices during multiple attempts at completing a PSPA. In the SSE preference assessment, we placed each stimulus one at a time in-sight and in-reach of Doug for 2 min and recorded the duration of engagement (i.e., physical contact) throughout the 2 min. We selected the three stimuli with the highest engagement across both SSE assessments. Doug’s most preferred items were the iPad, computer, and car toys. Arnold’s most preferred items were goldfish, chips, and gummies. Desmond’s most preferred items were a noise-producing pig toy, a globe light that projected stars and planets in various colors (i.e., space light), and a toy with multiple colorful lights that spun around quickly (i.e., spinning light). Miguel’s most preferred items were the same spinning light Desmond preferred, a wooden guiro which is a musical instrument shaped and painted like a fish, and a clear plastic cylinder filled with plastic beads that moved around when the cylinder was moved (i.e., bead toy).

To ensure that stimuli identified via preference assessments functioned as reinforcers, the experimenter conducted a reinforcer assessment in a multielement design consisting of one baseline (i.e., control) condition and one reinforcement condition for each of the top three stimuli identified in the preference assessments. During baseline, task completion did not result in any programmed consequences.
During reinforcement sessions, task completion resulted in the immediate delivery of a bite-sized edible or 15-s access to a leisure item. All sessions were 5 minutes. If a reinforcer assessment indicated that a preferred item did not function as a reinforcer, we evaluated other stimuli from the preference assessment or suggested by caregivers or direct observation. Doug’s reinforcer assessment indicated that the iPad was a reinforcer; however, responding was not maintained by the car and was maintained at low levels by the computer. Based on caregiver nomination, we evaluated pudding as a reinforcer and selected this item along with the iPad for use in the remainder of the experiment. Arnold’s reinforcer assessment confirmed that goldfish, chips, and gummies functioned as reinforcers. Desmond’s results confirmed that the spinning light and pig functioned as reinforcers. The space light was not evaluated because Desmond engaged with the stimulus in a manner that was potentially dangerous. Miguel’s reinforcer assessment revealed that only the light functioned as a reinforcer. Based on caregiver nomination, we then evaluated slime which the reinforcer assessment confirmed as a reinforcer. As such, we used the light and slime when we began the experiment. However, upon beginning the experiment, Miguel began stating that he did not want the light or slime during pairing sessions. Thus, we began the experiment again and conducted an abbreviated eight-item multiple stimulus without replacement preference assessment (MSWO; DeLeon & Iwata, 1996) before each pairing session. During these assessments, we presented all eight stimuli in-sight and in-reach of the participant. We then instructed Miguel to “pick one.” When Miguel selected an item, we provided access for 30 s. Then, we represented the remaining seven items in the same manner and also delivered access to the selected stimulus for 30 s. Because we only
used two stimuli per session, the assessment ended after Miguel’s second selection. All pre-session MSWOs included the same eight stimuli which were a fan, book, small squishy toys, music cube, water snake, coloring, as well as the light and slime from previous assessment. We used the first two stimuli selected during the pre-session MSWO during the subsequent pairing session.

**Verbal screener**

Experimenters completed a verbal screener to identify and exclude learners for whom verbal descriptions were likely sufficient to establish a token system. The experimenter administered the screener after the reinforcer assessment was complete to ensure that the stimuli included in the verbal description of the token economy functioned as reinforcers. During the screener, the experimenter described the contingencies for earning and exchanging tokens. Then, the experimenter asked the participant, “What do you get for [insert task]?” and “What can you exchange a token for?” Each participant completed the screener twice. If a participant had answered both questions correctly on both screeners they would have been excluded; however, no participants met exclusion criteria.

**General Procedure**

All sessions were 10 minutes. Task materials were present during all conditions except S-S and S-E-S pairing sessions. Prior to sessions during which task materials were available, the experimenter conducted three pre-session exposure trials using least-to-most (LTM) prompting to ensure participants contacted the upcoming contingencies for task completion.

Following baseline and an initial token test, all participants experienced S-S pairing first. Then, the experimenter randomly assigned participants to either R-S or S-
E-S pairing. Random assignment was used to control for potential sequence effects. Consequently, Doug and Arnold experienced R-S followed by S-E-S pairing whereas Desmond and Miguel experienced S-E-S pairing first and then R-S pairing. Next, participants experienced the other procedure to which they were not initially randomly assigned (e.g., S-E-S pairing). Last, participants were exposed to R-S-E pairing. If R-S-E pairing was not effective as indicated by the token test, participants began the thinned token-exchange test. However, if a given procedure had been effective as indicated by the token test condition, the evaluation would have been terminated as tokens would have been successfully established as conditioned reinforcers.

For participants whose reinforcers included leisure items, the experimenter provided access to each item for 15 s. The session clock was paused during reinforcer intervals for leisure items but continued to run when edible reinforcers were delivered.

**Baseline**

During baseline, the experimenter began sessions by placing task materials in view and in reach of the participant and informing participants that they could but did not have to work. Thereafter, the experimenter did not deliver any prompts for task completion. Instances of task completion resulted in no programmed consequences. For Desmond only, we provided noncontingent attention on a fixed-time (FT) 30-s schedule (i.e., every 30 seconds). When the experimenter first attempted to conduct baseline, Desmond was having toileting accidents during sessions. His clinical team reported that he engaged in such behaviors for attention and expressed concerns that our experimental sessions were negatively impacting his toilet training. As such, we started baseline over with the aforementioned modification in place.
**Token baseline**

The token baseline was identical to baseline except that task completion resulted in immediate token delivery on a fixed-ratio (FR) 1 schedule. The experimenter blocked attempts to exchange or otherwise engage (e.g., play) with the tokens. A token baseline phase immediately preceded each token-pairing procedure to ensure that the stimuli selected as tokens (e.g., plastic coins) for each procedure did not already function as conditioned reinforcers prior to pairing.

**Token test**

Token test was identical to token baseline and functioned to assess whether tokens were established as conditioned reinforcers following each pairing procedure. However, one session was conducted after every five S-S pairing and S-E-S pairing sessions. A series of sessions were conducted following completion of the R-S and R-S-E pairing conditions. To remain consistent with Dozier et al. (2012), R-S token test sessions alternated with baseline sessions Token test phases ended after at least three consecutive sessions with stable responding.

**S-S pairing**

During S-S pairing sessions, task materials were not present. On a FT 15-s schedule (i.e., every 15 s), the experimenter delivered a token immediately followed by one of the two to three identified reinforcers rotated quasi-randomly across deliveries.

**R-S pairing**

R-S pairing sessions began with the experimenter presenting task materials and instructions exactly as in the baseline condition. However, contingent on task completion, participants received a token immediately followed by one of the two or three quasi-randomly rotated reinforcers on a FR 1 schedule.
S-E-S pairing

S-E-S pairing sessions were similar to S-S pairing sessions in that participants received a token on a FT 15-s schedule. However, immediately following token delivery, the experimenter prompted the participant to exchange the token using LTM prompting. Regardless of prompting level, token exchanges immediately resulted in the delivery of one of three semi-randomly rotated reinforcers.

R-S-E pairing

R-S-E pairing sessions were similar to R-S pairing sessions in that participants earned tokens on an FR 1 schedule contingent on task completion. However, immediately following token deliveries, the experimenter prompted the participant to exchange the token using LTM prompting. Token exchanges immediately resulted in the delivery of one of three quasi-randomly rotated reinforcers regardless of prompting level.

Thinned token-exchange test

The purpose of this condition was to determine if the token test condition might produce false negatives. If tokens are only valuable insofar as they are predictive of forthcoming reinforcement (Shahan, 2010), low levels of responding during post-pairing token test conditions might reflect participants' sensitivity to discriminable changes in the exchangeability (i.e., predictiveness) of tokens. Thus, post-pairing token test conditions might function as an extinction condition that prevents detection of any value imparted by pairing procedures. We conducted the thinned token-exchange test only following only R-S-E pairing after we observed low levels of responding during token tests for all procedures for the first two participants who completed R-S-E pairing.
Throughout the thinned token-exchange test conditions, participants earned tokens for task completion on an FR 1 schedule. Following each token delivery, the experimenter immediately prompted participants to exchange the token using LTM prompting. Contingent on token exchanges, the experimenter delivered a reinforcer on a variable-ratio (VR) schedule. No pre-session exposure trials occurred prior to thinned token-exchange test conditions.

**Modified VR 3.** During modified VR 3 sessions, token exchanges resulted in one of three semi-randomly rotated reinforcers on a VR 3 schedule. However, the first token exchange in each session always resulted in a reinforcer delivery.

**Modified VR 6.** Modified VR 6 sessions were identical to modified VR 3 sessions, except that token exchanges produced reinforcers on a VR 6 schedule.

**Modified VR 6 with hidden reinforcers.** Sessions during the modified VR 6 with hidden reinforcers condition were identical to modified VR 6 sessions except that the experimenter hid reinforcers out of participants’ view throughout the session. Participants only saw reinforcers when they were delivered. The purpose of this condition was to determine if the visual presence of the reinforcers functioned as an SD that signaled their forthcoming delivery and, thus, evoked responding.

**True VR 6.** True VR 6 sessions were identical to sessions during VR 6 with hidden reinforcers condition except for that the first token exchange was no longer programmed to produce a reinforcer. The purpose of this condition was to determine whether the consequence (i.e., reinforcer delivery) for the first token exchange functioned as an SD for reinforcer availability and thereby evoked responding.
Results

Figure 3-1 depicts the results for Doug, Arnold, and Desmond. Note that the figure does not depict pairing sessions but that the number of pairings conducted during each procedure are written vertically just to the right of each token test phase line. Doug's data appear in the top panel of Figure 3-1. During baseline, S-S token baseline, and S-S token test, Doug engaged in zero responses per minutes (RPM). During R-S token baseline, Doug also engaged in zero RPM. In the following phase, Doug's response rate was zero during the R-S token test condition except for the first session during which his rate was .5 RPM. During the baseline condition that alternated with the R-S token test condition, responding remained at zero RPM throughout all sessions. Throughout the S-E-S token baseline and S-E-S token test phases, Doug engaged in a mean of .12 and 0 RPM, respectively. He again engaged in zero responses during both R-S-E baseline and the R-S-E token test. Doug's rate of responding immediately increased during the modified VR 3 condition. However, after the third session in that condition, responding began trending downward. Anecdotally, the experimenter observed that Doug was attempting the target response consistently throughout sessions. However, he was engaging in errors and, thus, his responses did not meet the operational definition for the target response. As such, we began recording incorrect responses during session 41 during which the rate of incorrect responses was similar to the rate of target responses throughout the first three sessions of the modified VR 3 phase. During session 42, incorrect responses decreased while correct responses increased to previous levels. The overall mean for the target response during the modified VR 3 phase was 2.9 RPM. During the following modified VR 6 phase, tracing and incorrect responses averaged 4 and 1 RPM, respectively. Throughout the modified
VR 6 with hidden reinforcers phase, correct tracing and incorrect responses averaged 3.58 and .70 RPM across sessions, respectively. Finally, during the true VR 6 condition, Doug engaged in 3.7 and .60 correct and incorrect tracing RPM, respectively.

Arnold’s results appear in the middle panel of Figure 3-1. During baseline, responding started out high but gradually diminished across sessions. He generally engaged in low levels of responding with an average of .58 RPM across sessions. Throughout the S-S token baseline phase, Arnold’s rate of responding averaged .10 RPM. Following 600 individual S-S pairings, he engaged in an average of .17 RPM during S-S token test sessions. Due to experimenter error, Arnold did not experience an R-S token baseline, but a baseline control condition was added to the R-S token test phase. Following R-S pairing, he engaged in a mean of .06 RPM and .26 RPM during the R-S token test and baseline conditions, respectively. Throughout the S-E-S baseline, Arnold engaged in zero responses during all sessions. In the S-E-S token test phase, he engaged in a mean of .1 RPM. Arnold engaged in no responses during the R-S-E baseline. During R-S-E pairing sessions, Arnold began engaging in stereotypy with the ring that he placed on the peg, resulting in the target response decreasing to zero levels across three consecutive sessions. Thus, we identified and collected baseline for a new target response (i.e., stacking Legos™), depicted in the phase labeled New Task. Following the new task phase, we continued R-S-E pairing and the remainder of the experiment using the new target response. During the subsequent R-S-E token test phase, Arnold’s rate of responding initially increased above R-S-E baseline levels. However, responding continued in a decreasing trend before leveling out at zero or near-zero levels during the last four sessions. Overall, Arnold engaged in a mean of .56
RPM across sessions. During the modified VR 3 condition, Arnold’s rate of responding immediately increased beyond all previously observed levels of responding with a mean of 5.57 RPM. In the following modified VR 6 phase, responding remained elevated at an average of 4.55 RPM across sessions. In the modified VR 6 with hidden reinforcers phase, Arnold’s rate of responding continued at similar rates with a mean of 5.6 RPM. In the final true VR 6 phase, Arnold’s RPM decreased relative to the previous phase but was higher than the initial R-S-E baseline at an average of 3.63 RPM.

Desmond’s results appear in the bottom panel of Figure 3-1. During baseline, his rate of responding was zero across all sessions. During the S-S baseline, he engaged in an average of .07 RPM. In the subsequent S-S token test phase, Desmond also engaged in a mean of .07 RPM. During S-E-S baseline and the S-E-S token test, Desmond engaged in .07 and .1 RPM, respectively. The rate during the R-S baseline was .05 RPM. In the next phase, Desmond engaged in a mean of .2 RPM during both the R-S token test and baseline conditions. During R-S-E baseline, Desmond engaged in .56 RPM across sessions with an outlier of 1.8 RPM during the first R-S-E baseline session. During the R-S-E test, Desmond engaged in lower levels of responding than in R-S-E baseline with a mean rate of .07 RPM across sessions. During the modified VR 3 condition, Desmond’s rate of responding immediately increased with a mean of 3.03 RPM. During the subsequent modified VR 6 condition, his responding decreased to a mean of .95 RPM. In the modified VR 6 with hidden reinforcers condition, Desmond’s rate of responding was similar to the previous condition with a mean of 1.03 RPM. During the true VR 6 condition, his response rate had a mean of 0.8 RPM.
Miguel’s results appear in Figure 3-2. During baseline, responding initially occurred on a downward trend before leveling out at zero RPM during the last three sessions. Mean responding during baseline was .68 RPM. During S-S token baseline and S-S token test, Miguel engaged in zero responses across all sessions. He also engaged in zero responses across all sessions of both the S-E-S token baseline and S-E-S token test phases. During the R-S baseline, he engaged in .05 RPM. In the subsequent phase during the R-S token test conditions, Miguel’s rate of responding increased above that observed in all previous sessions and phases at 3.2 and 3 RPM during the first and second sessions, respectively. Thereafter, responding in the R-S token test condition decreased and remained at or below .8 RPM similar to the rates observed in the baseline condition. During the latter condition, Miguel engaged in an average of .26 RPM across sessions. Miguel did not complete the experiment after he did not assent to participant across three consecutive days.

**Discussion**

As evidenced by token test results, S-S, R-S, S-E-S, and R-S-E pairing generally were not effective at establishing tokens as conditioned reinforcers as they are traditionally conceptualized (i.e., as stimuli that acquire the ability to independently support behavior because of their prior association with existing reinforcers; Shahan, 2010). Except for Miguel’s R-S token test, all participants’ token test results indicated that none of the evaluated pairing procedures were effective, but even Miguel’s effects were weak. However, under the thinned token-exchange test following R-S-E pairing and the initial R-S-E token test, Doug, Arnold, and Desmond demonstrated an immediate increase in responding. This level of responding maintained across all phases of the thinned token-exchange test (e.g., modified VR 6) with Doug and Arnold.
Desmond’s rate of responding decreased during the modified VR 6 condition and did not return to VR 3 levels during subsequent conditions. However, his mean RPM was generally above R-S-E baseline and token test levels in these conditions. Thus, this study found that, following pairing, tokens supported responding only when they remained at least intermittently exchangeable for backup reinforcers and that they did not support responding under extinction (i.e., token test) conditions.

The results for these participants suggested that tokens might not function as conditioned reinforcers in that pairing does not impart independent value or reinforcing efficacy upon tokens in a manner in which they could support behavior in and of themselves, as would have been indicated by continued responding during token tests. Instead, their results are more consistent with the notion that tokens might function more like discriminative stimuli in that they evoke behavior only when their delivery is predictive of the delivery of other existing reinforcers. This interpretation is more consistent with information theory which posits that conditioned reinforcers, like tokens, do not acquire response strengthening properties and, instead, support behavior by signaling forthcoming reinforcement (Shahan, 2010; Shahan & Cunningham, 2015). As such, it is possible that the three pre-session exposure trials that preceded all token test sessions were sufficient for participants to discriminate that tokens no longer signaled forthcoming backup reinforcers, thus, suppressing responding. When tokens became intermittently exchangeable during the thinned token-exchange test, they presumably regained their signaling function and, thus, increased and maintained responding in a manner that might otherwise be interpreted as a conditioned reinforcement effect. In sum, these results suggest that tokens quickly lose their “value” when they are no
longer exchangeable. This is perhaps not surprising, but it is inconsistent with the commonly held assumption that neutral stimuli paired with established primary or conditioned reinforcers assume some independent reinforcing value themselves (Williams, 1994).

However, as aforementioned, Miguel's R-S token test results demonstrated that R-S pairing resulted in tokens temporarily supporting behavior. His results may be interpreted as evidence that R-S pairing resulted in a transient conditioned reinforcement effect. An alternative explanation is that the temporary increase in responding might be evidence of an extinction process in which tokens gradually lost their discriminative function when their delivery was no longer correlated with backup reinforcers. Therefore, Miguel's results do not invalidate the possibility that tokens might serve more as discriminative stimuli than conditioned reinforcers. In fact, participants in prior studies for whom R-S pairing was effective demonstrated sustained reinforcing effects unlike the transient effect observed with Miguel (Dozier et al., 2012; Rodriguez & Gutierrez, 2017). The difference in the maintenance of the response-supporting effect might be indicative of tokens functioning more or primarily as discriminative rather than reinforcing stimuli. Nonetheless, it is possible that tokens might have both discriminative and reinforcing functions. However, more research is needed to further elucidate the function(s) of tokens.

Overall, the results of this study are consistent with previous research that showed that S-S pairing is not effective at establishing conditioned reinforcers (e.g., Dozier et al., 2012) when tested under extinction conditions. Moreover, Miguel's results are somewhat consistent with prior research that demonstrated that R-S pairing can
effectively establish conditioned reinforcers (e.g., Rodriguez & Gutierrez, 2017) in some but not all cases. However, as noted, the transient nature of the effect observed for Miguel is not characteristic of R-S pairing in prior research.

The results of this study must be interpreted in light of several limitations. First, we did not conduct a thinned token-exchange test following any pairing procedure prior to R-S-E pairing and, therefore, cannot conclude that tokens did not support behavior under similar circumstances prior to R-S-E pairing. Therefore, one cannot make conclusions about whether these procedures would have resulted in tokens supporting responding when they were intermittently exchangeable. Additionally, it is possible that the effects observed in the thinned token-exchange test following R-S-E might have been a sequence effect produced by the recent history with the previous four pairing procedures. As such, future research should evaluate the effects of S-S, R-S, and S-E-S pairing on tokens under thinned token-exchange test conditions. It is possible that, if tokens are primarily discriminative in function, any effects of these pairing procedures may only be detectable under such circumstances. Researchers should also consider comparing effects under thinned token-exchange test conditions to extinction test conditions to further inform whether thinned token-exchange test conditions are more sensitive than extinction test conditions. Relatedly, future researchers should design studies to identify the function(s) of tokens, which might allow for the design of more effective and efficient token pairing and testing procedures.

In sum, the results of this study indicated that R-S and R-S-E pairing might result in tokens supporting behavior. However, it is possible that other pairing procedures, including but not limited to S-S and S-E-S pairing, might also result in tokens supporting
behavior if effects are tested under more sensitive circumstances. Ultimately, the results of this study highlight that, as a field, we do not yet understand token pairing procedures nor the function of tokens for individuals with IDDs very well.
Figure 3-1. Results for S-S, R-S, S-E-S, and R-S-E pairing for Doug (top panel), Arnold (middle panel), and Desmond (bottom panel).
Figure 3-2. Results for S-S, R-S, S-E-S, and R-S-E pairing for Miguel.
CHAPTER 4
STUDY 3

The purpose of Study 3 was to compare the rate of acquisition of tasks reinforced with tokens in a token economy with a FR vs. VR exchange-production schedule.

Method

Participants, Setting, and Materials

Participants included two children diagnosed with ASD who attend a behavioral services clinic. Ignacio was a 6-year-old White non-Hispanic male who could follow one-step instructions, emit one- to two-word mands and tacts, and selected sight words from an array of three. Zola was a 4-year-old White non-Hispanic female who could follow one-step instructions, engage in one-word mands and tacts, and engaged in listener responses requiring the selection of a stimulus from an array of three. Sessions occurred in the clinic’s therapy or conference rooms containing a table and chairs. Task materials were individualized for each participant and consisted of materials for ten acquisition tasks. Ignacio’s task materials were laminated sight-word flashcards and Zola’s were laminated single uppercase letter flashcards.

Each participant’s materials also included two laminated token boards and tokens in the form of laminated shapes. One token board had five separate pieces of Velcro (one for each token) affixed to it for the fixed-ratio (FR) condition and the second board had one long strip of Velcro running across it for the variable-ratio (VR) condition. Additionally, each board had different color border around the edge and the tokens associated with each condition were the same color as the border. Ignacio’s FR and VR token boards and tokens were red and yellow, and the tokens were in the shape of crosses and diamonds, respectively. Zola’s FR and VR token boards and tokens were
black and orange, and the shape of her tokens were diamonds and hearts, respectively. Additionally, during each condition, we covered the table the participant and experimenter sat at with plastic tablecloths in the color corresponding to each condition when it was conducted. Backup reinforcers varied across participants according to preference and reinforcer assessment results (below).

**Response Measurement**

During preference assessments, the experimenter collected paper and pencil data on item selections, defined as touching a presented stimulus. During reinforcer assessments, the experimenter used the Countee app to collect live data on task completion, defined on an individual basis for each participant. During token training, the experimenter collected data on correct exchange responses, defined as independently releasing tokens into the experimenter’s hand. The experimenter also collected data on prompted exchange responses, defined as a correct exchange response following a gestural, model, or full physical prompt. We also collected data on correct production responses, defined as independent completion of the target response and prompted production responses, defined as correct production response follow a gestural, model, or fully physical prompt. During task probes, we collected paper and pencil data on correct responses, defined as independent completion of the target response. We also collected data on incorrect responses, defined as engaging in a response other than the pre-determined target response or emitting no response. During the experiment proper, we collected data on correct responses and incorrect responses defined in the same manner as in the task probes. We also collected data on prompted responses, defined as correct responses that occurred following a gestural, model, or physical prompt.
Pre-Experimental Procedures

Preference and reinforcer assessments

To identify backup reinforcers for which tokens were exchanged, participants completed a preference and reinforcer assessment. First, participants completed an eight-item PSPA (Fisher et al., 1992) conducted in the same manner as in Study 2. The assessment was conducted twice using the same stimuli and the two items with the highest selection percentages across the two PSPAs were included in a reinforcer assessment, also conducted in the same manner as Study 2, to ensure the stimuli selected from the PSPA functioned as reinforcers. Ignacio’s reinforcers were Wild Berry Skittles and mini-Oreos. Zola’s reinforcers were fruit gummies and M&Ms.

To control for the potential effects of existing color and shape preferences, we conducted color and shape multiple stimulus without replacement preference assessments (MSWO; DeLeon & Iwata, 1996) with nine color flash cards and nine laminated shapes. We randomly selected and assigned two moderately preferred colors and shapes to each the FR and VR condition’s token boards and tokens.

Task probes

We conducted task probes to identify 10 similar acquisition tasks (e.g., ten sight words) per participant based on their current behavioral program. During probe sessions, the experimenter presented the same task across at least five consecutive trials. For each trial, the experimenter presented the relevant discriminative stimulus and contingent on both correct and incorrect responses provided a neutral statement (e.g., “okay”) and proceeded to the next trial. Acquisition tasks were defined as those with 0-20% correct responding across three consecutive sessions. Tasks that met the definition were included in the experiment proper. If a participant responded correctly
during 40% or more of trials in a probe session, the experimenter did not conduct further
probes with that task and the task was not included in the experiment. We continued to
conduct probes until ten acquisition tasks were identified. Five acquisition tasks were
assigned to each of the two conditions. Ignacio’s selected tasks were reading sight
words with the words not, sat, went, our, and the assigned to the FR condition and new,
saw, with, out, and ten assigned to the VR condition. Zola’s tasks were reading single
uppercase letters with the targets B, D, Q, K, and M assigned to the FR condition and
G, Z, U, J, and N assigned to the VR condition.

**Token training**

All participants experienced token training, which started with ten pairing trials,
during which the experimenter delivered a token noncontingently immediately followed
by one of two semi-randomly rotated reinforcers. Participants accessed one edible or
one minute of access to a leisure item for each pairing trial. To ensure similar histories
with each of the token boards, the experimenter delivered tokens on the VR token board
during half of pairing trials and on the FR token board during the other half of trials. After
pairing, participants began exchange-production training during which the experimenter
delivered tokens noncontingently on a token board and used LTM prompting to evoke
an exchange response (i.e., release the token board into the experimenter’s hand).
Following an exchange response, the participant received one of two semi-randomly
rotated edibles. Sessions consisted of 10 total trials with five trials conducted with each
of the two token boards. Exchange training ended when the participant completed two
consecutive sessions with 100% independent exchanges.

Following exchange production training, participants began token production
training, during which the experimenter presented a mastered task and used LTM
prompting to evoke task completion. Contingent on task completion, the experimenter
delivered a token on a token board. Sessions consisted of 10 trials each. Initially,
participants immediately exchanged each token for an edible reinforcer (i.e., an FR 1
exchange-production schedule). During this initial stage, tokens were delivered on the
VR and FR token boards during five trials per session. Once the participant met the
same mastery criteria described above, the number of tokens required to access an
exchange opportunity increased to an FR and VR 2. During FR 2 sessions, the
experimenter delivered tokens only on the FR token board and participants exchanged
tokens only after they earned two tokens. Upon an exchange, participants accessed two
small pieces of an edible or two minutes of access to the leisure item. During VR 2
sessions, the experimenter delivered tokens only on the VR token board and
participants exchanged tokens only after they earned an average of two tokens across
opportunities. Regardless of the number of tokens earned during VR 2 trials,
participants always earned exactly two edibles or two minutes of access to leisure
items. FR and VR 2 sessions alternated and consisted of five trials per session. Once
the participant met the above mastery criteria in both the FR and VR conditions, the
exchange-production schedules increased to FR and VR 5. FR and VR 5 sessions were
conducted identically to FR and VR 2 sessions except participants exchanged tokens
when they earned exactly five or an average of five tokens across opportunities,
respectively. Additionally, participants earned five small pieces of edibles or five minutes
of access to leisure items for each exchange in both conditions. We terminated token
training when participants met the mastery criteria in the FR and VR 5 conditions.
General Procedures: Rate of Acquisition Comparison

We used a nonconcurrent multiple-baseline design across participants, with an embedded multielement design, to compare the rate of acquisition of tasks reinforced with tokens in a token economy with an FR vs. a VR exchange-production schedule. Participants sat at a table with relevant task materials present, including corresponding tablecloths and token boards. To enhance discriminability between conditions, participants sat on opposite sides of the table during FR and VR conditions. One to five sessions were conducted per day. Sessions started when the experimenter presented a token board and began presenting acquisition tasks. Sessions ended when participants earned the requisite number of tokens to produce an exchange opportunity for backup reinforcers or if 20 consecutive trials with no correct independent responses elapsed. A 30-s inter-trial interval was programmed between consecutive trials to further enhance the difference between FR and VR schedules.

Baseline

The data from the response probe sessions were used as baseline data for the experiment.

FR-VR comparison

Upon completing the probes and prior to beginning the FR-VR comparison, the experimenter semi-randomly assigned five acquisition tasks to the FR condition and the other five acquisition responses to the VR condition. While assigning tasks, the experimenter attempted to equate for difficulty across conditions as much as possible. Based on recommendations from Cariveau et al. (2021), the experimenter assigned sight words and letters that rhymed and that started with the same or similar sounds to different conditions whenever possible. We also attempted to equate for familiarity by
including one less-frequently used letter in each condition for Zola (i.e., Q and Z). Additionally, for Ignacio, we equated for the number of letters and syllables in the words in each condition. With these considerations in mind, we preemptively identified which targets would need to be assigned to different conditions and then used excel random number generator to assign each of these targets within the constraints described above.

**FR 5.** During the FR condition, the token board with the five separate Velcro dots was placed on the table in sight and in reach of the participant throughout the entire session. For each trial, the experimenter presented the relevant SD (e.g., “What does this say?”) and, contingent on correct responding, delivered a praise statement (e.g., “That’s right! You got a token!”) and a token. To promote independent correct responding during the initial stages of acquisition, the experimenter initially delivered errorless prompts by immediately following the SD with a model prompt for the first two trials for each acquisition response. The experimenter delivered tokens for each of these two initially prompted responses. Thereafter, tokens were only delivered contingent on independent correct responses. Contingent on an incorrect response, the experimenter used LTM prompting to evoke a correct response. Following prompted responses, the experimenter delivered praise only and presented the next trial. The experimenter continued to present tasks until the participant earned five tokens or until the participant emitted only incorrect or prompted responses across 20 consecutive trials. If the participant earned all five tokens, the experimenter instructed the participant to exchange their token board. Exchange responses resulted in five edibles for participants whose backup reinforcers were edibles or in 5 minutes of access to leisure
items for participants whose backup reinforcers were leisure items. If the participant emitted incorrect or prompted responses across 20 consecutive trials, the experimenter informed the participant the sessions was “all done” and the participant left the session room.

Throughout this condition, the five acquisition tasks were presented semi-randomly such that all tasks were presented at least once every five trials. Acquisition responses were considered mastered once the participant emitted 15 consecutive independent and correct responses.

**VR 5.** VR 5 sessions were conducted exactly like FR 5 sessions except that the number of tokens required to produce an exchange opportunity varied between 1 and 9 tokens from one session to the next. The number of tokens required for each session was randomly determined using excel prior to each session with the constraint that all nine possibilities were experienced once before any was repeated. To equate the presentation of acquisition tasks in the VR condition as much as possible to the FR condition, the five acquisition tasks in the VR condition were presented semi-randomly such that each task was presented once before any other task was repeated.

The FR-VR comparison ended when participants met the mastery criteria in both conditions or when 25 sessions elapsed without meeting mastery criteria in one or both conditions, whichever came first.

**Results**

Figure 4-1 shows cumulative correct responses across the FR and VR conditions for Ignacio and Zola. The top panel of Figure 4-1 depicts Ignacio’s results. During baseline, Ignacio engaged in zero correct responses across both conditions. During the FR-VR comparison, he has consistently responded correctly at a steeper rate during the
VR schedule. In the FR condition, Ignacio has engaged in four series of 20 consecutive trials during which he did not emit any correct responses. Ignacio has engaged in a total of 58 correct responses in the VR condition and 21 in the FR condition across thirteen sessions each. He has completed a total of 283 FR trials and 227 VR trials. At this rate, Ignacio is more likely to meet mastery criteria in the VR condition.

The bottom panel of figure 4-1 shows Zola’s results. During baseline, Zola engaged in zero correct responses in both conditions. During the FR-VR comparison, she initially engaged in similar rates of independent correct responses in both conditions. However, after approximately trial 70, she began to earn all her tokens in fewer trials and was responding correctly in the FR condition. However, her rate of correct responding in the VR condition began increasing after trial 130. Zola has emitted a total of 39 and 38 correct responses in the FR and VR conditions across 13 and 14 sessions, respectively. Zola has completed a total of 189 and 244 FR and VR trials, respectively.

Figure 4-2 depicts percent correct response across sessions for Ignacio and Zola. The top panel shows Ignacio’s results. During baseline, Ignacio engaged in 0% correct responses. During the FR-VR comparison, the VR condition produced greater percentages of correct responses with a slight increasing trend during the last five sessions. Across sessions, Ignacio’s percent correct responding had a mean of 22.62% and 4.15% during FR and VR sessions, respectively. Notably, during the last five sessions of FR, percent correct responding was consistently at zero or near-zero levels.

The bottom panel of Figure 4-2 shows Zola’s percent correct responses across sessions. During baseline, Zola engaged in zero correct responses. During the FR-VR
comparison, the percent correct responding across conditions generally overlapped. Mean percent correct response was 18.08\% and 20.36\% in FR and VR, respectively. Of note, Zola achieved 100\% correct responding in one VR session and none in the FR condition. However, that particular VR session consisted of only 3 trials.

**Discussion**

The results suggest that the effects of FR vs. VR exchange-production schedules in a token economy on the rate of skill acquisition likely varies by learner. For example, Ignacio reliably performed better under the VR schedule, while Zola has generally performed better under the FR schedule until a recent increase in rate under the VR schedule. Of note, both Ignacio and Zola’s target responses were vocal responses. Therefore, it is possible that one or more targets in a given condition (i.e., FR for Ignacio and VR for Zola) were more difficult to emit than the targets in the other condition. This possibility might be more likely for Ignacio who consistently performed better in the VR condition. However, anecdotally, the experimenter has previously observed both participants emit all targeted vocals as echoics or spontaneous vocalizations outside of the experimental context. Additionally, we attempted to equate for target difficulty across conditions by following recommendations from Cariveau et al. (2021) whenever possible. For example, we assigned targets that rhymed and or started with the same letter or phonetic sound to different conditions when possible. However, for Zola, it was not always possible to strictly adhere to these guidelines because many of the target letters identified via response probes rhymed or started with similar sounds. In these cases, we attempted to balance the number of similarities across conditions. For example, there were two targets in each condition that rhymed with each other (e.g., Q and U) and two that began with a similar phonetic sound (e.g., G and J). Despite our
efforts to equate for task difficulty across conditions, it is possible that some targets might have been inherently more difficult than others. The challenge of equating target difficulty across conditions is a prevalent one in research that implements alternating treatment designs to compare intervention effects in acquisition (Cariveau et al., 2021).

Nonetheless, the current results have important implications for how clinicians arrange exchange-production schedules in token economies. Prior research has found that FR and VR exchange-production schedules do not differentially impact rate of mastered responses (Argueta et al., 2019), suggesting that clinicians may consider other variables (e.g., client preference, environmental constraints) when selecting an exchange-production schedule. However, clinicians report using token economies to target both maintenance and skill acquisition (Fernandez et al., 2021). As such, when using token economies to target skill acquisition with or without targeting maintenance, individual differences in the rate of acquisition under FR and VR exchange-production schedules should inform which schedule clinicians select. However, given the constraints of clinical practice, it might not be feasible to conduct the type of extended analysis involved in this study. On the other hand, initially conducting an extended analysis of this type might save valuable instructional time that might be lost using a less effective exchange-production schedule in the future.

The current results are interesting given basic research findings that tokens delivered early on in an FR exchange-production schedule suppress responding likely because they signal a large delay to backup reinforcers and a large response requirement (Bullock & Hackenberg, 2015). If the same were true with the current population, one might expect VR schedules to produce less suppression and perhaps
faster acquisition. However, Zola’s results indicate that FR schedules might sometimes produce a higher rate of acquisition, although generally, the FR schedule has performed only slightly better. It is possible that individuals like Zola might lack sufficient discrimination skills, preventing early tokens in an FR exchange-production schedule from suppressing responding. However, we did not assess participants discrimination skills and, therefore, cannot ascertain the extent to which such skills might influence results. Anecdotally, Zola appeared to visually attend less to task materials and to the experimenter after earning at least a few tokens in the VR schedule. As suggested in Study 2, it appears that tokens likely have discriminative functions that have an important impact on behavior reinforced with tokens. Therefore, future researchers should investigate the potential discriminative and/or suppressive effects of tokens earned at different points in FR and VR exchange-production schedules to better inform how these schedules are arranged in clinical practice.

Relationally, a limitation that might have impacted this study is how we selected the values used in the VR schedule. We used values between one and $2n-1$ (e.g., Argueta et al., 2019; Foster et al., 2001), which resulted in a requirement as high as nine tokens in some VR sessions. It is possible that values at the higher end of our range (e.g., eight and nine) might have been too high and approached ratio strain for Zola. However, the later increase in her rate of acquisition in that context suggest this might not necessarily be a problem after all. Nonetheless, in the future, researchers should consider selecting lower VR values to reduce such potential effects and should consider evaluating the best way to select the range of VR values in an exchange-production schedule.
In sum, it appears the FR and VR exchange-production schedules might differentially affect the rate of acquisition for target responses, but that the effect likely varies by learner. As such, when utilizing token economies exclusively for skill acquisition or for both skill acquisition and maintenance, the effects of FR and VR exchange-production schedules on skill acquisition should be carefully evaluated to select the most efficacious arrangement for a given learner.
Figure 4-1. Cumulative correct responses for FR-VR comparison for Ignacio (top panel) and Zola (bottom panel).
Figure 4-2. Percent correct responses across sessions for the FR-VR comparison for Ignacio (top panel) and Zola (bottom panel).
CHAPTER 5
GENERAL DISCUSSION

Individuals with ASD and/or IDDs sometimes have restricted interests (American Psychiatric Association, 2013; Boyd et al., 2012; Rodriguez & Thompson, 2015) and few reinforcers. Unfortunately, a lack of reinforcers can interfere with treatment by limiting the individual’s motivation to learn (Egel, 1981; Koegel & Egel, 1979). Additionally, the few existing reinforcers that an individual might have can be of limited utility due to ethical and practical considerations. For example, some reinforcers might be unhealthy or potentially harmful, requiring minimal use (Behavior Analyst Certification Board, 2014). Moreover, some reinforcers might not always be available, or their delivery might be disruptive to the environment (e.g., playing music might distract classmates) or interfere with ongoing responses. Therefore, establishing neutral stimuli as conditioned reinforcers is often a crucial objective in treatment for individuals with ASD and/or IDDs.

Tokens are one type of neutral stimuli that one might attempt to establish as a conditioned reinforcer, which requires implementing a pairing procedure. Although several authors have made recommendations for pairing tokens, there is a gap in the applied behavior analysis literature examining pairing procedures for establishing tokens as conditioned reinforcers. Therefore, Studies 1 and 2 attempted to identify best practices for establishing tokens as conditioned reinforcers.

Specifically, Study 1 was a systematic review of S-S pairing, R-S pairing, and ODT for establishing any stimulus, including tokens, as a conditioned reinforcer for individuals with IDDs. However, none of the included studies
attempted to establish tokens as conditioned reinforcers, further highlighting the lack of token pairing research in our literature. Additionally, the findings of Study 1 revealed that, together, S-S pairing, R-S pairing, and ODT are effective at establishing conditioned reinforcers only in half of implementations. Although R-S pairing was the most effective of the three procedures, it was successful only in 61% of implementations. As such, these findings indicate that we do not yet have a well-developed applied technology for establishing neutral stimuli as conditioned reinforcers for individuals with IDDs. As highlighted by the lack of token pairing research found in this review, our pairing technology is further underdeveloped with respect to tokens specifically. Therefore, the results of Study 1 underscore the need for future research evaluating methods to improve the effectiveness of pairing procedures and methods for pairing tokens.

Therefore, Study 2 sought to evaluate S-S, R-S, S-E-S, and R-S-E pairing for establishing tokens as conditioned reinforcers. The results of the token test condition in Study 2 indicated that none of these procedures established tokens as conditioned reinforcers, at least not in the manner in which tokens are typically defined. However, when we tested the effects of R-S-E pairing in a condition in which tokens continued to be intermittently paired with backup reinforcers, we saw an immediate reinforcement effect. As such, Study 2’s results suggest that tokens might not function as conditioned reinforcers as they are traditionally conceptualized. Instead, tokens appear to function more like discriminative stimuli in that they only support responding if they are predictive of forthcoming reinforcement. The results of Study 2 also have implications for the conditions under which we evaluate the effects of token pairing procedures. If tokens’ ability
to support behavior depends on whether they are currently predictive of the delivery of reinforcers, then testing under extinction conditions, in which tokens no longer signal the delivery of reinforcers, might impede detecting the effects of pairing procedures. As such, future research should investigate the ideal conditions under which to test for the effects of token pairing procedures. Relatedly, given that the thinned token test condition only followed R-S-E pairing, we cannot rule out that S-S, R-S, and S-E-S pairing might have supported behavior under the same test conditions. Therefore, future research should evaluate the effects of S-S, R-S, and S-E-S pairing by using a test condition similar to the thinned token test condition included in this Study.

Given the results of Study 2 and prior basic research (i.e., Bullock & Hackenberg, 2015) suggesting that tokens have a discriminative function, Study 3 sought to compare the effects of FR and VR exchange-production schedules in a token economy on rate of skill acquisition. Given that tokens delivered early in an FR exchange-production schedule suppress responding because they signal a large response requirement and delay to backup reinforcers, learners might find tokens delivered early in an FR exchange-production schedule less reinforcing. Consequently, learners might be less motivated to learn targets reinforced with early tokens which might reduce the overall rate of acquisition. As such, learners might acquire skills faster under VR exchange-production schedules, in which early tokens do not predict large work requirements or delays to reinforcers.

The results of Study 3 suggest that VR exchange-production schedules might produce faster acquisition for some learners, but perhaps not all. As such,
when making the determination to use a FR or VR exchange-production schedule, clinicians should select the schedule that produces faster acquisition, given prior research indicating these schedules do not differentially impact maintenance responses. Regarding individual differences, it is possible that differences in which procedure produces faster acquisition might be tied to each learner’s existing discrimination skills. For individuals with less developed discrimination skills, tokens delivered early in an FR schedule might not be particularly discriminative of work requirements and delays to backup reinforcers. However, we did not assess discrimination skills and, therefore, future researchers should evaluate learners’ discrimination skills to determine to what extent they might impact how FR and VR exchange-production schedule affect acquisition. Like Study 2, the current results for Study 3 suggest that tokens have a function similar to discriminative stimuli that we do not yet fully understand. Future research should evaluate the discriminative function of tokens and explore how they impact responding in FR and VR exchange-production schedules.

To summarize, the results of these three studies highlight a gap in our understanding of conditioned reinforcers in general and tokens specifically. We do not currently have a well-developed technology for establishing conditioned reinforcers, especially tokens. In fact, it appears that tokens might not function as conditioned reinforcers after all and, instead, function more akin to discriminative stimuli. Moreover, this function might account for how different schedules and other components of token economies impact target behavior. Therefore, further research is needed investigating the function of tokens and how their function might inform how to best arrange token economies to maximize learning.
APPENDIX A
SEARCH STRATEGY FOR PSYCINFO AND ERIC

1. Stimulus-stimulus pairing
2. Stimulus stimulus pairing
3. Stimulus-stimulus training
4. Stimulus stimulus training
5. Stimulus pairing
6. Response-independent pairing
7. Response independent pairing
8. Response noncontingent pairing
9. Response-noncontingent pairing
10. Direct pairing
11. Stimulus-pairing
12. Response-stimulus pairing
13. Response stimulus pairing
14. Response stimulus training
15. Response-stimulus pairing
16. Response-contingent pairing
17. Response contingent pairing
18. Response dependent pairing
19. Response-dependent pairing
20. Discrim* training
21. Discrim* procedure*
22. Discrim* pairing
23. Operant discrim* pairing
24. Operant discrim* training
25. Operant discrim* procedure*
26. Establish* conditioned reinforce*
27. Establish* reinforce*
28. Condition* reinforce*
29. Novel reinforce*
30. New reinforce*
31. Autis*
32. Autis* spectrum disorder*
33. Asperger*
34. Mental* retard*
35. Intellectual disabilit*
36. Intellectual disorder*
37. Intellectual differen*
38. Developmental disabilit*
39. Developmental disorder*
40. Developmental differen*
41. Learning disorder*
42. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25
43. 26 or 27 or 28 or 29 or 30
44. 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41
45. 42 and 43 and 44
LIST OF REFERENCES

References marked with an asterisk indicate studies included in the systematic review.


BIOGRAPHICAL SKETCH

Tracy Argueta received a bachelor's degree in psychology and a bachelor's degree in criminal justice from Florida International University in 2014. She went on to receive her master's in professional behavior analysis from the Florida Institute of Technology in 2016. She received a Ph.D. in psychology, in the area of behavior analysis, from the University of Florida in 2022.