

This is the pre-peer reviewed version of the following article: Frank-Crawford, Michelle A., John C. Borrero, Alyssa Fisher, Paige Talhelm, and Nathalie Fernandez. "Discrete-Trial Teaching: A Scoping Review." *Behavioral Interventions* (23 March 2024). <https://doi.org/10.1002/bin.2012>., which has been published in final form at <https://doi.org/10.1002/bin.2012> . This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

This item may be protected under Title 17 of the U.S. Copyright Law. It is made available by UMBC for non-commercial research and education. For permission to publish or reproduce, please contact the author.

Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

Please provide feedback

Please support the ScholarWorks@UMBC repository by emailing scholarworks-group@umbc.edu and telling us what having access to this work means to you and why it's important to you. Thank you.

Discrete Trial Teaching: A Scoping Review

Michelle A. Frank-Crawford^{1, 2, 3}, John C. Borrero¹, Alyssa Fisher², Paige Talhelm²,
and Nathalie Fernandez^{2, 3}

¹Department of Psychology, University of Maryland, Baltimore County, Baltimore, MD, USA

²Department of Behavioral Psychology, Kennedy Krieger Institute, Baltimore, MD, USA

³Department of Psychiatry and Behavioral Science, Johns Hopkins University, School of
Medicine, Baltimore, MD, USA

Author Note

This research was conducted as part of the first author's requirements for a Doctoral degree in Applied Developmental Psychology at UMBC. Paige Talhelm is now at the University of South Florida. Conflict of Interest: The authors have no conflict of interest to declare. Data Availability: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Please direct correspondence to Michelle Frank-Crawford, Neurobehavioral Unit, Kennedy Krieger Institute, 707 N. Broadway, Baltimore, MD 21205. E-mail: crawfordm@kennedykrieger.org

Abstract

Discrete trial teaching (DTT) is an arrangement used in skill acquisition. The components that comprise DTT vary widely across applications, and previous reviews evaluating its efficacy have largely reported on DTT as part of a comprehensive intervention package. The purpose of this scoping review was twofold: to describe the component variations of DTT (descriptive analysis) and to evaluate the general efficacy of DTT in teaching new skills to individuals with disabilities (efficacy analysis). One hundred and thirty-four studies were included in the descriptive analysis of DTT and eighty-two were included in the efficacy analysis. Results indicated that many of the components of DTT align well with best practice recommendations, including that reinforcers be delivered continuously and immediately following correct responses. Overall, DTT was efficacious in teaching new skills; however, there were limited evaluations of the maintenance, generality, and social validity of the findings. The outcomes are discussed in light of best practice recommendations, and as a guide future practice and research.

Keywords: discrete trial teaching, effectiveness, efficacy

Discrete Trial Teaching: A Scoping Review

Research has demonstrated the efficacy of several behavior analytic strategies for teaching skills to individuals with intellectual and/or developmental disabilities (IDD). One strategy commonly used for providing instruction includes the massed presentation of discrete learning trials encompassing distinct antecedent, response, and consequent components. This method of instruction is referred to as discrete-trial teaching (DTT, also frequently called discrete-trial training and discrete-trial instruction). Discrete-trial teaching includes highly structured, teacher-led instruction in which complex skills are broken down into simplified target responses taught in a fast-paced, trial-based method (Smith, 2001; Tarbox & Najdowski, 2008). During each discrete trial, the instructor presents a discriminative stimulus (i.e., the antecedent component) and, when necessary, uses a predetermined prompting strategy that varies according to the learner's response. Following the response, the instructor then implements an appropriate consequence, such as access to a reinforcer for a correct response or additional prompts for an incorrect response (i.e., consequent component). The instructor implements the components relatively quickly and arranges an intertrial interval (i.e., typically only a few seconds) between the end of one trial and the initiation of the next.

Although perhaps most popularized by Lovaas's research on early intensive behavioral intervention (EIBI) with children with autism spectrum disorder (e.g., Lovaas, 1987), the use of DTT as an instructional method for children with IDD dates as far back as the 1950s and 1960s (Ghezzi, 2007; Lindsley, 1996; Tarbox & Najdowski, 2008; Wolf et al., 1963). Since that time, research has repeatedly supported the use of DTT as both part of a comprehensive intervention package (e.g., Reichow, 2012; Virués-Ortega, 2010; also see Lerman et al., 2016, for a review) and as a stand-alone treatment to teach individuals with IDD a variety of skills such as imitation

(MacDonald & Ahearn, 2015), receptive and expressive language (e.g., Grow et al., 2014; Petursdottir et al., 2005; Plaisance et al., 2016), play (e.g., Coe et al., 1990; Grow et al., 2016; Jahr et al., 2000), and functional communication (e.g., Schmidt et al., 2014).

In general, DTT programming includes the critical antecedent, response, consequent, prompting, and intertrial interval components. However, clinical practice and outcomes from research have led to the identification of several procedural variations to the implementation of these specific components (Lerman et al., 2016). For example, research has evaluated a number of antecedent manipulations (e.g., Petursdottir & Aguilar, 2016), procedures to fade prompts (e.g., Cengher et al., 2016), and intertrial interval durations (e.g., Cariveau et al., 2016). The consequent component of DTT, in particular, has received considerable research attention. Some researchers have focused on consequences that follow incorrect responses, such as error correction procedures (e.g., Carroll et al., 2015), whereas others have manipulated various reinforcer dimensions previously shown to impact responding, such as reinforcer delay (e.g., Carroll et al., 2016; Majdalany et al., 2016; Sy & Vollmer, 2012), magnitude (e.g., Paden & Kodak, 2015), and quality (e.g., Lang et al., 2014). Recommendations regarding the use of reinforcers in the context of DTT often include the delivery of a directly consumable reinforcer, such as a piece of food or brief access to an activity, immediately after each correct response—at least during the acquisition of new skills (e.g., Lerman et al., 2016; Smith, 2001; Tarbox & Najdowski, 2008). This is representative of what some have termed a distributed response-reinforcer arrangement (e.g., DeLeon et al., 2014).

Discrete-trial teaching procedures have been previously summarized in the literature. For example, a number of narrative reviews exist that describe procedural variations within DTT (e.g., Lerman et al., 2016), and there have been numerous papers and chapters published

providing clinical recommendations for how best to arrange DTT (e.g., Smith, 2001; Tarbox & Najdowski, 2008). Other researchers have conducted meta-analyses of EIBI, of which DTT was only one of several components. Outcomes of such reviews generally suggest that these comprehensive behavioral interventions are effective (e.g., Virués-Ortega, 2010). However, because these reviews focused on multi-component EIBI programs, they lack specific details regarding the arrangement of each component of the discrete trials. Thus, we currently have no summary of the types of procedural modifications commonly applied in DTT, nor do we know the types of responses typically taught via DTT. Further, although the meta-analyses provide a quantitative measure of the efficacy of interventions, DTT was implemented simultaneously with other behavioral intervention components; thus, it is difficult to determine the unique effects of DTT from the outcomes of these studies.

As noted, the DTT literature covers a broad range of studies—some studies have focused on modifications to DTT components (e.g., Cariveau et al., 2016; Cengher et al., 2016; DeLeon et al., 2014; Petursdottir & Aguilar, 2016), and other studies have focused on the application of DTT to a new target response or population (e.g., das Neves et al., 2018; Dixon et al., 2017; Lambert et al., 2016). In cases for which a literature base is quite wide-ranging, scoping reviews can provide one means of synthesis to allow researchers to describe current practices, provide guides for decision-making, and identify gaps in the literature; furthermore, scoping reviews can serve as a precursor to a more comprehensive or narrowly focused systematic review (Munn et al., 2018). Given that there is currently no summary unique to DTT in the literature, a scoping review may prove particularly useful to practitioners and researchers. First, it may allow us to determine whether research aligns with best-practice recommendations. Second, a scoping review could provide practitioners with information to guide future practice. For example, there

are many different teaching strategies outside of DTT, including pivotal response training (e.g., Koegel et al., 1987), fluency training (e.g., Nopprapun & Holloway, 2014), and naturalistic instruction (e.g., LeBlanc et al., 2006). It may not always be clear to practitioners which strategy is best for a teaching particular target response; a scoping review of DTT can help to identify which types of responses are most commonly taught using DTT. Finally, some researchers have conducted narrowly focused systematic reviews on a specific component of discrete trials, such as the consequence component (e.g., Weinsztok et al., 2023). A scoping review of DTT may help guide other, similarly focused systematic reviews.

Therefore, the purpose of the current study was to conduct a scoping review of DTT practices. We aimed to (a) describe the components of DTT as it has been employed to teach new skills to persons with IDD, and (b) to report on the general efficacy of DTT in the context of skill acquisition. We then discuss how the outcomes of the review relate to best practice recommendations, can be used to guide practice, and can inform future research.

Method

Study Identification

A scoping review of studies that employed DTT in the context of skill acquisition in the last two decades, between the years 2000 and 2021, was conducted following guidelines outlined in the Preferred Reporting Items for a Systematic Review and Meta-Analysis of Individual Participant Data Statement (Stewart et al., 2015). The purpose of the search was to identify as many relevant studies as possible to provide a large and representative (not necessarily comprehensive) sample of the literature on DTT. Two methods were used to identify relevant articles. First, an electronic database search of *Springer Link*, *PsychINFO*, *PubMed Central*, and *ERIC* was conducted using the search terms “discrete trial training,” “discrete trial teaching,” or

“discrete trial instruction” to identify relevant publications and unpublished dissertations. The rationale for starting the search in the year 2000 was threefold. An initial search employing this method for the years 1987 (year of publication for one of Lovaas’s seminal papers on EIBI; Lovaas, 1987) through 2021 indicated that over 90% of papers that reported using DTT in the context of skill acquisition with individuals with IDD were published after the year 2000. In addition, this period is consistent with the publication of early manuals on DTT (e.g., Leaf & McEachin, 1999). Finally, it is also consistent with other reviews that evaluated the use of DTT in the context of EIBI (e.g., Lerman et al., 2016). The second search method consisted of an ancestral search reviewing the references of all papers identified using the database search that meet inclusion criteria (outlined below). This second step increased the chances of identifying studies that may have used DTT but did not refer to it as DTT and would therefore not have been identified during the database search.

Inclusion Criteria

The study consisted of two levels of analysis, a descriptive analysis, and an efficacy analysis. All studies included in the review were required to meet the following criteria: (a) the study was conducted with at least one individual with IDD as a participant (if the study included those with and without an IDD, only those with the IDD who met other inclusion criteria were included in the review); (b) DTT was assessed in isolation in the study (e.g., outside of a comprehensive EIBI program or other teaching strategy such as shaping or chaining);¹ (c) DTT was used to teach the participant a previously unmastered behavior; and (d) the antecedent, response, consequence, and prompting components of DTT were described with enough detail to

¹ It is not uncommon for a variety of teaching strategies to be used throughout an individual’s educational programming. Studies were only excluded if additional teaching strategies were included in the study itself and not if the authors mentioned the use of other strategies outside the context of the research study.

distinguish among the specific procedures in place (e.g., to distinguish between the different types of prompting procedures that could have been employed; see Supporting Information A for definitions of each type of procedure coded). For the purposes of the review, *DTT* was defined as a teaching procedure characterized by the massed presentation of at least two learning trials where each trial was controlled by the instructor. That is, the instructor initiated each trial by presenting a clearly defined antecedent, provided predetermined prompts following no response or an incorrect response, consequated correct responses, and implemented an intertrial interval prior to initiating the next trial. A minimum of two trials was necessary as DTT inherently involves the repeated presentation of trials and implementation of an intertrial interval, neither of which was satisfied with only a single trial.

If any of the antecedent, response, consequence, or prompting components of DTT were not clearly defined, the study was excluded. The antecedent information was necessary to determine what initiated a trial, and an example of insufficient information included vague statements that “an antecedent” was issued to start the trial. The response description was required as it was tied to defining the dependent variable. Although no studies were excluded solely because they lacked sufficient information on the response, some studies failed to provide adequate information on several components, including the response, and were thus excluded for those reasons. The consequence description was necessary to determine the response-reinforcer arrangement used, and the prompting methods and sequence allowed for the determination of how the response was prompted and what consequences followed correct and incorrect responses. Examples of insufficient information regarding the consequent component included descriptions of delivering reinforcers without defining the reinforcer type or magnitude and/or reinforcement schedule or noting that prompting was used when needed without describing the

type or sequence of prompts employed. Studies in which the authors did not define the intertrial interval used during DTT were not excluded for several reasons. Excluding studies in which the intertrial interval was not reported would have led to the elimination of a substantial number of studies (i.e., 63% of the studies included in this review did not define the interval), thus limiting the generality of the findings.

The purpose of the descriptive analysis was to describe how researchers typically arranged components of DTT during skill acquisition. Studies were included in the descriptive analysis if they met the criteria outlined above. We were also interested in how well DTT worked. Thus, the efficacy analysis was conducted to examine the general efficacy of DTT in teaching new skills. Studies included in the efficacy analysis had to meet the inclusion criteria outlined above and were subjected to two additional inclusion criteria: (e) the article depicted data from the DTT phase of the study and (f) the study met experimental standards for single case research designs as described by the What Works Clearinghouse (Kratochwill et al., 2010). To be clear, all studies meeting Criteria a–d outlined above were included in the descriptive analysis; the subset of those studies that also met criteria (e) and (f) were included in the efficacy analysis.

Study Coding Procedures

A coding system was developed to synthesize data from various aspects of the studies included in each analysis (see Supporting Information A). Generally, the aspects included (a) descriptive characteristics of the participants and applications of DTT (descriptive analysis) and (b) dependent variable features and efficacy (efficacy analysis).

Descriptive Analysis

The purposes of the descriptive analysis were to provide a description of the population for whom DTT was employed, the DTT components commonly used, and whether procedural fidelity, generalization, maintenance, and social validity were evaluated for DTT arrangements.

Participant and Setting Characteristics. For each participant included in the study, data collectors recorded the individual's gender, age, presence or absence of a noted autism spectrum disorder, level of intellectual disability (if applicable), and other noted genetic or medical conditions. Data collectors also recorded whether the participant was reported to have a history with the use of DTT, and information on the intervention setting and the interventionist.

Intervention settings for the primary DTT intervention included a clinic or center, school, home or residential placement, hospital setting (separated by inpatient and outpatient), or unknown if the authors did not explicitly note the setting type. In some cases, multiple settings were noted. If generalization occurred in a separate setting, this was not included in the intervention setting and was instead noted in the description of generalization practices. *Interventionists* were categorized as the experimenter if DTT was noted to be implemented by (a) an “experimenter” or a study author, (b) electronic if the DTT program was presented primarily through a computer program or application, (c) teacher if conducted by an individual who regularly provided educational instruction to the participant, (d) peer if DDT was conducted by an individual in the same age group (with or without an IDD), (e) college student if the authors did not also note that the student fit one of the other categories (e.g., if the student was also not an experimenter), (f) parent if the individual was noted to be a primary caregiver for the participant, or as (g) multiple implementers if persons from more than one category took part in implementing the sessions

(e.g., the teaching sessions were completed electronically and by an experimenter; note that this excluded the use of novel interventionists for use during generalization).

Discrete-Trial Teaching Component Characteristics. We also wanted to examine how researchers have labeled and described the use of DTT. To do so, data collectors searched each paper to determine whether the terms “discrete-trial teaching,” “discrete-trial training,” “discrete-trial instruction,” or “discrete trial(s)” were included in the study to describe the teaching procedure employed. Each of the five components that encompassed DTT were then recorded, including the antecedent used to initiate a trial, the participant response, the consequence delivered for correct responses, the prompting strategy employed, and the duration of the intertrial interval. The *antecedent* was categorized as physical (a stimulus was presented in the absence of a vocal verbal prompt), vocal (a vocal verbal stimulus was presented in the absence of a physical stimulus), physical and vocal, or electronic (the stimulus was presented via a computer program or application). The *response* was characterized as receptive (identification or instruction following), verbal operants (mand, tact, echoic, intraverbal, textual, or transcription), play, or imitation. *Consequences* included attention (e.g., praise), food, tangible (e.g., toys or activities such as completing a puzzle), token, auditory (i.e., computer-generated tones), or multiple. In some cases, no explicit consequences were arranged, and the next trial was initiated following a response. In other cases, reinforcers from different classes were included and could vary unsystematically across sessions (i.e., both food and tangible stimuli were included in some studies, and it was unclear which reinforcer was used on a given trial). These were recorded as “nondescript (food or tangible).”

Corrective prompting included the predetermined prompting strategy employed in the event of an incorrect or no response on the part of the participant. Many studies used the term

“error correction” to describe several procedurally different strategies; thus, to differentiate between prompting strategies employed in the current review, *error correction* was defined as the representation of the entire trial until the participant responded correctly to the initial antecedent prompt. For example, contingent on an incorrect response, the experimenter provided feedback, removed the stimulus materials, and then represented the trial by initiating the antecedent prompt; this continued until the participant responded correctly to the antecedent. Error correction was differentiated from *repeated corrective prompting* in which a corrective prompt was represented until the participant responded correctly to the corrective prompt itself (e.g., contingent on an incorrect response, the experimenter continually presented a model prompt until the participant responded correctly following the model). A *single corrective prompt* included the provision of feedback for an incorrect response (e.g., a model prompt or “no, that is not correct”), followed by the initiation of the next trial. *Prompt delay* included the presentation of the antecedent followed by some predetermined delay to initiate a prompt; the delay then increased either within or across sessions (e.g., in a progressive prompt delay strategy, the antecedent would be immediately followed by a prompt to increase the likelihood of a correct response and the duration to the prompt would increase incrementally across sessions, such as to 2 s and then 5 s). *Stimulus prompt fading* included the modification of some feature of the task stimuli to increase the likelihood of a correct response (e.g., altering the positioning of the correct stimulus). A *prompt hierarchy* was recorded if successive response prompts were issued where they either increased or decreased in level of assistance administered (e.g., least-to-most prompting consisting of sequentially implemented vocal, model, and physical guidance prompts). *Errorless learning* included immediate prompting of the target following the presentation of the antecedent. *Flexible prompting* was defined as a dynamic prompting strategy

that relied on the instructor using judgment to decide when and how to prompt. In some cases, corrective prompts were not used, and incorrect responses were followed by the presentation of the next trial; these were coded as “no corrective prompts used.”

Finally, data collectors recorded the numerical (e.g., “5 s”) or narrative (e.g., “brief”) duration of the intertrial interval. If the authors provided no information on the interval duration, data collectors recorded “unknown.”

Other Descriptive Information. For each study, the number of trials in each DTT session was categorized as 2–5, 6–10, 11–15, 16–20, > 20, or varied. *Varied* was coded if the number of trials differed across phases of the study and met criteria for two or more categories. For example, if the study included eight trials one phase and 10 trials in another phase, the study would have been categorized as having 6–10 trials. Alternatively, if the study included 10 trials in one phase and 20 trials were in another phase, “varied” would have been coded. “Varied” would also be coded if the session was time-based and thus the number of trials completed could have differed across sessions.

Data collectors also recorded the presence or absence of data on procedural fidelity, generalization, maintenance, and social validity. *Procedural fidelity* included any measure of how well DTT was implemented as prescribed. Several other terms have been used synonymously with procedural fidelity (e.g., treatment fidelity, treatment integrity; see Essig et al., 2023, and St. Peter et al., 2023); thus, any use of a variation of this term was taken to mean “fidelity” and was coded as such. *Generalization* included an evaluation of the extent to which the effects of the intervention occurred in novel conditions (i.e., in a new environment, with new instructors, or with new task stimuli). *Maintenance* included evaluations of whether the effects of DTT continued in the absence of the intervention after mastery had been achieved. Finally,

social validity included evaluations of the acceptability of the intervention typically measured indirectly through questionnaires administered to stakeholders (e.g., caregivers of the person receiving the intervention) or directly by examining the preference of the individual receiving the intervention.

We also coded the type of study that included the DTT evaluation. Specifically, studies included in this review were categorized according to one of four study types: (a) the application of DTT to a new target (e.g., using DTT with an understudied population or to teach an understudied response), (b) comparisons of variations of the components of DTT (e.g., comparing different prompting strategies or different response-reinforcer arrangements), (c) comparisons of DTT to other intervention types (e.g., comparisons of DTT to fluency-based teaching), and (d) other study types (e.g., evaluations of data collection or applications of different mastery criteria).

Efficacy Analysis

The purposes of the efficacy analysis were to describe how the dependent variable was evaluated and, in general, how efficacious DTT was in teaching new skills. All studies were included in the descriptive analysis; only the subset of applications of DTT that met What Works Clearinghouse design standards (Kratochwill et al., 2010) were also included in the efficacy analysis. Specifically, each application within each study was evaluated to determine whether: (a) the independent variable was systematically manipulated by the experimenter, (b) the effects of DTT on the dependent variable were measured systematically over time, (c) interobserver agreement was assessed for at least 20% of sessions, (d) each application of DTT included at least three attempts to demonstrate an intervention effect or four if using an alternating treatment design (i.e., a minimum of three baselines for a multiple baseline or multiple-probe designs, three

reversals for a reversal design, or four alternations between conditions for an alternating treatment design), and (e) each attempt included at least three data points (four if using an alternating treatment design). The type of design was also recorded. Specifically, single-case designs were recorded as a *multiple baseline* if the effects of DTT were evaluated across at least three participants, settings, or responses. *Multiple-probe* designs included a multiple baseline design in which the baseline sessions were conducted intermittently or probed. A *reversal* included at least two repetitions each of a baseline and the DTT intervention phase. *Alternating treatments* included rapid alternation within the same timeframe with two or more conditions. A *combination* design was recorded for designs that included two or more of the above single-case designs.

Data collectors recorded how the dependent variable was measured and the mastery criteria employed. *Measurement* included the metric by which the dependent variable was captured and was categorized as percentage of correct responses, frequency of targets mastered, trials to reach mastery criterion, or response rate. *Mastery criteria*, if noted, were transcribed narratively and then categorized as the percentage of trials with correct responses. Categories included: 100% correct responses, 90%–99% correct responding, 80%–89% correct responding, and less than 80% correct responding. In some cases, multiple criteria were noted; for example, Reichow and Wolery (2011) required participants to respond correctly on (a) 100% of trials (b) under a thinned schedule of reinforcement.

As a general measure of efficacy, data collectors recorded whether a given application of DTT met the mastery criteria documented in the study from which the dataset being coded were extracted. Efficacy outcomes were recorded dichotomously as either “yes” if the mastery criteria

were satisfied or “no” if the given application of DTT did not meet the criteria. Efficacy was evaluated across all studies.

Intercoder Agreement

Intercoder agreement was evaluated for abstract reviews, inclusion criteria, and data extraction. Point-by-point agreement was assessed for each item coded. An agreement included both coders recording the same response (a “Yes” or a “No”) for each item. A disagreement included coder responses that differed across the same item (e.g., one data collector recording a “Yes” and the other recording a “No” for any given item). Prior to initiating any level of review, data collectors reviewed the definitions for each inclusion criterion and item for which data were to be extracted in the descriptive and efficacy analyses. The data collectors then met with the first author who provided training in the form of vocal instruction, modeling (e.g., reviewing the first three papers and providing step-by-step instructions for each variable to be coded), and feedback after the data collectors coded the first three studies without any assistance.

Disagreements were resolved prior to the data collectors continuing with data extraction.

For the abstract review, a second data collector independently reviewed 25.1% of abstracts from all non-duplicate papers identified in the initial electronic database search and the ancestral search. Agreement on inclusion of an article for full-text review equaled 96.2% (range: 92.8%–100%). The papers that were not excluded based on the abstract review were then subjected to full-text review for inclusion criteria. For this step, a second data collector independently reviewed 20.7% of studies. Mean agreement across inclusion criteria equaled 97.2% (range: 92%–100%). Studies that met inclusion criteria were then subjected to data extraction for the descriptive analysis and efficacy analysis. A second data collector reviewed and coded 20.9% of studies that met inclusion criteria for these analyses. Mean agreement

equaled 98.4% (range: 85.7%–100%) across all extracted data (see Supporting Information B for intercoder agreement scores).

Results

Figure 1 depicts the results of each step of the search process and application of inclusion criteria. Initial results of the literature search identified 1,111 non-duplicate citations, 753 of which were excluded after screening the titles and abstracts. Three hundred and fifty-eight studies were retrieved and reviewed for inclusion criteria for each of the analyses. The number of studies that met inclusion criteria for the descriptive analysis and the efficacy analysis equaled 134 and 82, respectively. Supporting Information C includes a reference list of all studies included in the descriptive and efficacy analyses.

Descriptive Analysis

Participant and Setting Characteristics

The 134 studies that met criteria for the descriptive analysis included a total of 1,860 individual applications of DTT across 652 participants. Table 1 depicts the participant, intervention setting, and interventionist characteristics for the 652 participants. For those individuals for whom age was reported, the largest age group for participants receiving instruction via DTT was less than 6 years of age. These results are perhaps not surprising because DTT is often incorporated into comprehensive EIBI programs (e.g., Lerman et al., 2016), the recipients of which are typically under the age of 6 (e.g., Love et al., 2009; Peters-Scheffer et al., 2011). The second largest group included grade-school-aged children between 6 and 11 years of age.

Most participants were male. The high percentage of male participants is understandable given that almost 77% of the participants had a diagnosis of autism spectrum disorder, as autism

is more prevalent among males (Baio et al., 2018). Approximately 31% of participants were also reported to have an intellectual disability, the majority of which were either unspecified in level of severity (40.4%) or in the mild to borderline range (28.1%). Seventeen percent of all participants were reported to have a genetic or medical condition, with 39.3% of those diagnosed with fragile X syndrome. Other diagnoses included Down syndrome (16.1%), an unspecified language disorder (10.7%), and attention deficit hyperactivity disorder (8%). The authors of the studies included in the review reported that 25.8% of participants had some history with the use of DTT prior to participating in the research study. In only 2.3% of cases was it explicitly noted that the participant(s) had never been exposed to DTT prior to the study. In the remaining cases, the authors did not report on participant history with DTT.

Most settings included a clinic or center that provided services to children with autism spectrum disorder and/or related disabilities (44.6%) or in the participant's school (29.4%). For 11.5% of participants, DTT was conducted in the home. More than half of the interventionists included an experimenter (62.1%), whereas only 13% included a teacher. DTT was implemented electronically in 18.3% of cases and by a parent in less than 1% of cases.

Discrete-Trial Teaching Component Characteristics

Figure 2 depicts the cumulative use of the term DTT (or some variant thereof) over the years evaluated in the current study. Of the 134 studies included, 92 (68.7%) explicitly reported using DTT to teach a target response. The other studies described the components of DTT without referring to it as such. These studies typically noted that "trials" or "learning trials" were conducted. Authors were equally likely to refer to the teaching arrangement as DTT as they were to refer to it as something else until approximately 2012; thereafter, the use of DTT to describe the teaching arrangement increased more rapidly.

Table 2 depicts the DTT component characteristics of each of the 1,860 applications. Most antecedents included the presentation of both a physical and vocal stimulus (75.1%). Receptive identification tasks were taught in 44.8% of analyses, followed by tacts in 32.2% of analyses. The remaining target responses were taught in less than 10% of applications. Table 3 documents the number of analyses conducted for each response type according to the category of antecedent provided. A vocal stimulus presented simultaneously with a physical stimulus was used for most response types; the exception was when teaching intraverbals. With intraverbals, a vocal stimulus alone was used as the antecedent more often than the presentation of the physical and vocal stimuli simultaneously.

Corrective prompting strategies in DTT varied widely. The most commonly used prompting strategies included the provision of a single corrective prompt following an incorrect or no response (23.1% of applications) and the use of a prompt hierarchy (22.4%). Less than 10% of applications included errorless learning as we have defined it (5.3%), repeated corrective prompting (5.0%), and flexible prompting (2.2%). Authors reported not using corrective prompting in 6.2% of applications.

As noted previously, it is recommended that practitioners keep the intertrial intervals relatively short, less than 5 s (e.g., Ghezzi, 2007; Smith, 2001). Authors did not report the duration of the intertrial interval for 63.2% of applications of DTT; however, most studies that did report the intertrial intervals in the current review were consistent with this recommendation. In 7.6% of applications, the authors' narrative description of the intertrial interval suggested that it was relatively short in duration (e.g., they noted the intertrial interval was "brief"). Intertrial intervals that were reported by authors mostly aligned with clinical recommendations (e.g., Ghezzi, 2007) in that they were 1–5 s in 17.4% of applications or between 6–15 s in 7.5% of

applications. Applications that included longer intervals (e.g., 20–30 s) were those that systematically manipulated the intertrial interval to determine the effects of different durations on acquisition (e.g., Cariveau et al., 2016). In a small minority of studies, the researchers presented new trials on a fixed-time schedule every 30 s (e.g., Neil & Jones, 2015). In those cases, the intertrial interval could not exceed 30 s and was therefore categorized as such.

Other DTT Arrangements

Table 4 depicts other characteristics of the response-reinforcer arrangements and how consequences were identified. Most studies included between 6–10 trials (39.6%) or 11–15 trials (21.6%); very few studies included 5 or fewer trials (8.2%) or more than 20 trials (3%). Trials varied across phases (or due to time-based sessions) for 16.4% of studies.

Stimuli used as reinforcers were most frequently identified via a preference assessment that was either conducted prior to each session (29.1%) or at the start of the study by the experimenters (25%). In 10.2% of applications, the stimulus was identified based on its historical use outside of the study. In 31.6% of the analyses, the experimenters did not describe the methods used to identify the stimuli included as consequences; in most of these cases, attention was the consequence provided.

Most studies implemented a fixed ratio 1 schedule of reinforcement during the initial stages of DTT (i.e., 95.3%; see Table 4). Leaner fixed ratio and variable ratio schedules were used in 3% of applications, and lag reinforcement schedules were used in 0.1% of applications. These data align with recommended practice regarding the use of continuous schedules of reinforcement during the initial stages of skill acquisition (e.g., Ghezzi, 2007; Lerman et al., 2016). It is also recommended that practitioners thin the reinforcement schedule after the learner acquires the response to promote generalization and maintenance of the skill (e.g., Ghezzi,

2007). However, schedule thinning was evaluated in only 2.3% of analyses in the current review. Although schedule thinning may have been conducted outside the confines of the research studies included in this review, the limited published data on schedule thinning during DTT presents an important area for future research.

Procedural Fidelity, Generalization, Maintenance, and Social Validity

Data collectors for the current review also recorded how often experimenters collected data on the accurate implementation of DTT, generalization and maintenance of the skills learned, and acceptability of the intervention. Slightly more than half of the analyses included some data on procedural fidelity (61.9%). Generalization data were only collected for 17.6% of applications. However, experimenters collected maintenance data in 43.7% of DTT applications. Social validity was assessed infrequently, in only 14.9% of cases in which DTT was applied. This finding is consistent with the results of other systematic reviews examining the use of social validity measures in behavior analytic research, which have similarly reported that a relatively low percentage of studies include such measures (e.g., Carr et al., 1999; Kennedy, 1992; Snodgrass et al., 2018).

In addition to evaluating the percentage of studies that included these data, we also evaluated trends in reporting practices. Specifically, Figure 3 depicts the cumulative number of studies that evaluated procedural fidelity, generalization, maintenance, and social validity from 2000 through 2021. The data indicate there has been a rapid increase in the number of studies that have included data on procedural fidelity since 2014. Specifically, the number of studies reporting procedural fidelity data through 2013 was 31 and the number of studies that did not report on procedural fidelity in this time frame was 29. In 2014, the cumulative number that reported on procedural fidelity increased to 37 whereas the number that did not report on these

data equaled 32. The number that reported procedural fidelity data rose steadily through 2021 ($n = 86$), whereas the number of studies that did not report on fidelity increased more slowly ($n = 48$). This increasing trend corresponds with publications of papers specifically focused on the importance of procedural fidelity (e.g., Fryling et al., 2012) and a special issue on procedural fidelity in the *Journal of Behavioral Education* in 2014. However, as St. Peter et al. (2023) have recently noted, reporting of fidelity data is still not commonplace across all behavior-analytic research. A similar, but arguably more modest, increase was observed regarding the number of studies that reported on the maintenance of effects. By the end of 2021, the number of studies reporting on maintenance of effects ($n = 63$) was roughly equal to the number of studies that did not report on maintenance ($n = 71$). Unfortunately, studies that reported on the generality of the findings and social validity lagged considerably. Only 35 studies included data on generality of findings, and only 21 studies included some measure of social validity.

Study Types

Table 5 includes the four types of studies coded. The most common study type included a comparison DTT components, which occurred in 59.0% of all DTT studies. This was followed by the application of DTT to a new target, which comprised 26.9% of DTT studies. Less common were comparisons of DTT to other instruction types, occurring in 5.9% of studies. Finally, studies that could not be classified as one of the above noted study types were coded as “other,” which was coded for 8.2% of studies.

The largest percentage of DTT component comparison studies (36.7%) included comparisons of different prompting strategies. For example, Carroll et al. (2018) conducted an abbreviated assessment to compare the effects of different prompting strategies on correct responding for four individuals with disabilities. Manipulations to the presentation of trials was

the second most common type of DTT component comparison, occurring in 32.9% of applications. This included a comparison of individual- vs. group-trial format (e.g., Jameson et al., 2007), table-top vs. naturalistic-trial presentation (e.g., Geiger et al., 2012), massed vs. distributed trials (e.g. Majdalany et al., 2014), and interspersal of mastered- and unmastered-task trials (Volkert et al., 2008). Evaluations of different consequences comprised 26.6% of DTT components comparisons. These studies included comparisons of different reinforcement parameters, such as delay or magnitude (e.g., Carroll et al., 2016; Paden & Kodak, 2015), and response-reinforcer arrangements (e.g., Frank-Crawford et al., 2019; Joachim & Carroll, 2018). Finally, a small percentage of DTT comparison studies (3.8%) evaluated different intertrial intervals (e.g., Cariveau et al., 2016; Neil et al., 2020). Among the applications to new targets, most (63.9%) were new types of responses being taught. For example, researchers used DTT to teach responses related to different types of sports (e.g., basketball; Lambert et al., 2016), different types of tacts (e.g., private events; Belisle et al., 2020), and attending responses (e.g., Tarbox et al., 2006). For 13.9% of applications, DTT was implemented with a novel implementer or on a novel platform, such as with peers (e.g., Radley et al., 2015) or via telehealth (e.g., Ferguson et al., 2020), respectively. In a smaller percentage of applications (11.1%), DTT was implemented with a novel population, such as with high school students (e.g., Parker & Schuster, 2002) or children with cochlear implants (das Neves et al., 2018). The most common comparison to another instructional strategy was between DTT and fluency training (37.5%; Holding et al., 2011; Nopprapun & Holloway, 2014), followed by mand training (e.g., Jennett et al., 2008), and incidental teaching (e.g., Charlop-Christy & Carpenter, 2000). Finally, in the “other” category, 27.3% of these studies included measurement of DTT (e.g., Cummings & Carr, 2009), and another 27.3% included evaluations of mastery criteria (e.g., Richling et al.,

2019). The effects of procedural fidelity errors were evaluated in 18.1% of applications (e.g., Carroll et al., 2013). The remaining studies evaluated neural correlates of responding in DTT (Klabunde et al., 2015), treatment dose (Neil & Jones, 2015), and potentially adverse outcomes related to DTT (i.e., task interspersal arranged in DTT; Chong & Carr, 2005).

Efficacy Analysis

There were 256 participants and 1,026 applications of DTT in the 82 studies that met the inclusion criteria for the efficacy analysis. Single-subject designs² most commonly employed to evaluate DTT included multiple baselines (38.8% of studies) and alternating treatments (27.5%). Multiple probe designs were used in 7.5% of studies. Combinations of designs were used across 27.5% of studies and most commonly included combining a multiple baseline design with an alternating treatments design. The percentage of correct responses was the most common measure of skill acquisition (87.1% of analyses), followed by frequency (8.2%), trials to criterion (3.9%), and response rate (0.8%).

Mastery criteria varied widely across studies. Approximately 29% of the studies required correct responding during 100% of trials across one to five consecutive sessions. In roughly one-quarter of the studies, mastery criteria fell between 90% and 99% (26.9%) or between 80% and 89% (25.8%) for two to three consecutive sessions. The mastery criterion was set below 80% in only two studies. A fluency aim (e.g., reaching a pre-specified number of correct responses per min), rather than the percentage of correct responses, was used in 4.3% of studies. Multiple criteria were reported in 7.5% of studies. Only four studies included in the efficacy analysis failed to provide clear mastery criteria.

² Although studies incorporating group design were not purposefully excluded from the efficacy analysis, no such study met the necessary criteria for inclusion, largely because they did not provide sufficient descriptions of DTT or they did not include individual participant data on the implementation of DTT.

Efficacy for a given application of DTT was evaluated based on the mastery criteria provided in the study from which the dataset was extracted. Because no universal mastery criteria exist and the criteria employed across studies evaluated in the current review varied so greatly, we used three consecutive sessions with at least 80% correct responding to assess efficacy for the four studies that did not provide their own measure of mastery. This criterion was selected because (a) three consecutive sessions was most commonly reported across studies, (b) most mastery criteria allowed for at least one error, and (c) most studies included five to 10 trials. Thus, to allow at least one error for studies that included fewer than 10 trials, the mastery criterion was set at 80% correct. Furthermore, outcomes of the survey conducted by Richling et al. (2019) indicated that the most commonly reported mastery criterion was 80% of trials with correct responding across three consecutive sessions. Overall, DTT was efficacious in teaching new skills, with 86% of applications meeting or exceeding mastery criteria.

All reinforcer types also tended to be highly efficacious. Specifically, food, toys/activities, and tokens were efficacious in at least 90% of applications. Attention and the use of multiple, nondescript consequences were efficacious in a slightly smaller percentage of applications, 84.3% and 85.4%, respectively.

Discussion

As noted, scoping reviews can serve several important functions for practitioners and researchers: (a) they can serve as a benchmark to determine the extent to which research aligns with practice, (b) they can guide future practice, and (c) they can be used to direct more focused research and systematic reviews. We will discuss each of these in relation to the outcomes of the current study.

Alignment with Practice

The findings of the current review are encouraging; they are consistent in many ways with best practice recommendations regarding the arrangement of DTT (e.g., Ghezzi, 2007; Lerman et al., 2016; Smith, 2001). Discrete trials were initiated by the presentation of a clear antecedent delivered by the instructor, prompting was used in most cases to assist the learner when needed, consequences were largely delivered continuously and immediately, and the intertrial interval (when reported) was relatively short. Most studies used distributed reinforcement where a relatively small amount of the reinforcer (e.g., brief praise or a piece of food) was delivered immediately following each correct response; this too aligns with recommendations that reinforcement be continuous and immediate during skill acquisition.

Most antecedents included both a physical and vocal stimulus presented simultaneously; this is perhaps not surprising given the types of the responses taught. That is, by their very nature, most of the task types involve the presentation of a physical stimulus. For example, receptive identification includes matching-to-sample tasks where the participant matches a target stimulus (e.g., picture of an apple) with a corresponding sample stimulus (e.g., picture of an apple), often presented amongst an array of other non-target sample stimuli (e.g., picture of an orange). A tact is a verbal operant in which the response form is “evoked (or at least strengthened) by a particular object or event or property of an object or event” (Skinner, 1957, p. 82). For example, the speaker may say “apple” after seeing an apple sitting on a desk. Experimenters typically provided a vocal prompt at the start of the trial, perhaps to clarify expectations given that the tasks were in the acquisition phase. Exceptions to the inclusion of vocal prompts to initiate a trial largely included the use of DTT to teach intraverbals. This too again aligns well with the type of response taught; an intraverbal is a verbal response to a verbal

stimulus. Thus, a physical stimulus is not often required (unless, for example, one is responding to a written verbal stimulus).

Although attention alone was only delivered as a consequence for correct responses in slightly less than one-quarter of analyses in the current review, it was very frequently paired with other reinforcers. The prevalence of attention, both alone and combined with other reinforcers, is consistent with a survey examining practices in assessing preferences for individuals with IDD administered to 406 professionals both within and outside the field of applied behavior analysis (Graff & Karsten, 2012). Specifically, 91.5% of respondents indicated that they used praise or attention as a reinforcer. The high percentage of applications with attention also makes sense in light of the types of responses commonly taught using DTT. That is, many of the responses taught in DTT arrangements, such as tacts and intraverbals, are maintained by generalized conditioned reinforcement (i.e., attention). Thus, teaching these responses may necessitate the use of attention as a consequence.

In that same survey by Graff and Karsten (2012), researchers asked respondents to report on the types of assessments used to identify stimuli that may function as reinforcers. Roughly half (51.6%) of the respondents reported using at least one preference assessment method. In the current review, preference assessments (implemented either prior to the study or at the start of the study) were used to identify the stimuli arranged in DTT for 35.2% of applications. Graff and Karsten also found that 61% of Board Certified Behavior Analyst's reported using "mini" preference assessments, where participants were offered a choice between two or more stimuli before a session, at least once per day. In the current scoping review, we found that daily preference assessments or choices before a given session were arranged for 29.1% of DTT applications. Although these findings suggest that preference assessments (full or mini) are used

less often in the published literature than in practice, it is important to note that praise was used in many applications and formal preference assessments were not always done in these cases. In addition, the method used to identify the stimuli arranged as consequences in DTT was unknown in almost one-third of applications. It is possible that preference assessments were used but were not reported in some of the studies included in this review.

Results of the current review also differ from those obtained by Graff and Karsten (2012) in several other ways. For example, a large percentage of practitioners reported using tokens (i.e., 65.6%) in the study by Graff and Karsten; however, tokens were only included in 18.5% of DTT analyses identified in the current review. Graff and Karsten also reported that roughly half of the respondents included toys as reinforcers; results of the current review indicate that considerably fewer analyses of DTT included toys (i.e., 13.3%, although tokens were also often exchanged for toys, and it was sometimes unclear whether toys or food were used). Although these differences seem to suggest that the consequences arranged in research on DTT may not always align well with the consequences reportedly arranged in practice, it is important to note that the survey conducted by Graff and Karsten focused on the general use of stimuli as reinforcers and not their use in DTT specifically. This difference in focus may at least partially account for the discrepant findings across our study and theirs.

Although the use of preference assessments and the types of stimuli arranged as reinforcers in the current review do seem to differ from practice, other outcomes described in this scoping review appear to align well with expectations regarding DTT arrangements for acquisition of new skills. For example, reversal designs were rarely used; when they were, reversals were used in combination with a multiple baseline design. It is not surprising that reversal designs were not used in isolation because demonstration of experimental control would

require the acquired response to decrease upon the removal of the intervention, which is unlikely to occur with learned skills. In addition, the studies included in this review rarely used response rate to measure the dependent variable likely because completion of the discrete trials is controlled by the opportunity to emit the response, which is, in turn, controlled by the practitioner. Responding in DTT is opportunity bound and, as such, measures such as percentage of opportunities are more appropriate. Finally, the findings of the current review regarding how mastery was measured (e.g., as percentage of trials correct or trials to criterion) and the variability in mastery criteria used are consistent with those reported by Richling et al. (2019) in their survey of practitioner practices related to mastery criteria. Richling et al. observed that most respondents reported measuring mastery based on the percentage of trials (68%), with relatively few indicating they used response rate (4%). Furthermore, mastery criteria varied widely (as was also observed in this review) but was generally set to at least 80% correct for one or more consecutive sessions. This is again consistent with the findings of this scoping review.

Guidance for Practice

Discrete trial teaching is a commonly used method of instruction. As such, it is important to know how and when DTT is most frequently arranged as this information can be used in preparing educational programming. Specifically, the outcomes of this review provide information on (a) the population with whom DTT has been used, (b) how to arrange DTT components, and (c) the types of responses for which DTT may be most appropriate given the current state of research on DTT. This information may be useful in guiding practice.

Discrete trial teaching is a common component of EIBI programs for young children with autism spectrum disorder. However, DTT need not be restricted to this population. The results of this review also support the use of DTT for individuals of varying ages and with a wide variety

of diagnoses. Just over 17% of the participants included in this review were adolescents or adults and approximately 23% did not have a reported autism spectrum diagnosis. Although DTT is most commonly used in clinics/centers or schools, this scoping review also demonstrated that DTT can be implemented in the home or residential placements and in hospital settings.

Furthermore, DTT can be implemented by a wide variety of individuals, including researchers (experimenters), teachers, peers, and caregivers, as well as electronically. This is consistent with a recent scoping review of procedures used to train individuals to implement DTT (Briggs et al., 2023). Briggs et al. (2023) found that (when appropriate data were reported) behavior skills training was used successfully to teach DTT to individuals across a wide range of ages (i.e., 18–70 years) and with different backgrounds (e.g., students and clinicians). Thus, practitioners should feel comfortable implementing DTT with a diverse clientele, across various settings, and with a variety of implementers.

The arrangement of DTT components will likely depend on a variety of factors, including practices common to the setting in which it is being conducted and prior information gathered about an individual learner. In some cases, though, practitioners may find themselves having to develop DTT programs without such structure from their institution or for a new learner for whom they have limited information. Results of this scoping review may provide guidance in those situations by outlining the most commonly assessed components. Specifically, the results indicate that DTT is most often arranged using 6–15 trials, and trials are separated by brief intertrial intervals, typically 1 to 5 s. The trials most commonly begin with a vocal stimulus as part of the antecedent, and errors are corrected using either a single corrective prompt or a prompt hierarchy. The consequence for a correct response is typically delivered on a fixed ratio 1 schedule and arranged using distributed reinforcement. And a prompt delay is used in about half

of applications. As a practitioner, one may consider this information when designing a DTT program for a new learner until more information is gathered and programming can be tailored to the individual learner's needs.

Finally, practitioners could also use the information in this review to guide their decision-making as to when to use DTT (relative to another type of instructional programming) for teaching specific responses. Results suggest that DTT has been most commonly assessed to teach receptive identification, tacts, and intraverbals; DTT was less commonly used when teaching mands, play, transcription, echoics, and imitation. These data do not suggest that DTT cannot or should not be used to teach these latter responses; they simply indicate that we have more research on the use of DTT with receptive identification, tacts, and intraverbals. Thus, practitioners should feel comfortable using DTT with receptive identification, tacts, and intraverbals. These findings also suggest we need more research on conditions under which DTT may be preferable to other instructional strategies, which we will discuss next.

Guidance for Future Research

Outcomes of this scoping review may help to guide future research in a number of ways. First, like the guidance this review may provide for practitioners, outcomes of this review could also direct the design of research protocols that include DTT. If researchers plan to incorporate DTT and wish to model the components of DTT after commonly employed arrangements, they could use the outcome of this scoping review as guidance. This may be particularly useful for research that aims to compare DTT to another instructional strategy. It would be wise for such researchers to arrange DTT in a manner that aligns both with best practice recommendations and its most frequent use in the literature; this scoping review could guide researchers in the latter.

Second, the outcomes of this review could be useful in directing the topics of future research. Such research could focus on comparisons of DTT to other instructional strategies, evaluations of different components of the DTT arrangement, and syntheses of the literature on a particular aspect of DTT. As noted, DTT has been most commonly used to teach receptive identification, tacts, and intraverbals and has been less commonly used with other types of responses such as mands and echoics. It is not clear whether or under what conditions other approaches would prove more efficacious than DTT for teaching certain responses. Some research exists in which DTT has been compared to other instructional strategies. For example, Jennett et al. (2008) compared the effects of DTT and mand training to teach children with autism spectrum disorder to make requests. Results indicated that most participants made more independent requests and acquired requesting faster in the mand training condition. Furthermore, although DTT has been very commonly used with tacts, some comparative research has suggested that other instructional strategies, like fluency training, may result in better maintenance after acquisition (e.g., Holding et al., 2011). Other comparative studies exist (e.g., Charlop-Christy, & Carpenter, 2000; Nopprapun & Holloway, J., 2014), however research in this area is relatively limited as compared to that focusing on comparisons of DTT components. Thus, this line of study is certainly one that warrants additional consideration by researchers.

Figure 4 depicts the cumulative number of each study type included in this review across the years. Studies comparing components of DTT accelerated more quickly than those that applied DTT to a new target, especially in the last 6 to 7 years. These data suggest a shift in focus away from novel applications of DTT to evaluations of how best to arrange DTT. Acceleration of this line of research presents opportunities for researchers to conduct more in-depth, quantitative evaluations of the efficacy of DTT across component variations than the more

cursory glimpse into DTT efficacy provided in this review. For instance, the results of the efficacy analysis of this scoping review indicated that DTT was efficacious in 86% of applications. Although encouraging, these data should be interpreted with caution for several reasons. Specifically, it is not clear whether prior history with DTT may have had an effect on the overall outcomes as history with DTT was not reported for most participants. In addition, many studies used mastery criteria that may not be considered very stringent or that may not produce long-lasting change. Further, very few studies evaluated the generality of the findings from the DTT applications. Thus, additional studies are warranted for which (a) more stringent mastery criteria are applied; (b) generality of outcomes are assessed across different caregivers, settings, and target skills; and (c) maintenance of the skills is more thoroughly assessed.

Regarding mastery criteria, Fienup and Carr (2021) and Richling et al. (2019) both noted the wide variety of mastery criteria used in behavior-analytic research. However, they also indicated that there is a lack of research on mastery criteria as the independent variable for skill acquisition. Results of our scoping review confirm this as we identified only three studies that met our inclusion criteria where the focus was on mastery criteria as the independent variable (see Table 5). Thus, this is an area ripe for further research particularly as it may relate to maintenance of skills. For example, Richling et al. conducted a series of studies examining the impact of differing levels of mastery criteria on maintenance of the learned skill. Across the studies, better maintenance was observed when the mastery criterion was relatively stringent, 100% correct responding across three consecutive sessions. Additional research is warranted to evaluate the efficacy of differing mastery criteria.

Other aspects of DTT arrangements also vary considerably and could be targets of further research and review. For example, there are numerous prompting strategies used in DTT to

improve learner accuracy. It is unclear whether any particular strategy is more efficacious than others or whether certain strategies may be better for specific responses. Future research could compare the efficacy of the most commonly applied prompting strategies (e.g., single corrective prompt and prompt hierarchy) across a variety of types of responses for the same participants. Systematic reviews of these common prompting strategies may also prove particularly informative. The number of trials included in a given session also varied quite a bit across DTT applications. Although repetition is a key aspect of DTT, the ideal number of trials required to achieve mastery is unknown.

Response-reinforcer arrangements in DTT could also be further explored. Weinsztok et al. (2023) conducted a systematic review of reinforcement parameters arranged in DTT and found that delivery of tangible or food reinforcers generally resulted in more efficient skill acquisition than contingent attention alone. We similarly found a higher percentage of efficacious applications of DTT when they included tangibles, tokens, and food relative to attention alone. Furthermore, Weinsztok et al. also found that accumulated-reinforcement arrangements were associated with more efficient acquisition relative to distributed arrangements in most evaluations. Variables contributing the general superiority of accumulated reinforcement are unknown; thus, future research may consider examining variables that predict preference for and efficacy of distributed and accumulated response-reinforcer arrangements in skill acquisition. Ward-Horner et al. (2017) suggested several variables that may predict preference and performance under these arrangements. For example, it is possible that individuals with attention-deficit hyperactivity disorder (who, by definition, behave more impulsively) or individuals with escape-maintained problem behavior (for which more immediate access to the functional reinforcer, a break, is a valuable consequence) may prefer and perform better under

distributed reinforcement arrangements (Chen et al., 2022; Frank-Crawford et al., 2021; Fulton et al., 2020; Robinson & St. Peter, 2019). It is also possible that reinforcer type matters; that is, reinforcer accumulation may be more valuable with activity reinforcers and less so with food (e.g., DeLeon et al., 2014). More research on preference and efficacy outcomes with the same individuals will also allow researchers to directly examine how well preference predicts efficacy. This is an important avenue for research as studies have not always demonstrated correspondence between preference and efficacy outcomes (e.g., Leaf et al., 2010; Slocum & Tiger, 2011; Winborn et al., 2002), and research on response-reinforcer arrangements is no exception (e.g., Frank-Crawford et al., 2019; Joachim & Carroll, 2018; Kocher et al., 2015). These are but a few examples of more direct lines of systematic research; others certainly exist and should also be explored.

On a separate note, but related to future research, Baer et al. (1968) called for the technological description of the procedures used in studies. Technological descriptions include identifying and labeling procedures as well as thoroughly describing the relevant features of the procedure. Noteworthy in the current review is that 31.3% of studies did not describe the teaching arrangement as DTT (or some variant thereof). Multiple studies also failed to explicitly describe person variables (e.g., age, level of intellectual disability, or setting), specific reinforcers used across sessions, and intertrial interval durations. As noted in other reviews of behavior analytic procedures (e.g., Briggs et al., 2023; Jones et al., 2020; Rooker et al., 2018), failure to adequately describe demographic variables or procedures employed in each study can impact the quality of the literature and may make it difficult to speak to the generality of procedures across a given line of research (e.g., Walker & Carr, 2021). It is therefore recommended that authors provide more technologically precise descriptions of study procedures and that journal editorial

practice provide the space and contexts for technologically richer descriptions. Results of Figure 2 do suggest improvements in overall reporting on the use of DTT in studies, which is promising.

Limitations and Conclusions

The current review is not without limitations. Although multiple search methods were used to identify studies that met inclusion criteria, it is probable that some studies were missed. Furthermore, studies that did not adequately describe each component of DTT were excluded from the review. Thus, it is unclear to what extent data from any missing studies or from those we excluded because they did not describe all of the DTT components would have impacted our overall findings. In addition, the findings showed that DTT was generally efficacious in promoting skill acquisition. However, we used a relatively elementary measure of efficacy; that is, a dichotomous “yes” or “no” was recorded based on whether the DTT application met the mastery criteria reported on in the respective studies. Thus, it is unclear whether a more sensitive analysis, such as measures of effect sizes or using a standardized mastery criterion across all studies, would yield similar results.

It is perhaps not surprising that most studies reported at least one positive application of DTT given the tendency for studies with more favorable outcomes to be both submitted and accepted for publication. Such publication bias has been demonstrated to occur in single subject research (e.g., Shadish et al., 2016; Sham & Smith, 2014; Tincani & Travers, 2019). The current scoping review attempted to minimize such bias by also including unpublished dissertations and studies that directly compared DTT to other instructional strategies (as these studies were more likely to include negative findings); however, most studies included in this review were peer reviewed publications (98.5%; $n = 132$ of the 134 studies).

Finally, this study was merely intended to provide an overview of the literature on DTT, to generally describe the current state of research and provide a brief summary of the efficacy of DTT. Thus, there are a number of important questions that we cannot answer in this scoping review. For example, we cannot speak to how well DTT fares relative to other modes of instruction (e.g., such as naturalistic instruction), and we cannot evaluate the efficacy of DTT using a standardized mastery criterion (instead of the study-specific criteria used here). Such investigations would be more appropriate for systematic reviews and meta-analyses, which we do hope can be inspired by some of the findings in this scoping review.

Results of the current review support the findings of other, more broadly focused reviews on comprehensive EIBI programs (e.g., Virués-Ortega, 2010), suggesting that DTT is an efficacious instructional strategy to teach a variety of skills to learners with IDD. The current review has also documented that although many studies attempted to evaluate the maintenance of skills learned through DTT, data on the generality of these outcomes were more limited. Furthermore, the assessment of the degree to which stakeholders and/or the recipients of behavior-change procedures themselves found DTT to be an appropriate and efficacious teaching procedure was largely lacking. Although the results of this review suggest that, in general, DTT is an efficacious teaching strategy, data on outcomes from these additional variables would be required to speak more to its effectiveness (e.g., Tolin et al., 2015). Thus, continued evaluations of the maintenance, generality, and social validity of the outcomes of DTT represent important avenues for future research, particularly in pursuit of evidence of the effectiveness of this teaching procedure.

References

- Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 1(1), 91–97.
<https://doi.org/10.1901/jaba.1968.1-91>
- Baio, J., Wiggins, L., Christensen, D. L., Maenner, M. J., Daniels, J., Warren, Z., Kurzius-Spencer, M., Zahorodny, W., Rosenberg, C. R., White, T., Durkin, M. S., Imm, P., Nikolaou, L., Yeargin-Allsopp, M., Lee, L., Harrington, R., Lopez, M., Fitzgerald, R. T., Hewitt, A., ... & Durkin, M. S. (2018). Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2014. *MMWR Surveillance Summaries*, 67(6), 1–23.
<https://dx.doi.org/10.15585/mmwr.ss6706a1>
- Belisle, J., Dixon, M. R., Alholai, A., Ellenberger, L., Stanley, C., & Galliford, M. (2020). Teaching children with autism to tact the private events of others. *Behavior Analysis in Practice*, 13(1), 169–173. <https://doi.org/10.1007/s40617-019-00334-9>
- Briggs, A. M., Zohr, S. J., & Harvey, O. B. (2023). Training individuals to implement discrete-trial teaching procedures using behavioral skills training: A scoping review with implications for practice and research. *Journal of Applied Behavior Analysis*, 57(1), 86–103. <https://doi.org/10.1002/jaba.1024>
- Cariveau, T., Kodak, T., & Campbell, V. (2016). The effects of intertrial interval and instructional format on skill acquisition and maintenance for children with autism spectrum disorders. *Journal of Applied Behavior Analysis*, 49(4), 809–825.
<https://doi.org/10.1002/jaba.322>

- Carr, J. E., Austin, J. L., Britton, L. N., Kellum, K. K., & Bailey, J. S. (1999). An assessment of social validity trends in applied behavior analysis. *Behavioral Interventions*, 14(4), 223–231. [https://doi.org/10.1002/\(SICI\)1099-078X\(199910/12\)14:4%3C223::AID-BIN37%3E3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1099-078X(199910/12)14:4%3C223::AID-BIN37%3E3.0.CO;2-Y)
- Carroll, R. A., Joachim, B. T., St Peter, C. C., & Robinson, N. (2015). A comparison of error-correction procedures on skill acquisition during discrete-trial instruction. *Journal of Applied Behavior Analysis*, 48(2), 257–273. <https://doi.org/10.1002/jaba.205>
- Carroll, R. A., Kodak, T., & Adolf, K. J. (2016). Effect of delayed reinforcement on skill acquisition during discrete-trial instruction: Implications for treatment-integrity errors in academic settings. *Journal of Applied Behavior Analysis*, 49(1), 176–181. <https://doi.org/10.1002/jaba.268>
- Carroll, R. A., Kodak, T., & Fisher, W. W. (2013). An evaluation of programmed treatment-integrity errors during discrete-trial instruction. *Journal of Applied Behavior Analysis*, 46(2), 379–394. <https://doi.org/10.1002/jaba.49>
- Carroll, R. A., Owsiany, J., & Cheatham, J. M. (2018). Using an abbreviated assessment to identify effective error-correction procedures for individual learners during discrete-trial instruction. *Journal of Applied Behavior Analysis*, 51(3), 482–501. <https://doi.org/10.1002/jaba.460>
- Cengher, M., Shamoun, K., Moss, P., Roll, D., Feliciano, G., & Fienup, D. M. (2016). A comparison of the effects of two prompt-fading strategies on skill acquisition in children with autism spectrum disorders. *Behavior Analysis in Practice*, 9(2), 115–125. <https://doi.org/10.1007/s40617-015-0096-6>

- Charlop-Christy, M. H., & Carpenter, M. H. (2000). Modified incidental teaching sessions: A procedure for parents to increase spontaneous speech in their children with autism. *Journal of Positive Behavior Interventions*, 2(2), 98–112. <https://doi.org/10.1177%2F109830070000200203>
- Chen, E., Borrero, C. S., Frank-Crawford, M. A., & Borrero, J. C. (2022). Accumulated-and distributed-reinforcer arrangements in the treatment of challenging mealtime behavior. *Behavioral Interventions*, 37(4), 1058–1079. <https://doi.org/10.1002/bin.1899>
- Chong, I. M., & Carr, J. E. (2005). An investigation of the potentially adverse effects of task interspersal. *Behavioral Interventions*, 20(4), 285–300. <https://doi.org/10.1002/bin.202>
- Coe, D., Matson, J., Fee, V., Manikam, R., & Linarello, C. (1990). Training nonverbal and verbal play skills to mentally retarded and autistic children. *Journal of Autism and Developmental Disorders*, 20(2), 177–187. <https://doi.org/10.1007/BF02284717>
- Cummings, A. R., & Carr, J. E. (2009). Evaluating progress in behavioral programs for children with autism spectrum disorders via continuous and discontinuous measurement. *Journal of Applied Behavior Analysis*, 42(1), 57–71. <https://doi.org/10.1901/jaba.2009.42-57>
- das Neves, A. J. D., Almeida-Verdu, A. C. M., Assis, G. J. A. D., Silva, L. T. D. N., & Moret, A. L. M. (2018). Improving oral sentence production in children with cochlear implants: effects of equivalence-based instruction and matrix training. *Psicologia: Reflexão e Crítica*, 31, 1–17. <https://doi.org/10.1186/s41155-018-0095-y>
- DeLeon, I. G., Chase, J. A., Frank-Crawford, M. A., Carreau-Webster, A. B., Triggs, M. M., Bullock, C. E., & Jennett, H. K. (2014). Distributed and accumulated reinforcement arrangements: Evaluations of efficacy and preference. *Journal of Applied Behavior Analysis*, 47(2), 293–313. <https://doi.org/10.1002/jaba.116>

- Dixon, M. R., Belisle, J., Munoz, B. E., Stanley, C. R., & Rowsey, K. E. (2017). Teaching metaphorical extensions of private events through rival-model observation to children with autism. *Journal of Applied Behavior Analysis*, 50(4), 744–749.
<https://doi.org/10.1002/jaba.418>
- Essig, L., Rotta, K., & Poling, A. (2023). Interobserver agreement and procedural fidelity: An odd asymmetry. *Journal of Applied Behavior Analysis*, 56(1), 78–85.
<https://doi.org/10.1002/jaba.961>
- Ferguson, J. L., Majeski, M. J., McEachin, J., Leaf, R., Cihon, J. H., & Leaf, J. B. (2020). Evaluating discrete trial teaching with instructive feedback delivered in a dyad arrangement via telehealth. *Journal of Applied Behavior Analysis*, 53(4), 1876–1888. <https://doi.org/10.1002/jaba.773>
- Fienup, D. M., & Carr, J. E. (2021). The use of performance criteria for determining “mastery” in discrete-trial instruction: A call for research. *Behavioral Interventions*, 36(4), 756–763.
<https://doi.org/10.1002/bin.1827>
- Frank-Crawford, M. A., Borrero, J. C., Newcomb, E. T., Chen, T., & Schmidt, J. D. (2019). Preference for and efficacy of accumulated and distributed response–reinforcer arrangements during skill acquisition. *Journal of Behavioral Education*, 28(2), 227–257.
<https://doi.org/10.1007/s10864-018-09312-7>
- Frank-Crawford, M. A., Borrero, J. C., Newcomb, E. T., Doan, T., Fisher, A., & Rooker, G. W. (2021). Accumulated and distributed response–reinforcer arrangements during the treatment of escape-maintained problem behavior. *Journal of Applied Behavior Analysis*, 54(4), 1566–1585. <https://doi.org/10.1002/jaba.870>

- Fryling, M. J., Wallace, M. D., & Yassine, J. N. (2012). Impact of treatment integrity on intervention effectiveness. *Journal of Applied Behavior Analysis*, 45(2), 449–45. <https://doi.org/10.1901/jaba.2012.45-449>
- Fulton, C. J., Tiger, J. H., Meitzen, H. M., & Effertz, H. M. (2020). A comparison of accumulated and distributed reinforcement periods with children exhibiting escape-maintained problem behavior. *Journal of Applied Behavior Analysis*, 53(2), 782–795. <https://doi.org/10.1002/jaba.622>
- Geiger, K. B., Carr, J. E., LeBlanc, L. A., Hanney, N. M., Polick, A. S., & Heinicke, M. R. (2012). Teaching receptive discriminations to children with autism: A comparison of traditional and embedded discrete trial teaching. *Behavior Analysis in Practice*, 5(2), 49–59. <https://doi.org/10.1007/BF03391823>
- Ghezzi, P. M. (2007). Discrete trials teaching. *Psychology in the Schools*, 44(4), 667–679. <https://doi.org/10.1002/pits.20256>
- Graff, R. B., & Karsten, A. M. (2012). Assessing preferences of individuals with developmental disabilities: A survey of current practices. *Behavior Analysis in Practice*, 5(5), 37–48. <https://doi.org/10.1007/BF03391822>
- Grow, L. L., Kodak, T., & Carr, J. E. (2014). A comparison of methods for teaching receptive labeling to children with autism spectrum disorders: A systematic replication. *Journal of Applied Behavior Analysis*, 47(3), 600–605. <https://doi.org/10.1002/jaba.141>
- Grow, L. L., Kodak, T., & Clements, A. (2016). An evaluation of instructive feedback to teach play behavior to a child with autism spectrum disorder. *Behavior Analysis in Practice*, 10(3), 1–5. <https://doi.org/10.1007/s40617-016-0153-9>

- Holding, E., Bray, M. A., & Kehle, T. J. (2011). Does speed matter? A comparison of the effectiveness of fluency and discrete trial training for teaching noun labels to children with autism. *Psychology in the Schools*, 48(2), 166–183.
<https://doi.org/10.1002/pits.20535>
- Jahr, E., Eldevik, S., & Eikeseth, S. (2000). Teaching children with autism to initiate and sustain cooperative play. *Research in Developmental Disabilities*, 21(2), 151–169.
[https://doi.org/10.1016/S0891-4222\(00\)00031-7](https://doi.org/10.1016/S0891-4222(00)00031-7)
- Jameson, J. M., McDonnell, J., Johnson, J. W., Riesen, T., & Polychronis, S. (2007). A comparison of one-to-one embedded instruction in the general education classroom and one-to-one massed practice instruction in the special education classroom. *Education and Treatment of Children*, 30(1), 23-44. <https://www.jstor.org/stable/42899917>
- Jennett, H. K., Harris, S. L., & Delmolino, L. (2008). Discrete trial instruction vs. mand training for teaching children with autism to make requests. *The Analysis of Verbal Behavior*, 24(1), 69–85. <https://doi.org/10.1007/BF03393058>
- Joachim, B. T., & Carroll, R. A. (2018). A comparison of consequences for correct responses during discrete-trial instruction. *Learning and Motivation*, 62, 15–28.
<https://doi.org/10.1016/j.lmot.2017.01.002>
- Jones, S. H., St. Peter, C. C., & Ruckle, M. M. (2020). Reporting of demographic variables in the *Journal of Applied Behavior Analysis*. *Journal of Applied Behavior Analysis*, 53(3), 1304–1315. <https://doi.org/10.1002/jaba.722>
- Kennedy, C. H. (1992). Trends in the measurement of social validity. *The Behavior Analyst*, 15(2), 147–156. <https://doi.org/10.1007/BF03392597>

Klabunde, M., Saggar, M., Hustyi, K. M., Kelley, R. G., Reiss, A. L., & Hall, S. S. (2015).

Examining the neural correlates of emergent equivalence relations in fragile X syndrome.

Psychiatry Research: Neuroimaging, 233(3), 373–379.

<https://doi.org/10.1016/j.psychresns.2015.06.009>

Kocher, C. P., Howard, M. R., & Fienup, D. M. (2015). The effects of work-reinforcer schedules on skill acquisition for children with autism. *Behavior Modification*, 39(4), 600–621.

<https://doi.org/10.1177%2F0145445515583246>

Koegel, R. L., O'dell, M. C., & Koegel, L. K. (1987). A natural language teaching paradigm for nonverbal autistic children. *Journal of Autism and Developmental Disorders*, 17(2), 187–200. <https://doi.org/10.1007/BF01495055>

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). Single-case designs technical documentation. *What works clearinghouse*. <https://eric.ed.gov/?id=ED510743>

Lambert, J. M., Copeland, B. A., Karp, E. L., Finley, C. I., Houchins-Juarez, N. J., & Ledford, J. R. (2016). Chaining functional basketball sequences (with embedded conditional discriminations) in an adolescent with autism. *Behavior Analysis in Practice*, 9(3), 199–210. <https://doi.org/10.1007/s40617-016-0125-0>

Lang, R., van der Werff, M., Verbeek, K., Didden, R., Davenport, K., Moore, M., Lee, A., Rispoli, M., Machalicek, W., O'Reilly, M., Sigafos, J., & Lancioni, G. (2014). Comparison of high and low preferred topographies of contingent attention during discrete trial training. *Research in Autism Spectrum Disorders*, 8(10), 1279–1286. <https://doi.org/10.1016/j.rasd.2014.06.012>

- Leaf, J. B., Sheldon, J. B., & Sherman, J. A. (2010). Comparison of simultaneous prompting and no-no prompting in two-choice discrimination learning with children with autism. *Journal of Applied Behavior Analysis*, 43(2), 215–228.
<https://doi.org/10.1901/jaba.2010.43-215>
- Leaf, R., & McEachin, J. (1999). *A work in progress: Behavior management strategies and a curriculum for intensive behavioral treatment of autism*. DRL Books.
- LeBlanc, L. A., Esch, J., Sidener, T. M., & Firth, A. M. (2006). Behavioral language interventions for children with autism: Comparing applied verbal behavior and naturalistic teaching approaches. *The Analysis of Verbal Behavior*, 22(1), 49–60.
<https://doi.org/10.1007/BF03393026>
- Lerman, D. C., Valentino, A. L., & LeBlanc, L. A. (2016). Discrete trial training. In R. Lang, T. B. Hancock, & N. N. Singh (Eds.), *Early intervention for young children with autism spectrum disorder* (pp. 47–83). Springer International Publishing.
<https://doi.org/10.1007/978-3-319-30925-5>
- Lindsley, O. R. (1996). The four free-operant freedoms. *The Behavior Analyst*, 19(2), 199–210.
<https://doi.org/10.1007/BF03393164>
- Lovaas, I. O. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology*, 55(1), 3–9.
<https://doi.org/10.1037/0022-006X.55.1.3>
- Love, J. R., Carr, J. E., Almason, S. M., & Petursdottir, A. I. (2009). Early and intensive behavioral intervention for autism: A survey of clinical practices. *Research in Autism Spectrum Disorders*, 3(2), 421–428. <https://doi.org/10.1016/j.rasd.2008.08.008>

- MacDonald, J., & Ahearn, W. H. (2015). Teaching observational learning to children with autism. *Journal of Applied Behavior Analysis*, 48(4), 800–816.
<https://doi.org/10.1002/jaba.257>
- Majdalany, L., Wilder, D. A., Smeltz, L., & Lipschultz, J. (2016). The effect of brief delays to reinforcement on the acquisition of tacts in children with autism. *Journal of Applied Behavior Analysis*, 49(2), 411–415. <https://doi.org/10.1002/jaba.282>
- Majdalany, L. M., Wilder, D. A., Greif, A., Mathisen, D., & Saini, V. (2014). Comparing massed-trial instruction, distributed-trial instruction, and task interspersal to teach tacts to children with autism spectrum disorders. *Journal of Applied Behavior Analysis*, 47(3), 657–662. <https://doi.org/10.1002/jaba.149>
- Munn, Z., Peters, M. D., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, 18, 143.
<https://doi.org/10.1186/s12874-018-0611-x>
- Neil, N., Hansford, R., Young, K., & Zwick, L. (2020). The effect of instructional pacing on skill acquisition and maintenance for children with developmental disabilities. *Journal on Developmental Disabilities*, 25(1), 1–15.
- Neil, N., & Jones, E. A. (2015). Studying treatment intensity: Lessons from two preliminary studies. *Journal of Behavioral Education*, 24(1), 51–73. <https://doi.org/10.1007/s10864-014-9208-6>
- Nopprapun, M., & Holloway, J. (2014). A comparison of fluency training and discrete trial instruction to teach letter sounds to children with ASD: Acquisition and learning

- outcomes. *Research in Autism Spectrum Disorders*, 8(7), 788–802.
<https://doi.org/10.1016/j.rasd.2014.03.015>
- Paden, A. R., & Kodak, T. (2015). The effects of reinforcement magnitude on skill acquisition for children with autism. *Journal of Applied Behavior Analysis*, 48(4), 924–929.
<https://doi.org/10.1002/jaba.239>
- Parker, M. A., & Schuster, J. W. (2002). Effectiveness of simultaneous prompting on the acquisition of observational and instructive feedback stimuli when teaching a heterogeneous group of high school students. *Education and Training in Mental Retardation and Developmental Disabilities*, 37(1), 89–104.
<https://www.jstor.org/stable/23879585>
- Peters-Scheffer, N., Didden, R., Korzilius, H., & Sturmey, P. (2011). A meta-analytic study on the effectiveness of comprehensive ABA-based early intervention programs for children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, 5(1), 60–69.
<https://doi.org/10.1016/j.rasd.2010.03.011>
- Petursdottir, A. I., & Aguilar, G. (2016). Order of stimulus presentation influences children's acquisition in receptive identification tasks. *Journal of Applied Behavior Analysis*, 49(1), 58–68. <https://doi.org/10.1002/jaba.264>
- Petursdottir, A. I., Carr, J. E., & Michael, J. (2005). Emergence of mands and tacts of novel objects among preschool children. *The Analysis of Verbal Behavior*, 21(1), 59–74.
<https://doi.org/10.1007/BF03393010>
- Plaisance, L., Lerman, D. C., Laudont, C., & Wu, W. L. (2016). Inserting mastered targets during error correction when teaching skills to children with autism. *Journal of Applied Behavior Analysis*, 49(2), 251–264. <https://doi.org/10.1002/jaba.292>

- Radley, K. C., Dart, E. H., Furlow, C. M., & Ness, E. J. (2015). Peer-mediated discrete trial training within a school setting. *Research in Autism Spectrum Disorders*, 9, 53–67. <https://doi.org/10.1016/j.rasd.2014.10.001>
- Reichow, B. (2012). Overview of meta-analyses on early intensive behavioral intervention for young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 42(4), 512–520. <https://doi.org/10.1007/s10803-011-1218-9>
- Reichow, B., & Wolery, M. (2011). Comparison of progressive prompt delay with and without instructive feedback. *Journal of Applied Behavior Analysis*, 44(2), 327–340. <https://doi.org/10.1901/jaba.2011.44-327>
- Richling, S. M., Williams, W. L., & Carr, J. E. (2019). The effects of different mastery criteria on the skill maintenance of children with developmental disabilities. *Journal of Applied Behavior Analysis*, 52(3), 701–717. <https://doi.org/10.1002/jaba.580>
- Robinson, N., & St. Peter, C. C. (2019). Accumulated reinforcers increase academic responding and suppress problem behavior for students with attention-deficit hyperactivity disorder. *Journal of Applied Behavior Analysis*, 52(4), 1076–1088. <https://doi.org/10.1002/jaba.570>
- Rooker, G. W., Bonner, A. C., Dillon, C. M., & Zarcone, J. R. (2018). Behavioral treatment of automatically reinforced SIB: 1982–2015. *Journal of Applied Behavior Analysis*, 51(4), 974–997. <https://doi.org/10.1002/jaba.492>
- Schmidt, J. D., Drasgow, E., Halle, J. W., Martin, C. A., & Bliss, S. A. (2014). Discrete-trial functional analysis and functional communication training with three individuals with autism and severe problem behavior. *Journal of Positive Behavior Interventions*, 16(1), 44–55. <https://doi.org/10.1177/1098300712470519>

- Shadish, W. R., Zelinsky, N. A., Vevea, J. L., & Kratochwill, T. R. (2016). A survey of publication practices of single-case design researchers when treatments have small or large effects. *Journal of Applied Behavior Analysis, 49*(3), 656–673.
<https://doi.org/10.1002/jaba.308>
- Sham, E., & Smith, T. (2014). Publication bias in studies of an applied behavior-analytic intervention: An initial analysis. *Journal of Applied Behavior Analysis, 47*(3), 663–678.
<https://doi.org/10.1002/jaba.146>
- Skinner, B. F. (1957). *Verbal behavior*. Prentice Hall.
- Slocum, S. K., & Tiger, J. H. (2011). An assessment of the efficiency of and child preference for forward and backward chaining. *Journal of Applied Behavior Analysis, 44*(4), 793–805.
<https://doi.org/10.1901/jaba.2011.44-793>
- Smith, T. (2001). Discrete trial training in the treatment of autism. *Focus on Autism and Other Developmental Disabilities, 16*(2), 86–92.
<https://doi.org/10.1177%2F108835760101600204>
- Snodgrass, M. R., Chung, M. Y., Meadan, H., & Halle, J. W. (2018). Social validity in single-case research: A systematic literature review of prevalence and application. *Research in Developmental Disabilities, 74*, 160–173. <https://doi.org/10.1016/j.ridd.2018.01.007>
- St. Peter, C. C., Brand, D., Jones, S. H., Wolgemuth, J. R., & Lipien, L. (2023). On a persisting curious double standard in behavior analysis: Behavioral scholars' perspectives on procedural fidelity. *Journal of Applied Behavior Analysis, 56*(2), 336–351.
<https://doi.org/10.1002/jaba.974>
- Stewart, L. A., Clarke, M., Rovers, M., Riley, R. D., Simmonds, M., Stewart, G., & Tierney, J. F. (2015). Preferred reporting items for a systematic review and meta-analysis of individual

- participant data: the PRISMA-IPD statement. *JAMA*, 313(16), 1657–1665.
<https://doi.org/10.1001/jama.2015.3656>
- Sy, J. R., & Vollmer, T. R. (2012). Discrimination acquisition in children with developmental disabilities under immediate and delayed reinforcement. *Journal of Applied Behavior Analysis*, 45(5), 667–684. <https://doi.org/10.1901/jaba.2012.45-667>
- Tarbox, R. S. F., Ghezzi, P. M., & Wilson, G. (2006). The effects of token reinforcement on attending in a young child with autism. *Behavioral Interventions*, 21(3), 155–164.
<https://doi.org/10.1002/bin.213>
- Tarbox, R. S. F., & Najdowski, A. C. (2008). Discrete trial training as a teaching paradigm. In J. K. Luiselli, D. C. Russo, W. P. Christian & S. M. Wilczynski (Eds.), *Effective practices for children with autism: Educational and behavioral support interventions that work* (pp. 325–389). Oxford University Press.
- Tincani, M., & Travers, J. (2019). Replication research, publication bias, and applied behavior analysis. *Perspectives on Behavior Science*, 42(1), 59–75.
<https://doi.org/10.1007/s40614-019-00191-5>
- Tolin, D. F., McKay, D., Forman, E. M., Klonsky, E. D., & Thombs, B. D. (2015). Empirically supported treatment: Recommendations for a new model. *Clinical Psychology: Science and Practice*, 22(4), 317–338. <https://doi.org/10.1111/cpsp.12122>
- Virués-Ortega, J. (2010). Applied behavior analytic intervention for autism in early childhood: Meta-analysis, meta-regression and dose–response meta-analysis of multiple outcomes. *Clinical Psychology Review*, 30(4), 387–399. <https://doi.org/10.1016/j.cpr.2010.01.008>
- Volkert, V. M., Lerman, D. C., Trosclair, N., Addison, L., & Kodak, T. (2008). An exploratory analysis of task-interspersal procedures while teaching object labels to children with

- autism. *Journal of Applied Behavior Analysis*, 41(3), 335–350.
<https://doi.org/10.1901/jaba.2008.41-335>
- Walker, S. G., & Carr, J. E. (2021). Generality of findings from single-case designs: It's not all about the "N." *Behavior Analysis in Practice*, 14(4), 991–995.
<https://doi.org/10.1007/s40617-020-00547-3>
- Ward-Horner, J. C., Cengher, M., Ross, R. K., & Fienup, D. M. (2017). Arranging response requirements and the distribution of reinforcers: A brief review of preference and performance outcomes. *Journal of Applied Behavior Analysis*, 50(1), 181–185.
<https://doi.org/10.1002/jaba.350>
- Weinsztok, S. C., Goldman, K. J., & DeLeon, I. G. (2023). Assessing parameters of reinforcement on efficiency of acquisition: A systematic review. *Behavior Analysis in Practice*, 16(1), 76–92. <https://doi.org/10.1007/s40617-022-00715-7>
- Winborn, L., Wacker, D. P., Richman, D. M., Asmus, J., & Geier, D. (2002). Assessment of mand selection for functional communication training packages. *Journal of Applied Behavior Analysis*, 35(3), 295–298. <https://doi.org/10.1901/jaba.2002.35-295>
- Wolf, M. M., Risley, T. R., & Mees, H. L. (1963). Application of operant conditioning procedures to the behavior problems of an autistic child. *Behaviour Research and Therapy*, 1(2–4), 305–312. [https://doi.org/10.1016/0005-7967\(63\)90045-7](https://doi.org/10.1016/0005-7967(63)90045-7)

Table 1*Participant, Setting, and Interventionist Characteristics*

Participant Variables and Intervention Characteristics	Number of Participants (N = 652)	Percentage of Participants (%)
Age		
0–5 years	240	36.8
6–11 years	199	30.5
12–18 years	76	11.7
> 18 years	35	5.4
Unknown	102	15.6
Gender		
Female	118	18.1
Male	435	66.7
Unknown	99	15.2
Autism		
Yes	501	76.8
No/not reported	151	23.2
Level of intellectual disability (n = 203)		
Mild/borderline	57	28.1
Moderate	44	21.7
Severe	20	9.9
Unspecified	82	40.4
Genetic/medical condition (n = 112)		
Fragile X syndrome	44	39.3
Down syndrome	18	16.1
Language disorder	12	10.7
ADHD	9	8.0
Multiple other genetic/medical	14	12.5
Other genetic/medical	15	13.4
History with discrete trial teaching		
Yes	168	25.8
No	15	2.3
Not reported	469	71.9
Setting location		
Clinic/center	291	44.6
School	192	29.4
Home/residential	75	11.5
Hospital - outpatient	23	3.5

Hospital - inpatient	9	1.4
Multiple	34	5.2
Unknown	28	4.3
Implementer		
Experimenter	405	62.1
Electronic	119	18.3
Teacher	85	13.0
Peer	9	1.4
College student	9	1.4
Parent	3	0.5
Multiple implementers	22	3.4

Note. ADHD = attention deficit hyperactivity disorder. The percentages reported for level of intellectual disability and genetic/medical conditions are based on the subset of participants for whom these apply, *n* is noted in parentheses.

Table 2*Discrete Trial Teaching Component Characteristics*

Discrete Trial Teaching Components	Number of Applications (N = 1,860)	Percentage of Applications
Antecedent		
Physical + vocal stimulus	1,397	75.1
Vocal stimulus	177	9.5
Electronic stimulus	161	8.7
Physical stimulus	125	6.7
Target response		
Receptive identification	834	44.8
Tact	599	32.2
Intraverbal	183	9.8
Textual	90	4.8
Mand	65	3.5
Instruction following	60	3.2
Play	20	1.1
Transcription	5	0.3
Echoic	2	0.1
Imitation	2	0.1
Consequence correct response		
Multiple (food or tangible)	436	23.4
Food	402	21.6
Attention	391	21.0
Tokens	344	18.5
Tangible	248	13.3
Auditory	10	0.5
N/A, no consequence, next trial	29	1.6
Corrective prompting strategy		
Single corrective prompt	430	23.1
Prompt hierarchy	416	22.4
Stimulus prompt fading	244	13.1
Prompt delay	236	12.7
Error correction	188	10.1
Errorless learning	98	5.3
Repeated corrective prompt	93	5.0
Flexible prompting	40	2.2
No corrective prompts used	115	6.2

Prompt delay

Yes	750	40.3
No	1,037	55.8
Unknown	72	3.9

Inter-trial interval

1-5 s	324	17.4
6-15 s	140	7.5
> 15 s	71	3.8
Progressive (2–20 s)	8	0.4
Narrative (i.e., immediate, brief, short)	141	7.6
Unknown	1,176	63.2

Table 3*Antecedents Used for Each Target Response Type*

Target Response	Number of Applications	Physical + Vocal Stimulus	Vocal Stimulus	Electronic Stimulus	Physical Stimulus
Receptive identification	834	609	0	157	68
Tact	599	579	0	4	16
Intraverbal	183	43	134	0	6
Textual	90	90	0	0	0
Instruction following	60	20	37	0	3
Mand	65	48	4	0	13
Play	20	6	0	0	14
Transcription	5	0	0	0	5
Echoic	2	0	2	0	0
Imitation	2	2	0	0	0

Table 4*Trials and Consequence Characteristics*

Characteristics	Number of Applications (<i>N</i> = 1,860)	Percentage of Applications
Number of trials (<i>N</i> = 134 studies)		
2–5	11	8.2
6–10	53	39.6
11–15	29	21.6
16–20	15	11.2
> 20	4	3.0
Varied	22	16.4
Identification of reinforcer		
Daily preference assessment/choice	542	29.1
Preference assessment at start	465	25.0
History (prior PA or history as reinforcer)	189	10.1
Caregiver nomination	48	2.6
Unknown	587	31.6
N/A, no consequence, next trial	29	1.6
Initial Schedule		
FR 1	1,773	95.3
VR/FR 3	56	3.0
Lag	2	0.1
N/A, no consequence, next trial	29	1.6
Schedule Thinning		
Yes	42	2.2
No	1,789	96.2
N/A, no consequence, next trial	29	1.6

Note. FR = fixed ratio, VR = variable ratio. The percentages reported for number of trials are based on the total number of studies rather than applications, *N* is noted in the parentheses.

Table 5*Types of Studies*

Study Type	Number of Studies	Percentage of Studies
Comparison of DTT Components	79	59.0
Prompting	29	36.7
Trial Presentation	26	32.9
Consequence	21	26.6
Intertrial Interval	3	3.8
Application of DTT to New Target	36	26.9
Response Type	24	63.9
Implementer	5	13.9
Population	4	11.1
Other	4	11.1
Comparison of DTT to Other Instruction Type	8	5.9
Fluency-Based Instruction	3	37.5
Mand Training	2	25.0
Incidental Teaching	1	12.5
Other	2	25.0
Other	11	8.2
Measurement	3	27.3
Mastery Criteria	3	27.3
Procedural Fidelity	2	18.1
Other	3	27.3

Note. DTT = discrete trial teaching.

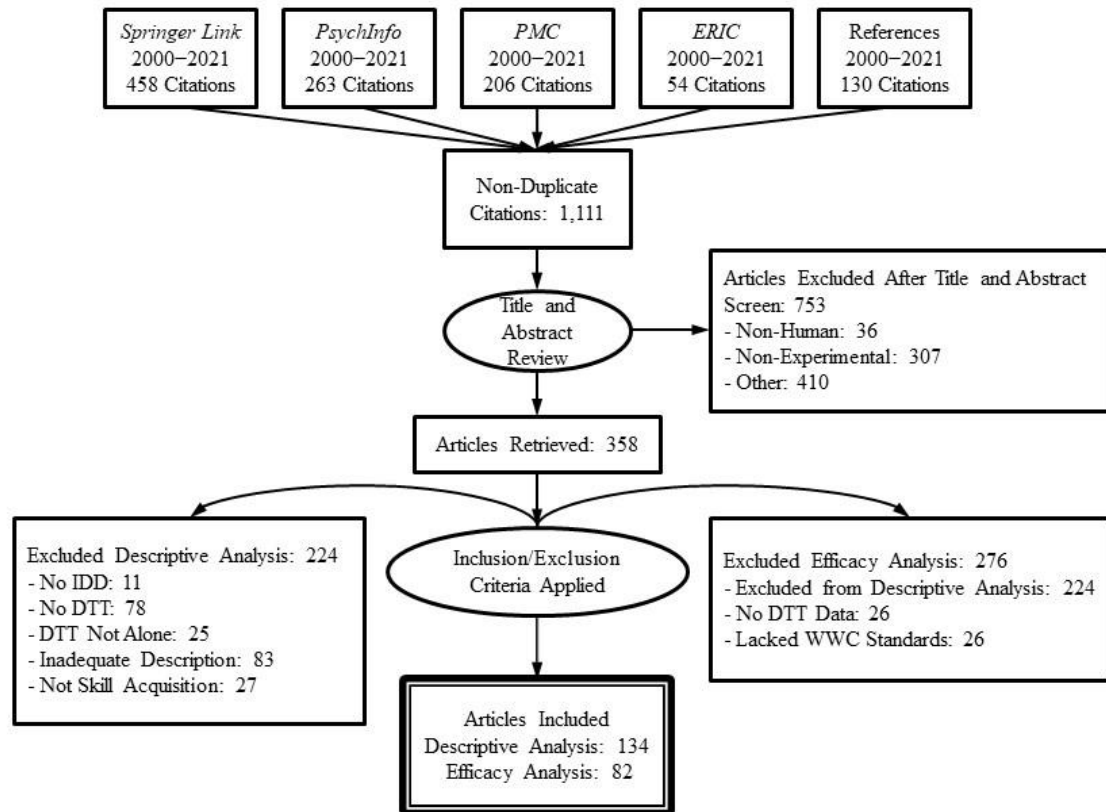
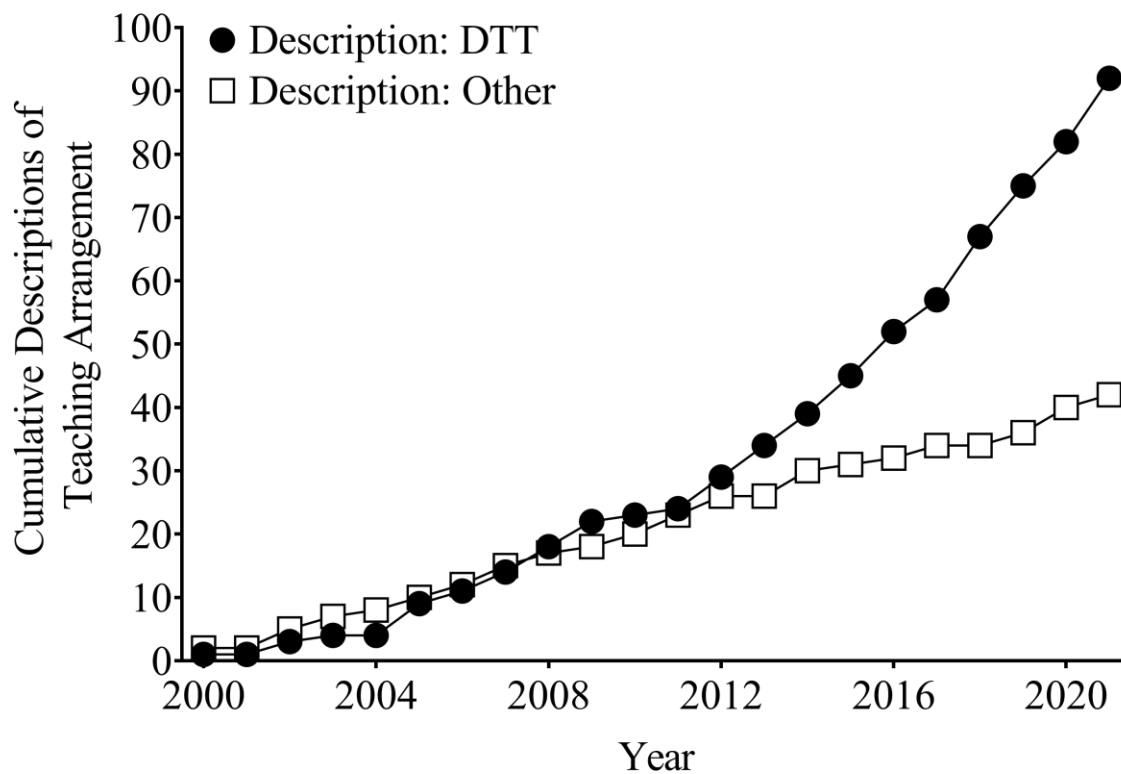
Figure 1*Flow Diagram of Search Results*

Figure 2

Cumulative Descriptions of the Teaching Arrangements as Discrete Trial or Other



Note. DTT = discrete trial teaching.

Figure 3

Cumulative Number of Studies that Evaluated Procedural Fidelity, Generalization, Maintenance, and Social Validity

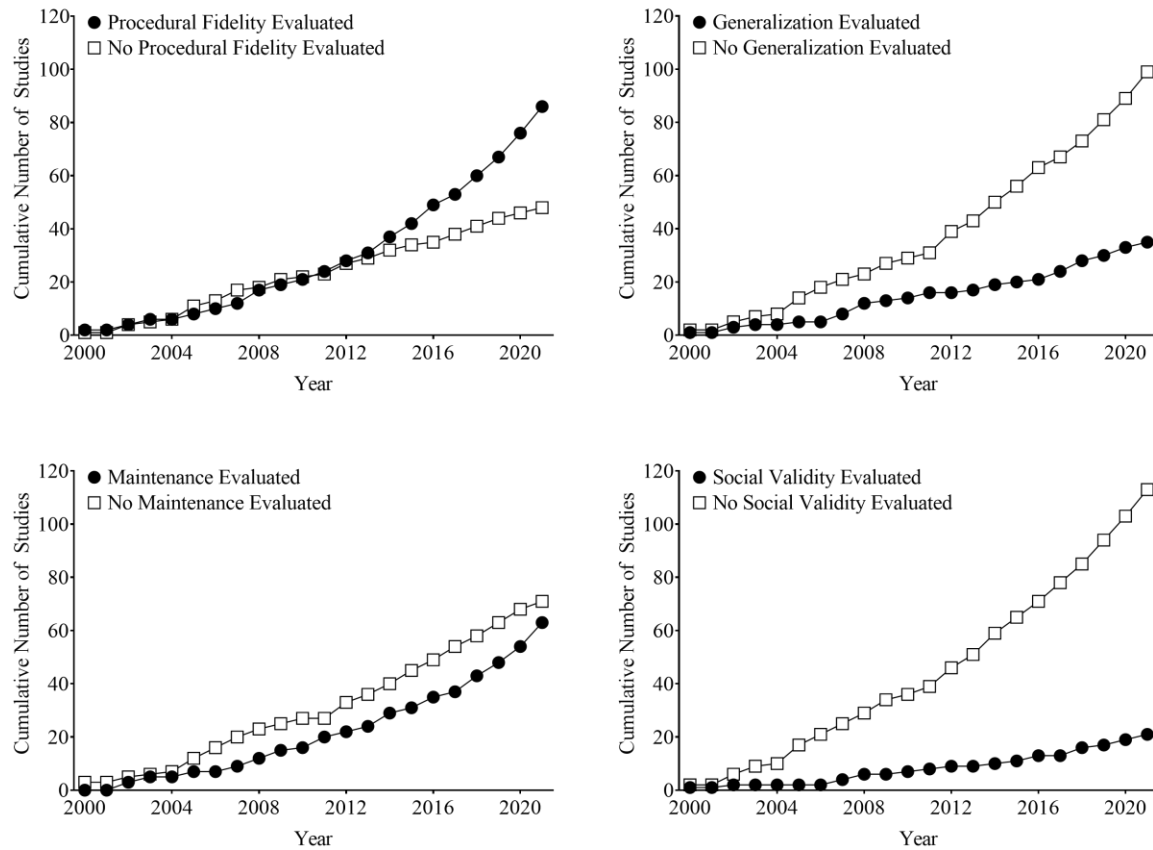
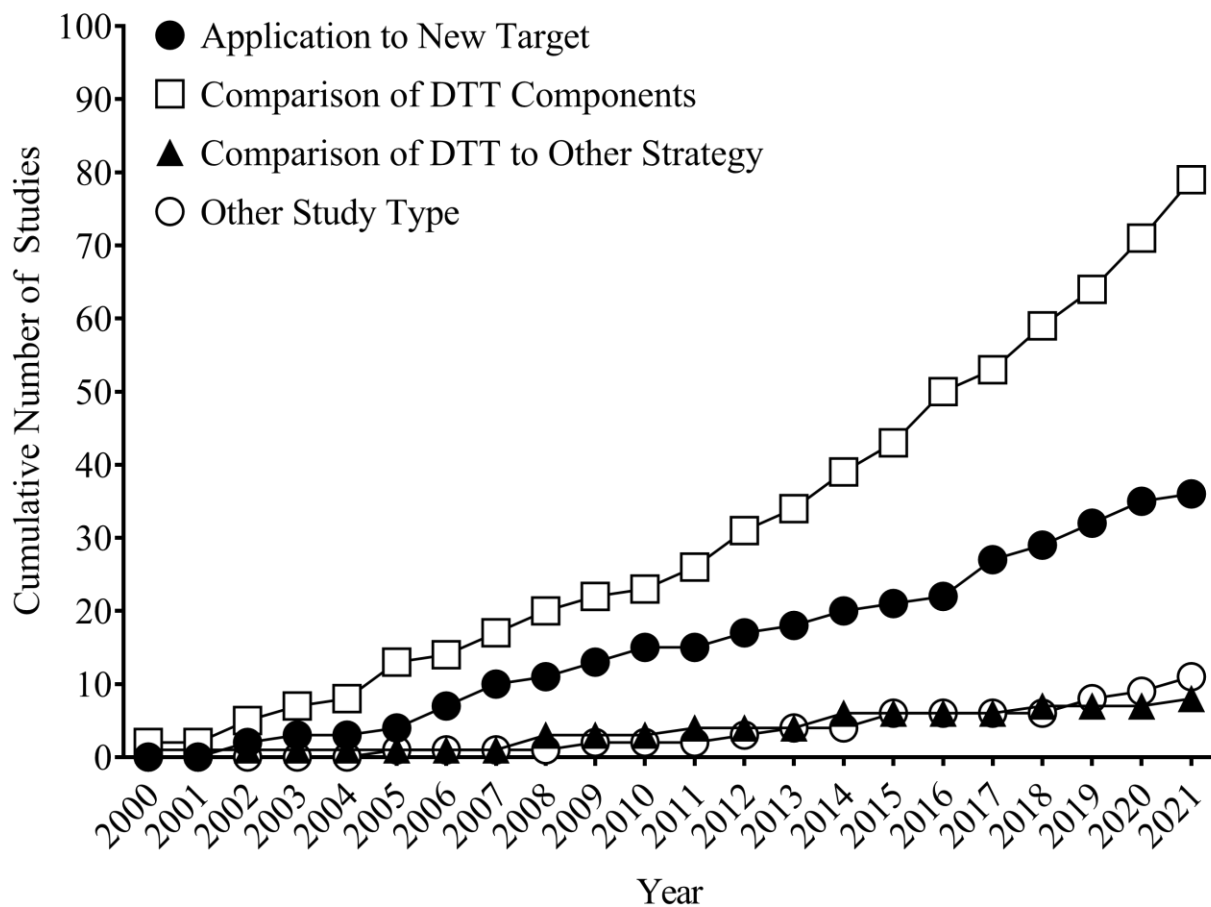


Figure 4*Cumulative Number of Each Type of Study*

Note. DTT = discrete trial teaching.