Hammill Institute on Disabilities

Focus on Autism and Other Developmental Disabilities
1–12

DOI: 10.1177/10883576221081076
focus.sagepub.com
SAGE

Training Individuals to Implement Discrete Trials with Fidelity: A Meta-Analysis

Joelle Fingerhut[1] and Mariola Moeyaert[2]

Abstract

Discrete trial training is a popular teaching method for individuals with autism, but is not easily implemented with fidelity due to its complexity. This is the first meta-analysis of single-case experimental design studies to quantify the impact of behavioral skills training on individuals' ability to implement discrete trials with fidelity. Furthermore, this meta-analysis examines the four training methods that make up behavioral skills training—feedback, instruction, modeling, and rehearsal—to determine the "active ingredients" of behavioral skills training. A total of 46 single-case experimental design studies are included in this meta-analysis. Hierarchical linear modeling, which has the ability to analyze clustered data, is the meta-analytic technique used to estimate the effectiveness of behavioral skills training across studies. Results show that behavioral skills training has a statistically significant positive effect on discrete trial training implementation fidelity; therefore, behavior skills training is recommended for discrete trial training implementation instruction.

Keywords

autism spectrum disorders, educational evidence-based practices, instruction, special education, parent training

Discrete trial training (DTT) or discrete trial instruction (e.g., Pollard et al., 2014) is a popular evidence-based practice, and it is frequently used when working with children with autism spectrum disorders (Majdalany et al., 2014). Defined as an intensive 1:1 teaching session that is meant to be implemented in a distraction-free environment (Pellecchia et al., 2015), DTT can be applied by mass implementing the teaching trials, interspersing trials with previously mastered skills, or distributing trials with breaks in between trials (Henrickson et al., 2015; Majdalany et al., 2014).

DTT can include various components and steps. The three essential components of DTT, known as the three-term contingency (Skinner, 1953), are the instructor presenting an instruction (i.e., the discriminative stimulus; the instructor says, "Point to the picture of a dog"), waiting for the learner to respond (i.e., the behavioral response; the child points to the picture of a dog), and the instructor providing a consequence (i.e., reinforcing stimulus; the instructor says, "Great job pointing to the picture!"). Examples of other DTT components include, but are not limited to, securing the learner's attention before beginning the session, establishing a contingent tangible, and presenting the instruction only once. One defining component of DTT is the use of a hierarchy of prompts (e.g., from most to least intrusive) to promote errorless learning (e.g., Severtson & Carr, 2012). More recently, researchers have advocated for the use of a "progressive" version of DTT that encourages individuals to incorporate natural language and flexible prompt fading (Leaf et al., 2016). The DTT evaluation form (Fazzio et al., 2007), which is frequently cited in DTT training literature (e.g., Wightman et al., 2012), lists a total of 21 DTT components for instructors to master, but other trainers may utilize more or less components for a successful DTT session. One reason to have less DTT components, for example, would be to simplify the implementation procedure for certain individuals (e.g., parents, peers, or teaching assistants) who may have less experience using complex interventions. Readers are advised to reference Fazzio et al. (2007) to see an example of the components that can be a part of DTT.

Implementation fidelity of DTT has been directly linked to student success, highlighting the importance of implementing DTT with fidelity when working with students. For example, Jenkins et al. (2015) conducted a parametric analysis of implementation fidelity errors, and they found commission errors to be detrimental for skills acquisition. In another study, by Carroll et al. (2013), children with autism

[1]Rutgers University Graduate School of Applied and Professional Psychology, Piscataway, NJ, USA
[2]University at Albany – State University of New York, USA

Corresponding Author:
Joelle Fingerhut, Rutgers University Graduate School of Applied and Professional Psychology, 152 Frelinghuysen Rd., Piscataway, NJ 08854-8020, USA.
Email: jf1051@rutgers.edu

acquired target skills at a slower rate when the DTT was implemented with low fidelity. Therefore, proper training is essential, and even more so when considering the wide range of contexts in which discrete trials are implemented. DTT can be implemented in multiple different settings, such as schools, clinics, university research centers, and at home, and can be implemented by people with various skill sets, such as special education teachers, paraprofessionals, and parents.

Each context in which DTT is implemented presents different challenges and advantages. Implementing interventions in a home setting can be difficult because the parent or caregiver may be required to be home during sessions and there may be lack of oversight from staff regarding the implementation fidelity of the intervention (Leaf et al., 2018). Some challenges of implementing evidence-based practices like DTT in a clinical setting include cost, lack of parental involvement, and the need for more intensive trainings (Leaf et al., 2018). Within a classroom setting, paraprofessionals may be ill-prepared to implement evidence-based practices, such as discrete trials, because they do often do not receive extensive training (Brock & Carter, 2016). It is important that discrete trials can be implemented by different individuals and across contexts because DTT implementation fidelity has been directly linked to learner success; West et al. (2013) found that when evidence-based practices are not delivered properly, students may display limited progress on academic goals. Reed et al. (2013) found DTT to be most effective when implemented consistently and with fidelity. Thus, regardless of the context or individual who is implementing discrete trials, effective trainings are essential so that DTT can be implemented with fidelity and have the greatest possible positive effect on the learner.

Different training methods are used when teaching individuals to implement DTT. These training methods include, but are not limited to, instruction (e.g., Salem et al., 2009), feedback (e.g., McKenney & Bristol, 2015), modeling (Catania et al., 2009), and rehearsal (Arnal et al., 2007). Instruction manuals are sometimes utilized due to their flexibility, cost-effectiveness, and self-guided nature (Thomson et al., 2012), although previous research has found instruction alone to be an ineffective training method (e.g., Ward-Horner & Sturmey, 2012). Feedback, specifically performance feedback, can help to ensure that mistakes made during DTT implementation are corrected. One previously conducted meta-analysis on the effect of performance feedback on teacher treatment integrity found performance feedback to have a moderate effect on an individual's ability to implement evidence-based practices with fidelity (Solomon et al., 2012). Modeling is a beneficial training method because it allows the individual to see the different components of DTT "in action," as verbal or written instruction might not convey how each component should be delivered. Modeling also allows for the demonstration of different skills across different contexts (Morgan & Salzberg, 1992), which is useful since DTT can be utilized to teach a range of skills. Rehearsal is a useful training method because it allows the participant to practice more difficult elements of DTT that might not be understood through other types of interventions. For example, certain prompting procedures should be used after the DTT recipient responds incorrectly two times in a row; rehearsal allows for the trainer to engage the participant in these unique situations (Gerencser et al., 2018).

One popular training package is *behavioral skills training* (BST), which includes the four training methods of feedback, instruction, modeling, and rehearsal delivered together and/or sequentially. BST has been used successfully to teach and maintain accurate implementation fidelity of discrete trials (Lerman et al., 2016) and has been an effective training method for teachers, clinicians, and parents (Day-Watkins et al., 2018). Other research supports the finding of BST to be effective; Sarokoff and Sturmey (2004) found BST to be highly effective in training individuals to use DTT, and Brock and colleagues (2017) found it to be effective for teaching implementation of other educational strategies as well (e.g., reinforcement, prompting, pivotal response training). It is also an effective method for learner acquisition and promotes the generalization of teaching skills in novel situations (Fetherston & Sturmey, 2014).

Several studies have attempted to determine the "active components" of BST. Ward-Horner and Sturmey (2012) and Madzharova et al. (2012) found modeling and feedback to be the most essential components of BST, while LaBrot et al. (2018) found feedback to be the most impactful component. Solomon et al (2012) also found performance feedback to have a moderate effect on an individual's ability to implement evidence-based practices with fidelity, and Cardinal et al. (2017) found video modeling and feedback to be effective when delivered together. However, as noted by LaBrot et al. (2018), results from the componential analysis are difficult to interpret because instruction is often delivered first, and it thus can be difficult to evaluate the differential impact of the individual components. More research is needed in this area.

Although previous research has investigated the effect of BST and the different training components of BST on evidence-based practices, including DTT implementation fidelity (e.g., Leaf et al., 2019; Shapiro & Kazemi, 2017), these previous studies are qualitative reviews and did not calculate an effect size (i.e., quantifying the effectiveness of BST and the different training components). No known meta-analyses to date have quantitatively examined the impact of BST and the individual training components of BST on individuals' ability to implement discrete trials with fidelity. Furthermore, previous systematic reviews did not focus on single-case experimental design (SCED) studies;

they included both group design studies and single-case experimental design studies. Due to the detailed information that SCEDs provide about participants and the ability for participants to serve as their own controls, it would be beneficial to conduct a meta-analysis that focuses solely on SCEDs. Previous studies also did not examine whether the effects of trainings last across time. Several studies on DTT have sought to determine whether the training effectiveness lasted during a later maintenance phase (Catania et al., 2009; Parnell et al., 2017). It would be helpful to compile these results and conduct one analysis across studies (i.e., meta-analysis) to determine whether the DTT trainings have lasting effects.

Based on the aforementioned research, the research questions for this study are as follows:

Research Question 1 (RQ1): Is complete BST (i.e., feedback, instruction, modeling, and rehearsal implemented together and/or sequentially) more effective than trainings that utilize some training components of BST (i.e., studies with only one, two, or three BST training components)? We hypothesized that across studies, BST would be statistically significantly more effective than trainings that only utilized BST components.

Research Question 2 (RQ2): Are any individual BST training components (e.g., instruction, feedback, modeling, and rehearsal) more effective than other components? We hypothesized that different individual training components (e.g., feedback; Solomon et al., 2012) would be more impactful than others.

Research Question 3 (RQ3): Are skills gained from the training maintained over time (e.g., after 1 month)? We hypothesized that the training effects would be maintained after the training is completed.

Method

The procedures of this meta-analysis were conducted in accordance with recommendations from Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Moher et al., 2009). PRISMA provides standards and evidence-based procedures for researchers conducting systematic reviews and/or meta-analyses.

Inclusion and Exclusion Criteria

The search was limited to articles investigating the effectiveness of training(s) to teach DTT implementation fidelity. The training(s) needed to include BST or any of the four components of BST to be eligible. The DTT implementation fidelity needed to be directly observed and measured on a percentage scale (e.g., 0–100) and calculated by determining the total number of steps that were performed correctly by the participant divided by the total possible number of DTT steps. The articles selected for inclusion were limited to those that utilized an SCED. The data needed to be provided in graphical or tabular form to be included in the meta-analysis, and the independent variable needed to be systematically manipulated, as recommended by What Works Clearinghouse (WWC; 2020). Studies needed to show three demonstrations of intervention effectiveness at three different points in time (e.g., a multiple-baseline design with three participants; WWC, 2020). Studies also needed to meet WWC (2020) standards with or without reservations in regard to number of data points per phase (e.g., a reversal design with a minimum of four phases per case and a minimum of three data points in each phase). Studies that included DTT trainings paired with trainings for another teaching method were also not included to eliminate the possibility of the other trainings serving as confounding variables (e.g., Nosik & Williams, 2011). Last, studies that included duplicate data (e.g., dissertations that were later published) were excluded.

Search and Study Selection

The following scientific databases were used to select primary level articles: Education Full Text, PsycINFO, Educational Resources Information Center, and Academic Search Complete. We decided that SCED studies published between 1977 and 2019 would be eligible for inclusion. This start date is the year of the earliest article (Koegel et al., 1977) included in the systematic reviews by Leaf et al. (2019) and Thomson et al. (2009). The keywords used in the scientific databases were "discrete trial" AND "autism," "discrete trial" AND "instructor," "discrete trial" AND "staff," "discrete trial" AND "paraprofessional," or "discrete trial" AND "parent."

Search terms were not restricted, could be present in any field (e.g., title, abstract, full text), and included gray literature such as dissertations and theses. Gray literature was included to help address issues of publication bias, as other methods commonly used to address publication bias such as funnel plots and/or robust Egger's regression are not suitable for SCEDs (Barton et al., 2017).

Two coders independently completed the primary search process. First, they read the title and/or the abstract of all the primary search studies to determine whether the study potentially met the inclusion criteria. If the title indicated that the study was potentially related to teaching discrete trial implementation, the coders read the study abstract as well. After this screening process, any articles that were not related to teaching discrete trial implementation were eliminated. The coders then screened the full text of the remaining articles, and articles were further eliminated if they did not fit the study eligibility criteria. Next, an ancestry search was conducted using the systematic reviews by Leaf et al. (2019), Shapiro and Kazemi (2017), and Thomson et al.

(2009) to determine additional studies. The title and abstract of all articles included within the systematic reviews were screened.

Total agreement among the two researchers for articles eligible for full-text review was 95%. Agreement between researchers was 100% for articles that were fully screened and met eligibility criteria. A graphical display of the selection process is displayed in Appendix A in the online supplemental materials.

The data software program WebPlot Digitizer 2.0 (Moeyaert et al., 2016; Rohatgi, 2014) was then used to retrieve raw SCED data from graphs of primary studies retained for inclusion in the meta-analysis.

Coding of Variables

A description of the different variables that were coded for each of the included articles is provided in the next section. Each variable was operationally defined so that the variables would be reliably coded by the researchers. These variables were used in inferential analyses.

BST. BST occurred when instruction, modeling, rehearsal, and feedback were implemented simultaneously or sequentially during the training.

Feedback. For the purpose of this study, the definition of feedback was limited to performance feedback. When a participant performed DTT and received information from an observer regarding the quality and/or accuracy of their DTT implementation, this was considered feedback. Feedback from the result of quizzes or tests (sometimes included as part of instruction) was excluded from this definition (e.g., Thiessen et al., 2009).

Instruction. This was defined as any form of information about implementing DTT that was given verbally, digitally, or in print form.

Modeling. Modeling was defined as any demonstration of DTT implementation by another person that occurred in real life or video recording (e.g., Catania et al., 2009).

Rehearsal. This was defined as structured role-play of DTT implementation that occurred with another person in the presence of a trainer. Rehearsal did not include instances where the participant independently practiced the skill or practiced the skill with an imaginary person (e.g., Thomson et al., 2012). Rehearsal could occur during data collection in the intervention phase if participants delivered DTT to a trainer role-playing a child (e.g., Higbee et al., 2016).

All of the aforementioned variables were dummy coded: 0 if the variable was not present and 1 if the variable was present.

Implementation fidelity. This was defined as the percentage of DTT steps implemented correctly. For example, if there were 10 DTT components identified in the study, and the participant completed eight of the 10 components correctly, the participant scores 80% implementation fidelity for that session.

Maintenance session: Maintenance was coded for each individual observation. Maintenance phases were any sessions that took place at least 1 day after the training had ended. If a maintenance phase occurred, it was recorded in number of days after the last intervention phase.

Six other variables were also coded for this study, including the participant's experience implementing DTT, the number of discrete trial steps included for measuring the dependent variable implementation fidelity, and if the participant practiced with a child or a confederate (adult acting as a child). The definitions of these other variables can be found in Appendix B in the online supplemental materials.

Interobserver Agreement

Interobserver agreement (IOA) was calculated for a subset of studies (30%). For this subset, each of the variables was double coded by the first author of this study and another doctoral student with past experience implementing DTT. The agreement among coders was divided by the total number of agreements plus disagreements. This was calculated for each individual variable. The total IOA for variables coded across observations, participants, and studies was 100%. A detailed version of the IOA for each individual variable is available upon request.

Risk of Bias

Risk of bias for individual studies was evaluated by the first author of this study. The domains that were evaluated for each study can be found in Appendix C in the online supplemental materials, and they are adapted from the risk of bias tool for single-case design studies (Reichow et al., 2018). This quality appraisal tool was chosen for use as it was based upon the Cochrane Risk of Bias tool, with additional SCED-specific considerations (Reichow et al., 2018). This tool is also more comprehensive than other tools available for use, such as the Single-Case Analysis and Review Framework (SCARF; Ledford et al., 2016). Using the nine domains and definitions of low, high, and unclear risk of bias provided by Reichow et al. (2018), each study was evaluated and rated as having either a low, high, or unclear risk of bias. The same second coder who conducted IOA also evaluated the risk of bias for each study, and the interrater reliability was 100%.

Analysis

Hierarchical linear modeling (HLM) is a meta-analytic method that can be used for the quantitative integration of primary level effect sizes. HLM has several benefits that make it a desirable analysis technique for single-case design meta-analyses. HLM is able to analyze clustered data, or observations within participants and participants within studies (Van den Noortgate & Onghena, 2003). The effect sizes are allowed to vary across participants and studies. The HLM method allows investigating the effectiveness of an intervention for multiple participants and studies using one analytic model.

Analysis Models

DTT implementation fidelity score ($DTTIA_{ijk}$) is the dependent variable and represents the outcome score on measurement occasion i for participant j being part of study k. The variable "Moderator" represents the moderating variable (BST or one of the individual training components). $Phase_{ijk}$ is a dummy-coded independent variable. This variable indicates whether observation i nested within participant j belonging to study k is part of the baseline (i.e., $Phase_{ijk}$ = 0) or treatment phase (i.e., $Phase_{ijk}$ = 1). $\text{Timec}_{\text{ijk}}$ was centered around the last training session of the treatment phase, that is, the last observation during the intervention phase was coded as 0 with previous observations coded in descending order and any following observations in ascending order. This was necessary to evaluate the effectiveness of BST and the individual training sessions at the end of the training (i.e., during the last observation data point during the training phase). The analysis model is introduced in Equation 1.

Level 1:

$$\begin{aligned} DTTIA_{ijk} = {} & \beta_{0jk} + \beta_{1jk} Phase_{ijk} + \\ & \beta_{2jk} Moderator * Phase_{ijk} + \\ & \beta_{3jk} \text{Timec}_{ijk} * Phase_{ijk} + \\ & \beta_{4ij} MT\text{Timec}_{ijk} * \\ & MTPhase_{ijk} + \phi + e_{ijk} \end{aligned} \tag{1}$$

In this equation, β_{0jk} indicates the baseline level for participant j nested within study k, and β_{1jk} indicates the difference between the baseline level and the last data point of the intervention phase for participant j nested within study k. β_{2jk} indicates whether the intervention effect differs across the different levels of the moderator variable. β_{3jk} indicates the slope during the treatment phase for participant j within study k. β_{4jk} indicates the change in slope between the treatment phase and maintenance phase for participant j within study k. In this equation, $\varnothing$ refers to the lag 1 autocorrelation, while e_{ijk} refers to the residuals, which are assumed to be identical and normally distributed. Autocorrelation is when observation points close to each other are more similar than observations that are further apart (Shadish & Sullivan, 2011). Lag 1 autocorrelation was added to the model because accounting for autocorrelation can help increase the control of type I error rate control, as well as improve the model's statistical power (Heyvaert et al., 2017)

Four different models were specified using Equation 1. Model 1 included the moderator BST, Model 2 included the moderator modeling, and Model 3 included the moderator rehearsal. Due to the large number of studies that included instruction in either the baseline or intervention phase (n = 44, 96%), instruction was unable to be isolated as a variable to determine the impact of this training component. Furthermore, too few studies included feedback as a training component separate from BST (n = 5, 11%), and so this training component could not be isolated and included in a separate analysis model. Equation 1 was also used without any moderators to find the overall effect across all training types, and this is referred to as Model 4. These four models used the same data set, so Bonferroni correction was used to control type I error rates ($\propto = .05/4 = .0125$).

The participant-specific parameters in Equation 1 were allowed to vary between participants, as it is unrealistic to assume that all participants and all studies would have the same parameter estimates. In the Level 2 equation introduced next, the case-specific population parameters from Equation 1 are allowed to vary around a study-specific parameter.

Level 2:

$$\begin{cases} \beta_{0jk} = \theta_{00k} + u_{ojk} \\ \beta_{1jk} = \theta_{10k} + u_{1jk} \\ \beta_{2jk} = \theta_{20k} + u_{2jk} \\ \beta_{3jk} = \theta_{30k} + u_{3jk} \end{cases} \text{with} \begin{bmatrix} u_{ojk} \\ u_{1jk} \\ u_{2jk} \\ u_{3jk} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{u_0} & & & \\ \sigma_{u_0 u1} & \sigma^2_{u1} & & \\ \sigma_{u_0 u2} & \sigma_{u2u1} & \sigma^2_{u2} & \\ \sigma_{u_0 u3} & \sigma_{u3u1} & \sigma_{u3u2} & \sigma^2_{u3} \end{bmatrix} \right) \tag{2}$$

In Equation 2, θ_{00k} indicates the mean baseline level across participants in study k. θ_{10k} indicates the difference between the baseline level and the last data point of the intervention phase across participants in study k. θ_{20k} indicates the effect of the moderator on the intervention effectiveness. θ_{30k} indicates the mean trend during the intervention phase across participants in study k. Furthermore, u_{0jk}, u_{1jk}, u_{2jk}, and u_{3jk} indicate how participant j from study k deviates from the study-specific baseline level, intervention effect, moderator effect, and time trend, respectively, during the intervention. The deviations are assumed to be multivariate normally distributed.

The study-specific parameters were allowed to vary between studies (see Equation 3).

Level 3:

$$\begin{cases}\theta_{00k} = \varphi_{000} + v_{00k} \\ \theta_{10k} = \varphi_{100} + v_{10k} \\ \theta_{20k} = \varphi_{200} + v_{20k} \\ \theta_{30k} = \varphi_{300} + v_{30k}\end{cases} \text{with} \begin{bmatrix} v_{00k} \\ v_{10k} \\ v_{20k} \\ v_{30k}\end{bmatrix} \sim N\left(\begin{bmatrix}0\\0\\0\\0\end{bmatrix}, \begin{bmatrix}\sigma^2_{v0} & & & \\ \sigma_{v1v0} & \sigma^2_{v1} & & \\ \sigma_{v2v0} & \sigma_{v2v1} & \sigma^2_{v2} & \\ \sigma_{v3v0} & \sigma_{v3v1} & \sigma_{v3v2} & \sigma^2_{v3}\end{bmatrix}\right) \quad (3)$$

In Equation 3, φ_{000} indicates the mean baseline level across participants and across studies. φ_{100} indicates the difference between the baseline level and the last data point of the intervention phase across participants and across studies. φ_{200} indicates the effect of the moderator on the intervention effectiveness across participants and studies, and φ_{300} indicates the trend in the intervention phase across participants and studies. v_{0jk}, v_{1jk}, v_{2jk}, and v_{3jk} indicate the deviation of study k from the mean baseline level, treatment level, moderator effect, and intervention trend, respectively. The deviations are assumed to be multivariate normally distributed. Meta-analysts are mainly interested in the coefficient estimates across studies, and so this is what is reported in the "Results" section, as answers to the research questions.

Study-specific effect sizes were also calculated. The effect sizes reflect the changes in level (change between baseline level and last observation of the intervention phase). This gives an idea about the variability in study-specific intervention effects. All the analyses were performed using the statistical software program RStudio (RStudio Team, 2015).

Results

Study Selection

The database search resulted in the identification of 3,394 articles, with 3,332 articles remaining after removing duplicates. After screening the title and abstracts of the latter, 86 articles remained eligible for full-text screening. After screening the full text of the 86 remaining articles, we eliminated articles if they did not fit the study eligibility criteria. After the full-text screening of the articles yielded from the database search, the ancestry search yielded a total of 92 additional articles, and two remained after duplicates were removed. By the end of the full-text screening of articles from the database search and the ancestry search, 43 articles remained eligible for inclusion. Finally, because some articles included more than one study, a total of 46 studies were included in the meta-analysis. A flowchart detailing the study selection procedure can be seen in Appendix A in the online supplemental materials.

Descriptive Statistics

The number of participants across all 46 studies was 224. A total of 130 (58%) participants received BST as their training. Detailed information regarding the variables and specific primary-level study characteristics can be found in Appendix D in the online supplemental materials, including the number of participants and participant type, the training(s) they received, and the criteria used within the study to determine if mastery of DTT implementation was reached. Table 1 demonstrates the combination of trainings found across studies and participants. A total of 51 participants (23%) participated in a maintenance phase. The maintenance phase occurred anywhere between 4 and 224 days after the last intervention session.

Risk of Bias

Risk of bias results can be found in Appendix C in the online supplemental materials. The studies included in this meta-analysis were mixed in terms of the category sequence generation. The majority of studies (n = 23, 53%) had unclear risk of bias for this category; many studies utilized a multiple-baseline design with staggered start points but did not specify how these start points were determined. Concerning the category of participant selection, clear inclusion criteria were used for 10 (23%) studies, indicating low risk of bias. There was an unclear risk of bias for the category of participant selection for 24 (56%) studies and high risk of bias for nine (21%) studies. In regard to blinding participants and personnel, only one study (Serna et al., 2016) explicitly stated that the participants were blind to the hypothesis during both the baseline and intervention phases, indicating a low risk of bias. For the category concerning procedural fidelity, 17 studies (40%) had a low risk of bias, and 24 studies (56%) had a high risk of bias, as they did not report procedural fidelity. As for the category of blinding outcome assessors, only two studies explicitly stated that the raters were naïve to the study hypothesis. Twenty-four studies (56%) had high risk of bias for this category for reasons such as the first author also recorded the data (e.g., Fazzio et al., 2009). For the category selective outcome reporting, most studies had a low risk of bias (n = 35, 81%), as they did not have missing data. For dependent variable

Table 1. Training Combinations.

Combination	Number of combinations within meta-analysis	
	Across all studies	Across all participants
BST	20	76
Extensive instruction	2	19
Extensive instruction + modeling	2	10
Extensive instruction + rehearsal	6	36
Extensive instruction + modeling + rehearsal	6	24
Feedback + instruction	2	6
Feedback + instruction + rehearsal	1	3
Feedback + modeling + rehearsal	1	9
Instruction	2	6
Instruction + modeling	1	4
Instruction + modeling + rehearsal	3	3
Modeling	1	3
Other	1	8

Note. BST = behavioral skills training.

reliability, all studies except one (O'Guin, 2010) had a low risk of bias. For the category data sampling, 15 studies (35%) had a low risk of bias, while 28 studies (65%) had an unclear risk of bias. Other possible risks of bias are reported in Appendix C.

Inferential Statistics

The effect size of the 46 individual studies can be seen in the forest plot in Appendix E in the online supplemental materials, which was created with the package *metaphor* (Viechtbauer, 2010) in RStudio (RStudio Team, 2015). This forest plot shows the intervention effect for each study along with a 95% confidence interval. The magnitude of the intervention effect varies between studies, and so HLM can offer more insights into the results as this analytical technique is able to separate within-case, between-case, and between-study variance.

The baseline level was consistent across all models and was estimated to be 40.61%, $\hat{\gamma}_{000} = 40.61$, $t(3400) = 16.76$, $p < .0125$. This indicates that before the intervention phase, participants were implementing DTT with 40.61% accuracy, which is considered to be low (see Appendix D in the online supplemental materials for mastery criteria). The mean change in DTT implementation fidelity between the baseline phase and the intervention phase at the last point of the intervention was estimated to be 49.97%, $\hat{\gamma}_{100} = 49.97$, $t(3693) = 18.25$, $p < .0125$. This was evaluated using Model 4, and it indicates that at the end of the trainings, the average participant was able to implement DTT with 90.58% accuracy. Previous studies (e.g., Gerencser et al., 2018; Nosik et al., 2013) used a criterion of 90% DTT implementation accuracy as a threshold for mastery of implementing DTT (see Appendix D in the online supplemental materials). Therefore, it can be concluded that overall, individuals can be trained to implement DTT with fidelity. An overall positive trend was detected during the intervention phase, $\hat{\gamma}_{200} = 1.33$, $t(3693) = 5.26$, $p < .001$, indicating that with each additional observation, participants' DTT implementation accuracy increased by 1.33.

RQ1. Model 1 was used to determine whether BST is more effective than trainings that utilize BST components (i.e., studies with only one, two, or three BST components). Results showed that the moderation effect of BST was 12.69%, $\hat{\gamma}_{200} = 12.69$, $t(3692) = 10.69$, $p < .0125$. Participants who received BST had an average estimated DTT implementation score of 96.06%, demonstrating that participants who received BST were statistically significantly more likely to have a higher DTT implementation score than those who did not receive BST (i.e., received only one, two, or three training components).

RQ2. Models 2 and 3 were used to determine whether modeling and rehearsal are effective training components. For Model 2 (modeling as the moderator), the moderation effect of modeling was 1.35%, $\hat{\gamma}200 = 1.35$, $t(3692) = 0.91$, $p = .36$. Participants who received modeling had an average DTT implementation score of 90.59%. This was not statistically significant, meaning that there is no difference in DTT implementation average for those who did and did not receive modeling as an intervention. For Model 3 (rehearsal as the moderator), the moderation effect of rehearsal was −1.79%, $\hat{\gamma}_{200} = -1.79$, $t(3692) = -0.99$, $p = .32$. Those who had rehearsal during the training implemented DTT with an average of 88.28% fidelity.

Participants who rehearsed during the training phase were not statistically significantly more likely to have a higher DTT implementation fidelity score than those who did not engage in rehearsal.

RQ3. Model 1 was used to determine whether the effects of training last across time. The change in rate between the slope of the intervention phase and the slope of the maintenance phase was −4.39, $\hat{\gamma}_{200} = -4.39$, $t(3399) = -2.86$, $p <$ .025. This indicates that there was a statistically significant difference between the intervention phase and maintenance slope. Participants had a slight decline in implementation fidelity after the intervention was finished.

Discussion

This meta-analysis attempts to address gaps in previous research regarding training methods' effectiveness for teaching participants how to use DTT. This is the first meta-analysis to the authors' knowledge that analyzed DTT training methods only while also estimating effect sizes and taking hierarchical structured data complexity into account by using an HLM technique. The aim of this study was to see if BST, which was previously demonstrated as an effective training for DTT (e.g., Sarokoff & Sturmey, 2004) would prove to be effective, as well as any of the four training components of BST (i.e., feedback, instruction, modeling, and rehearsal).

Through HLM analysis, it was affirmed that individuals can be trained to implement DTT with fidelity. Individuals implemented DTT with low levels of accuracy (40.61%) before receiving a training. Participants could likely implement DTT with low levels of accuracy since 61% of participants received instruction during baseline. After receiving a training, individuals were able to implement DTT with over 90% accuracy. As all studies had a criteria of at least 80% implementation fidelity as indicating mastery (59% of studies included in this meta-analysis set mastery at 90% or above), it can be concluded that, on average, individuals who undergo training are able to effectively implement DTT afterward. The statistically significant time trend ($\hat{\gamma}_{200} = 1.33$) likely reflects the impact of the different training types that were implemented during or between the observations in the intervention phase. For example, feedback was sometimes delivered between observations (e.g., Higbee et al., 2016); this may explain the finding that implementation fidelity increases with more observations.

When feedback, instruction, modeling, and rehearsal are implemented together (BST), they are statistically significantly effective. BST is demonstrated to be largely effective, as participants who received BST implemented DTT with 96.06% fidelity, on average. This finding is in line with findings from Brock et al. (2017), Sarokoff and Sturmey (2004), Lerman et al. (2016), and others who have demonstrated that BST is an effective training method for teaching practices. The results from this meta-analysis serve as supporting evidence that training methods that utilize BST are effective and recommended for use. Furthermore, most participants in the study (69%) had no prior experience implementing discrete trials before the training, demonstrating that BST is effective even for those with no previous experience.

The variables of modeling and rehearsal were analyzed separately to determine whether either of these training components was more effective than others. It is possible that this meta-analysis did not find modeling or rehearsal to be a statistically significantly impactful training component because these training methods were not often implemented with other methods that are historically known to work well together (e.g., feedback with modeling; Madzharova et al., 2012; Ward-Horner & Sturmey, 2012). When modeling was utilized as a training, it was used alone or with only one other training component 40% of the time, and only delivered together with feedback in one study. Similarly, rehearsal was implemented exclusively with instruction 55% of the time it was utilized; it is possible that for rehearsal to be demonstrated as an effective component, it needs to be delivered with a different training method besides instruction, like feedback.

Results demonstrate that after completing the training, DTT implementation skills drop slightly with time (−4.74). However, when accounting for outliers ($n = 14$), there does not appear to be a difference in DTT implementation fidelity during maintenance phases (e.g., 30 days later). Outliers were calculated by determining the interquartile range, multiplying the interquartile range by 1.5, and then adding this obtained number to the third quartile, while subtracting this number from the first quartile. Any number greater or less than this respective number was marked as an outlier. These results indicate that providing DTT implementers with summary guidelines or instructional checklists could help counter a potential drop in DTT ability with time.

The risk of bias results have implications regarding the quality of single-case design research. For example, only 17 of the 46 studies reported procedural fidelity of 80% or higher across at least 20% of the DTT sessions. Similarly, many of the studies ($n = 24$) included scorers who were not naïve to the study hypotheses. These results show areas that are in need of improvement within the general field of single-case design. Researchers who conduct single-case design studies are encouraged to reference the risk of bias domains as outlined by Reichow et al. (2018) so that they can conduct high-quality studies with low risk of bias across the different domains. Some of the Reichow et al. (2018) definitions are open to interpretation (e.g., the definition for data sampling bias). "Adequate number of data points" was interpreted to mean five data points, as the WWC (2020) document stated that five data points are needed to meet the standards *without reservation.* If instead three data points per phase are considered adequate, the number of studies with low risk of bias for data sampling increases from 17 (40%) to 45 (98%) studies. Researchers who use these risk of bias domains are encouraged to make clear their interpretations of ambiguous definitions (see note at bottom of Appendix C in the online supplemental materials for more details).

Limitations and Future Implications

This meta-analysis presents several limitations as well as implications for the direction of future research. Although HLM was determined to be the best approach for this data set due to the research questions and nesting characteristics, HLM has potential limitations, for example, distributional assumptions (e.g., assumption of homogeneous residuals). Nevertheless, future research is encouraged to use HLM as an analytical technique for meta-analyses due to its ability to consider the nesting of data (i.e., observations within cases and cases within studies). This is an important feature of HLM, since ignoring the natural hierarchy of data can lead to an underestimate of the standard error and an increase in Type I errors.

Several variables are important for consideration when determining the effectiveness of an intervention that could not be isolated and statistically analyzed. For example, the location of the training may impact the effectiveness of the training; most of the trainings were conducted at school (46%), university (28%), or private facility (14%). Due to the challenges that can occur when teaching at home (e.g., lack of oversight, issues related to generalization of skills Leaf et al., 2018]), more research is needed to determine if BST is effective for teaching DTT in a home environment. Relatedly, only 13% of participants in the study were parents; future research is needed to determine the effectiveness of BST to effectively teach parents, who may have less experience implementing evidence-based practices with fidelity. The variable "instruction" could not be analyzed in isolation due to its high prevalence in the studies (n = 44, 96%). Furthermore, instruction was frequently used by participants during the baseline phase (n = 126, 56%). Studies that had instruction during baseline were included because eliminating these studies would mean eliminating many studies from the analyses (n = 28, 61%). However, future research should examine the effectiveness of instruction alone and determine the extent to which this impacts DTT. This meta-analysis was also unable to isolate the variable feedback due to the small percentage of studies including this as a training component separate from BST (n = 5, 11%). It is, therefore, difficult to know the impact of these training methods in isolation on DTT implementation fidelity, and researchers in the future should continue to conduct research to determine the impact of these methods.

This meta-analysis did not look at student outcomes, as this information was not consistently provided in the primary studies. Although past research has demonstrated that student ability is correlated with DTT implementation fidelity (Carroll et al., 2013), it would be helpful for future research to examine the relationship between best practices and student outcomes.

Conclusion

As DTT is a frequently utilized training method for individuals with autism spectrum disorders, it is important to evaluate whether people are indeed being successfully trained to use these methods with fidelity. The purpose of this meta-analysis was to offer more insight into what makes an effective training for DTT. It supports previous research demonstrating the success of BST as a training method. An innovative and highly recommended technique, HLM, was utilized as it considers natural hierarchical data. As researchers continue to examine best practices in the future, more will be known about how to create the best trainings for different people and scenarios.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the Institute of Education Sciences, U.S. Department of Education, through grant R305D190022. The content is solely the responsibility of the author and does not necessarily represent the official views of the Institute of Education Sciences, or the U.S. Department of Education.

ORCID iD

Joelle Fingerhut https://orcid.org/0000-0002-8237-7578

Supplemental Material

Supplemental material is available on the webpage with the online version of the article.

References

References marked with an asterisk (*) indicate studies included in the meta-analysis.

*Arnal, L., Fazzio, D., Martin, G. L., Yu, C. T., Keilback, L., & Starke, M. (2007). Instructing university students to conduct discrete-trials teaching with confederates simulating children with Autism. *Developmental Disabilities Bulletin*, *35*(1), 131–147.

Barton, E. E., Pustejovsky, J. E., Maggin, D. M., & Reichow, B. (2017). Technology-aided instruction and intervention for students with ASD: A meta-analysis using novel methods of estimating effect sizes for single-case research.

Remedial and Special Education, *38*(6), 371–386. https://doi.org/10.1177/0741932517729508

*Belfiore, P. J., Fritts, K. M., & Herman, B. C. (2008). Using self-monitoring to enhance discrete trial. *Focus on Autism and Other Developmental Disabilities*, *23*(2), 95–102.

Brock, M. E., Cannella-Malone, H. I., Seaman, R. L., Andzik, N. R., Schaefer, J. M., Page, J. E., Barczak, M., & Dueker, S. A. (2017). Findings across practitioner training studies in special education: A comprehensive review and meta-analysis. *Exceptional Children*, *84*(1), 7–26.

Brock, M. E., & Carter, E. W. (2016). Efficacy of teachers training paras to implement peer support arrangements. *Exceptional Children*, *82*, 354–371. https://doi.org/10.1177/0014402915585564

*Cardinal, J. R., Gabrielsen, T. P., Young, E. L., Hansen, B. D., Kellems, R., Hoch, H., Nicksic-Springer, T., & Knorr, J. (2017). Discrete trial teaching interventions for students with autism. *Journal of Special Education Technology*, *32*(3), 138–148.

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.

*Catania, C. N., Almeida, D., Liu-Constant, B., & DiGennaro Reed, F. D. (2009). Video modeling to train staff to implement discrete-trial instruction. *Journal of Applied Behavior Analysis*, *42*(2), 387–392.

*Clayton, M., & Headley, A. (2019). The use of behavioral skills training to improve staff performance of discrete trial training. *Behavioral Interventions*, *34*(1), 136–143.

*Cowart, J. D., Iwata, B. A., & Poynter, H. (1984). Generalization and maintenance in training parents of the mentally retarded. *Applied Research in Mental Retardation*, *5*(2), 233–244.

Day-Watkins, J., Pallathra, A. A., Connell, J. E., & Brodkin, E. S. (2018). Behavior skills training with voice-over video modeling. *Journal of Organizational Behavior Management*, *38*(2–3), 258–273. https://doi.org/10.1080/01608061.2018.1454871

*Dib, N., & Sturmey, P. (2007). Reducing student stereotypy by improving teachers' implementation of discrete-trial teaching. *Journal of Applied Behavior Analysis*, *40*(2), 339–343.

*Downs, A., Downs, R., & Rau, K. (2008). Effects of training and feedback on discrete trial teaching skills and student performance. *Research in Developmental Disabilities*, *29*, 235–246.

*Eid, A. M., Aljaser, S. M., AlSaud, A. N., Asfahani, S. M., Alhaqbani, O. A., Mohtasib, R. S., Aldhalaan, H. M., & Fryling, M. (2017). Training parents in Saudi Arabia to implement discrete trial teaching with their children with autism spectrum disorder. *Behavior Analysis in Practice*, *10*(4), 402–406.

Fazzio, D., Arnal, L., & Martin, G. (2007). Discrete-Trials Teaching Evaluation Form (DTTEF). *Unpublished Scoring Manual*.

*Fazzio, D., Martin, G. L., Arnal, L., & Yu, D. C. T. (2009). Instructing university students to conduct discrete-trials teaching with children with autism. *Research in Autism Spectrum Disorders*, *3*, 57–66.

*Fetherston, A., & Sturmey, P. (2014). The effects of behavioral skills training on instructor and learner behavior across responses and skill sets. *Research in Developmental Disabilities*, *35*(2), 541–562.

*Garland, K. V., Vasquez, E., & Pearl, C. (2012). Efficacy of individualized clinical coaching in a virtual reality classroom for increasing teachers' fidelity of implementation of discrete trial teaching. *Education and Training in Autism and Developmental Disabilities*, *47*(4), 502–515.

*Gerencser, K. R., Higbee, T. S., Contreras, B. P., Pellegrino, A. J., & Gunn, S. L. (2018). Evaluation of interactive computerized training to teach paraprofessionals to implement errorless discrete trial instruction. *Journal of Behavioral Education*, *27*(4), 461–487.

*Gilligan, K., Luiselli, J., & Pace, G. (2007). Training paraprofessional staff to implement discrete trial instruction: Evaluation of a practical performance feedback intervention. *The Behavior Therapist*, *30*(3), 63–66. https://psycnet.apa.org/record/2007-11699-002

Henrickson, M., Rapp, J., & Ashbeck, H. (2015). Teaching with massed versus interspersed trials: Effects on acquisition, maintenance, and problem behavior. *Behavioral Interventions*, *30*, 36–50.

Heyvaert, M., Moeyaert, M., Verkempynck, P., Van den Noortgate, W., Vervloet, M., Ugille, M., & Onghena, P. (2017). Testing the intervention effect in single-case experiments: A Monte Carlo simulation study. *Journal of Experimental Education*, *85*(2), 175–196.

*Higbee, T., Aporta, A. P., Resende, A., Nogueira, M., Goyos, C., & Pollard, J. S. (2016). Interactive computer training to teach discrete-trial instruction to undergraduates and special educators in Brazil: A replication and extension. *Journal of Applied Behavior Analysis*, *49*(4), 780–793.

Jenkins, S. R., Hirst, J. M., & DiGennaro Reed, F. D. (2015). The effects of discrete-trial training commission errors on learner outcomes: An extension. *Journal of Behavioral Education*, *24*(2), 196–209. https://doi.org/10.1007/s10864-014-9215-7

*Jull, S., & Mirenda, P. (2016). Effects of a staff training program on community instructors' ability to teach swimming skills to children with autism. *Journal of Positive Behavior Interventions*, *18*(1), 29–40. https://doi.org/10.1177/1098300715576797

*Koegel, R. L., Glahn, T. J., & Nieminen, G. S. (1978). Generalization of parent-training results. *Journal of Applied Behavior Analysis*, *11*(1), 95–109. https://doi.org/10.1901/jaba.1978.11-95

*Koegel, R. L., Russo, D. C., & Rincover, A. (1977). Assessing and training teachers in the generalized use of behavior modification with autistic children. *Journal of Applied Behavior Analysis*, *10*(2), 197–205. https://doi.org/10.1901/jaba.1977.10-197

LaBrot, Z. C., Radley, K. C., Dart, E. H., Moore, J. W., & Cavell, H. J. (2018). A component analysis of behavioral skills training for effective instruction delivery. *Journal of Family Psychotherapy*, *29*, 122–141.

*Lafasakis, M., & Strumey, P. (2007). Training parent implementation of discrete-trial teaching: Effects on generalization of parent teaching and child correct responding. *Journal of Applied Behavior Analysis*, *40*(4), 685–689.

Leaf, J., Cihon, J., Leaf, R., McEachin, J., & Taubman, M. (2016). A progressive approach to discrete trial teaching: Some current guidelines. *International Electronic Journal of Elementary Education*, *9*(2), 361–372.

Leaf, J. B., Aljohani, W. A., Milne, C. M., Ferguson, J. L., Cihon, J. H., Oppenheim-Leaf, M. L., McEachin, J., & Leaf, R. (2019). Training behavior change agents and parents to implement discrete trial teaching: A literature review. *Review Journal of Autism and Developmental Disorders*, *6*(1), 26–39. https://doi.org/10.1007/s40489-018-0152-6

Leaf, J. B., Leaf, R., McEachin, J., Cihon, J. H., & Ferguson, J. L. (2018). Advantages and challenges of a home and clinic-based model of behavioral intervention for individuals diagnosed with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *48*(6), 2258–2266. https://doi.org/10.1007/s10803-017-3443-3

*Leblanc, M., Ricciardi, J. N., & Luiselli, J. K. (2005). Improving discrete trial instruction by paraprofessional staff through an abbreviated performance feedback intervention. *Education and Treatment of Children*, *28*(1), 76–82.

Ledford, J. R., Lane, J. D., Zimmerman, K. N., Chazin, K. T., & Ayres, K. A. (2016, April). *Single case analysis and review framework (SCARF)*. http://ebip.vkcsites.org/scarf/

*Lerman, D., Hawkins, L., Hoffman, R., & Caccavale, M. (2013). Training adults with an autism spectrum disorder to conduct discrete-trial training for young children with autism: A pilot study. *Journal of Applied Behavior Analysis*, *46*(2), 465–478. https://doi.org/10.1002/jaba.50

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.

Madzharova, M. S., Sturmey, P., & Jones, E. A. (2012). Training staff to increase manding in students with autism: Two preliminary case studies. *Behavioral Interventions*, *27*, 224–235.

Majdalany, L., Wilder, D., 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.

*Mason, R. A., Schnitz, A. G., Wills, H. P., Rosenbloom, R., Kamps, D. M., & Bast, D. (2017). Impact of a teacher-as-coach model: Improving paraprofessionals fidelity of implementation of discrete trial training for students with moderate-to-severe developmental disabilities. *Journal of Autism and Developmental Disorders*, *47*(6), 1696–1707.

*McKenney, E., & Bristol, R. (2015). Supporting intensive interventions for students with autism spectrum disorder: Performance feedback and discrete trial teaching. *School Psychology Quarterly*, *30*(1), 8–22.

*McKenney, E. L. W., Page, V., Lakota, J., Niekra, N., & Thompson, S. J. (2019). Supporting integrity of discrete trial teaching via tiered consultation: A pilot study. *Journal of Applied School Psychology*, *35*(1), 52–74. https://doi.org/10.1080/15377903.2018.1493555

*McKinney, T., & Vasquez, E. (2014). There's a bug in your ear! Using technology to increase the accuracy of DTT implementation. *Education and Training in Autism and Developmental Disabilities*, *49*, 594–601.

Moeyaert, M., Maggin, D., & Verkuilen, J. (2016). Reliability, validity, and usability of data extraction programs for single-case research designs. *Behavior Modification*, *40*(6), 874–900. https://doi.org/10.1177/0145445516645763

Moher D., Liberati A., Tetzlaff J., & Altman D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ, 6*(7), 1–6.

Morgan, R., & Salzberg, C. (1992). Effects of video-assisted training on employment-related social skills of adults with severe mental retardation. *Journal of Applied Behavior Analysis*, *25*, 365–383.

Nosik, M. R., & Williams, W. L. (2011). Component evaluation of a computer based format for teaching discrete trial and backward chaining. *Research in Developmental Disabilities*, *32*(5), 1694–1702. https://doi.org/10.1016/j.ridd.2011.02.02

*Nosik, M. R., Williams, W. L., Garrido, N., & Lee, S. (2013). Comparison of computer based instruction to behavior skills training for teaching staff implementation of discrete-trial instruction with an adult with autism. *Research in Developmental Disabilities*, *34*(1), 461–468.

*O'Guin, D. (2010). *Training paraprofessionals to implement a discrete trial language intervention for children with autism* [Unpublished doctoral dissertation]. Oklahoma State University, Stillwater.

*Parnell, A. M., Lorah, E. R., Karnes, A., & Schaefer-Whitby, P. (2017). Effectiveness of job aids and post performance review on staff implementation of discrete trial instruction. *Journal of Organizational Behavior Management*, *37*(2), 207–220.

Pellecchia, M., Connell, J. E., Beidas, R. S., Xie, M., Marcus, S. C., & Mandell, D. S. (2015). Dismantling the active ingredients of an intervention for children with autism. *Journal of Autism and Developmental Disorders*, *45*(9), 2917–2927.

*Pollard, J. S., Higbee, T. S., Akers, J. S., & Brodhead, M. T. (2014). An evaluation of interactive computer training to teach instructors to implement discrete trials with children with autism. *Journal of Applied Behavior Analysis*, *47*(4), 765–776.

*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.

Reed, F. D. D., Hirst, J. M., & Howard, V. J. (2013). Empirically supported staff selection, training, and management strategies. In D. D. Reed, R. F. D. DiGennaro, & J. K. Luiselli (Eds.), *Handbook of crisis intervention and developmental disabilities* (pp. 71–85). Springer.

Reichow, B., Barton, E. E., & Maggin, D. M. (2018). Development and applications of the single-case design risk of bias tool for evaluating single-case design research study reports. *Research in Developmental Disabilities*, *79*, 53–64. https://doi.org/10.1016/j.ridd.2018.05.008

Rohatgi, A. (2014). *WebPlotDigitizer user manual version 3.4*. http://arohatgi.info/WebPlotDigitizer/userManual.pdf

RStudio Team. (2015). *RStudio: Integrated development for R. RStudio, Inc* [Computer software]. http://www.rstudio.com

*Salem, S., Fazzio, D., Arnal, L., Fregeau, P., Thomson, K., Martin, G. L., & Yu, C. T. (2009). A self-instructional package for teaching university students to conduct discrete-trials teaching with children with autism. *Journal on Developmental Disabilities*, *15*(1), 21–29.

*Sarokoff, R. A., & Sturmey, P. (2004). The effects of behavioral skills training on staff implementation of discrete-trial teaching. *Journal of Applied Behavior Analysis*, *37*, 535–538.

*Sarokoff, R. A., & Sturmey, P. (2008). The effects of instructions, rehearsal, modeling, and feedback on acquisition and generalization of staff use of discrete trial teaching and student correct responses. *Research in Autism Spectrum Disorders*, *2*(1), 125–136.

*Serna, R. W., Foran, M. M., Cooke, C., Hurd, K. E., Tello, A. J., Vangapalli, R., & Hamad, C. D. (2016). Teaching discrete trial training. *Journal of Special Education Technology*, *31*(1), 39–49.

*Severtson, J. M., & Carr, J. E. (2012). Training novice instructors to implement errorless discrete-trial teaching: A sequential analysis. *Behavior Analysis in Practice*, *5*(2), 13–23.

Shadish, W. R., & Sullivan, K. J. (2011). Characteristics of single-case designs used to assess intervention effects in 2008. *Behavior Research Methods*, *43*, 971–980.

Shapiro, M., & Kazemi, E. (2017). A review of training strategies to teach individuals implementation of behavioral interventions. *Journal of Organizational Behavior Management*, *37*(1), 32–62.

Skinner, B. F. (1953). *Science and human behavior*. Macmillan.

Solomon, B. G., Klein, S. A., & Politylo, B. C. (2012). The effect of performance feedback on teachers' treatment. *School Psychology Review*, *41*(2), 160–175.

*Subramaniam, S., Brunson, L. Y., Cook, J. E., Larson, N. A., Poe, S. G., & St Peter, C. C. (2017). Maintenance of parent-implemented discrete-trial instruction during videoconferencing. *Journal of Behavioral Education*, *26*(1), 126. https://doi.org/10.1007/s10864-016-9258-z

*Thiessen, C., Fazzio, D., Arnal, L., Martin, G. L., Yu, C. T., & Keilback, L. (2009). Evaluation of a self-instructional manual for conducting discrete-trials teaching with children with autism. *Behavior Modification*, *33*(3), 360–373.

*Thomas, B. R. (2013). Effects of conducting peer behavioral observations on the observer's correct use of discrete trial teaching procedures. *Research in Developmental Disabilities*, *34*(7), 2143–2148. https://doi.org/10.1016/j.ridd.2013.03.033

Thomson, K., Martin, G. L., Arnal, L., Fazzio, D., & Yu, C. T. (2009). Research in autism spectrum disorders instructing individuals to deliver discrete-trials teaching to children with autism spectrum disorders: A review. *Research in Autism Spectrum Disorders*, *3*, 590–606.

*Thomson, K. M., Martin, G. L., Fazzio, D., Salem, S., Young, K., & Yu, C. T. (2012). Evaluation of a self-instructional package for teaching tutors to conduct discrete-trials teaching with children with autism. *Research in Autism Spectrum Disorders*, *6*(3), 1073–1082.

Van den Noortgate, W., & Onghena, P. (2003). Hierarchical linear models for the quantitative integration of effect sizes in single-case research. *Behavior Research Methods, Instruments & Computers*, *35*(1), 1–10.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. http://www.jstatsoft.org/v36/i03/

*Vladescu, J. C., Carroll, R., Paden, A., & Kodak, T. M. (2012). The effects of video modeling with voiceover instruction on accurate implementation of discrete-trial instruction. *Journal of Applied Behavior Analysis*, *45*(2), 419–423.

*Ward-Horner, J., & Sturmey, P. (2008). The effects of general-case training and behavioral skills training on the generalization of parents' use of discrete-trial teaching, child correct responses, and child. *Behavioral Interventions*, *284*, 271–284.

Ward-Horner, J., & Sturmey, P. (2012). Component analysis of behavior skills training in functional analysis. *Behavioral Interventions, 27*(2), 75–92. https://doi.org/10.1002/bin.1339

West, E. A., McCollow, M., Umbarger, G., Kidwell, J., & Cote, D. L. (2013). Current status of evidence-based practice for students with intellectual disability and autism spectrum disorders. *Education and Training in Autism and Developmental Disabilities*, *48*, 443–455.

What Works Clearinghouse (WWC). (2020). *What Works Clearinghouse standards handbook, Version 4.1*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. https://ies.ed.gov/ncee/wwc/handbooks

*Wightman, J., Boris, A., Thomson, K., Martin, G., Fazzio, D., & Yu, C. T. (2012). Evaluation of a self-instructional package for teaching tutors to conduct discrete-trials teaching with children with autism. *Journal of Developmental Disabilities*, *18*(3), 33–34.

*Young, K. L., Boris, A. L., Thomson, K. M., Martin, G. L., & Yu, C. T. (2012). Evaluation of a self-instructional package on discrete-trials teaching to parents of children with autism. *Research in Autism Spectrum Disorders*, *6*, 1321–1330.