Journal on Developmental Disabilities
Le journal sur les handicaps du développement
JoDD
Volume 23, Number 1, 2017

Teaching Staff Who Work With Children With Autism Spectrum Disorders to Evaluate the Treatment Integrity of Discrete-Trials Teaching Sessions

Authors

Jade Ellement,
Ashley Boris,
Garry Martin,
Daniela Fazzio

Department of Psychology,
University of Manitoba,
Winnipeg MB

Correspondence

jadekellement@gmail.com

Keywords

discrete-trials teaching,
treatment integrity,
self-instruction

Abstract

The current study evaluated the effectiveness of a self-instructional package to teach individuals to evaluate the treatment integrity (the degree to which intervention is implemented as planned) of discrete-trials teaching (DTT) sessions using the Discrete-Trials Teaching Evaluation Form (DTTEF). Participants were six staff from the St.Amant autism programs. We used a modified multiple-baseline design across a pair of participants, and replicated across two more pairs. At Baseline, a participant reviewed the 20-item DTTEF and a one-page summary of how to use it; they then attempted to assess the accuracy of DTT applied by a confederate role-playing an instructor teaching three tasks to a confederate role-playing a child with autism spectrum disorder (ASD). During training, participants studied a self-instructional package for using the DTTEF. At Post-training, participants were reassessed on the same tasks as during Baseline. During Generalization, participants were assessed evaluating the treatment integrity of three videos of an autism tutor administering DTT to a child with ASD. During a seven-month Follow-up, four participants were available and were reassessed. Training time averaged 1 hour and 16 minutes and mean accuracy increased from 47.6% in Baseline to 84.7% at Post-training. All participants showed excellent generalization results, and three of the four participants during Follow-up performed at a high level.

Treatment integrity is the degree to which an intervention is implemented as planned. If a behavioural intervention is administered with low treatment integrity, then the outcome cannot be interpreted with confidence (DiGennaro Reed & Codding, 2014). Although there have been numerous studies teaching individuals to administer discrete-trials teaching (DTT) (e.g., Arnal et al., 2007; Thiessen et al., 2009; Salem et al., 2009; Thomson et al., 2012; Wightman et al., 2012), few studies have examined methods to teach supervisors to evaluate the treatment integrity of such interventions applied by front-line staff (e.g., Peterson, Homer, & Wonderlich, 1982; Gresham, Gansle, & Noell, 1993; Wheeler, Baggett, Fox, & Blevins, 2006; McIntyre, Gresham, DiGennaro, & Reed, 2007).

Research has demonstrated that early intensive behavioural intervention (EIBI) is an effective treatment for some children with autism spectrum disorders (Matson & Smith, 2008; Matson & Sturmey, 2011). However, there is a paucity of

published information on instructor treatment integrity in EIBI, and specifically, on DTT, a main EIBI procedure.

In DTT an instructor administers a series of approximately 10-20 teaching trials, with brief inter-trial intervals, before providing a brief break. DTT has been effective for teaching a variety of behaviours to children with ASD (e.g., Smith, 2001). Across the research evaluating DTT training, it has been noted that the number and type of DTT components that are assessed for treatment integrity vary, are not always stated, are very brief in description, and are not consistently applied (e.g., Carroll, Kodak, & Fisher, 2013; Thomson, Martin, Arnal, Fazzio, & Yu, 2009).

To address the need for a valid and reliable method to evaluate the treatment integrity of a DTT session with a child with autism, Fazzio and colleagues observed a large number of training sessions administered by staff of the St.Amant autism programs in Manitoba, Canada, a government funded program that Provides E|IBI services to children with ASD. They then developed a 19-item checklist called the Discrete-Trials Teaching Evaluation Form (Fazzio, Arnal, and Martin, 2007a) along with the DTTEF Scoring Manual (Fazzio, Arnal, & Martin, 2007b). The 11-page manual provides correct and incorrect descriptions on instructor behaviour and is used to train DTTEF users. However, no research has been conducted to evaluate this manual.

The DTTEF is divided into five parts and has been modified to include 20 DTT items (Fazzio, Arnal, & Martin, 2012; see Figure 1). It has been demonstrated to: (a) possess high face validity; (b) possess high interobserver agreement; (c) distinguish between untrained and trained individuals who were implementing DTT; (d) possess concurrent validity; and (e) be high in social validity (Babel, Martin, Arnal, Fazzio, & Thomson, 2008; Jeanson et al., 2010).

Currently, the DTTEF is the only researched tool, proven to be valid and reliable, for assessing the treatment integrity of DTT sessions. In order to teach readers to use the DTTEF to reliably assess treatment integrity of DTT sessions, Wightman, Martin, Fazzio, and Arnal (2014) prepared the *Discrete-Trials Teaching Evaluation Form Self-Instructional Manual* (DTTEF-SIM). The purpose of the current study was to evaluate the effectiveness of the DTTEF-SIM with staff from the St.Amant autism programs who supervise the treatment integrity of DTT sessions conducted by staff working with children with ASD.

Materials and Methods

Participants

Ethics approval was obtained from the University of Manitoba Psychology and Sociology Research Ethics Board and St.Amant Research Centre. Participants consisted of six staff members recruited from the St.Amant autism programs. It was made clear that participation was voluntary and would in no way affect their job. Four participants were autism senior tutors and two participants were autism consultants. All participants had experience using DTT. Education level ranged from a bachelor's degree in psychology (all tutors) to a master's in psychology (all consultants). No participants had used the DTTEF prior to the study. All phases of the study were conducted at St.Amant in a private testing room during the day. Baseline and post-training sessions for a pair of participants were conducted during one day and were approximately 30 minutes each. Generalization sessions were conducted approximately one week after this and were 30 minutes long. Follow-up was conducted approximately one month following generalization and took approximately 30 min.

A trained research student at the doctoral level acted as a confederate role-playing an instructor and another trained research student at the doctoral level acted as a confederate role-playing a child with ASD. These individuals had several years of experience as confederates in such roles. Prior to the study, they were required to role-play with 100% accuracy over three sessions. In the study, several confederate/child pairs took part.

Materials

Baseline. A participant received the 20-item DTTEF and a one-page summary on how to use it. The confederate instructor received three

scripts, one for each teaching session, indicating how to instruct the session, trial by trial (see Table 1). The teaching sessions across phases consisted of one of three tasks: (a) matching pictures, (b) pointing-to-named-pictures, and (c) motor imitation. The confederate instructor also received teaching materials, which included a data sheet to record the responses of the confederate child, picture flash cards, edibles for reinforcement, and a pen. The confederate child received three scripts, one for each teaching session, indicating how to respond to the instructor, trial by trial (see Table 2).

Training. A participant received the DTTEF-SIM (described below), a computer to practice scoring videos, and a pen.

Post-training. The participant and confederates received the same materials as in Baseline.

High-integrity generalization. Participants received three DTTEFs to score three videos (described later) of a tutor administering DTT to a child with ASD. The videos consisted of 12 trials of the same three teaching tasks from Baseline.

Table 1. Example of Confederate Instructor Script

DTTEF SCORE FORM

COMPONENT	SCORE											
Part II: On Standard Trials, Manage Antecedents	1	2	3	4	5	6	7	8	9	10	11	12
7. Check the data sheet for the arrangement of teaching materials or response to be modeled.	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
8. Secure the child's attention	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
9. Present the teaching materials and/or model response	✔	✔	✔	✗	✗	✔	✗	✔	✔	✔	✔	✔
10. Present the correct instruction	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
11. Present Prompts	✗	✗	✗	✗	✗	✗	✔	✔	✔	✔	✔	✔
Part III: On Standard Trials, Manage Consequences & Record Data												
Score 12 or 13, NOT both — 12. Following a **correct response**, praise & present an additional reinforcer	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
Score 12 or 13, NOT both — 13. Following an **incorrect response**, block gently if possible, remove materials or stop gesturing & show a neutral expression for 2 or 3 seconds												
14a. Record the response immediately/accurately	✔	✗	✔	✗	✗	✔	✔	✔	✔	✔	✔	✔
15a. Allow brief intertrial interval of 3-10 seconds	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	/
Part IV: An Error Correction Trial Following An Error												
16. Secure the child's attention												
17. Re-present the materials												
18. Re-present the instruction & prompt immediately to guarantee correct response												
19. Praise only												
14b. Record the response immediately/accurately												
15b. Allow brief intertrial interval of 3-10 seconds												
Part V: Fade Prompts												
20. Fade prompts across trials	✗											

Follow-up. A participant received the same materials as in Baseline and training. The confederates received the same materials as in Baseline and training. All sessions were videotaped for data collection purposes.

Procedure

We used a concurrent modified multiple-baseline design across a pair of participants, repli-

Table 2. Example of Confederate Child Script

Pointing to Named Pictures Task

Pointing	1	**Pointing**	
Attending/Not attending	A	Attending/Not attending	A
Prompting level	FP	Prompting level	P1
Correct/Error	C	Correct/Error	C
Pointing	2	**Pointing**	9
Attending/Not attending	A	Attending/Not attending	A
Prompting level	FP	Prompting level	(P2)
Correct/Error	C	Correct/Error	E
Pointing	3	**Pointing – ERROR CORRECTION**	9
Attending/Not attending	Tap Table	Attending/Not attending	A
Prompting level	FP	Prompting level	P1
Correct/Error	C	Correct/Error	C
Pointing	4	**Pointing**	10
Attending/Not attending	A	Attending/Not attending	A
Prompting level	P1	Prompting level	(P2)
Correct/Error	C	Correct/Error	E
Pointing	5	**Pointing – ERROR CORRECTION**	10
Attending/Not attending	A	Attending/Not attending	A
Prompting level	(P1)	Prompting level	P1
Correct/Error	E	Correct/Error	C
Pointing – ERROR CORRECTION	5	**Pointing**	11
Attending/Not attending	A	Attending/Not attending	Look down
Prompting level	FP	Prompting level	P1
Correct/Error	C	Correct/Error	C
Pointing	6	**Pointing**	12
Attending/Not attending	A	Attending/Not attending	A
Prompting level	P1	Prompting level	P1
Correct/Error	C	Correct/Error	C
Pointing	7		
Attending/Not attending	A		
Prompting level	P1		
Correct/Error	C		

cated across two more pairs. Each pair of participants was randomly assigned to evaluate one of three levels of treatment integrity during Baseline and Post-training, low, moderate, or high described below. During a generalization phase, participants evaluated videos of a tutor administering DTT to a child with ASD. Finally during a follow-up phase, participants evaluated all three levels of treatment integrity.

Baseline. A participant filled out a background questionnaire that inquired about his or her experience using the DTTEF and DTT. Next, a participant was given one copy of the DTTEF in order to score a live simulated session of the confederate instructor teaching one of three tasks to a confederate role-playing a child with ASD. The teaching tasks were (a) matching pictures (e.g., when a picture of a dog, a house, and a tree are placed in front of a child, and the child is given a picture of the dog, the correct response is to match the pictures of the dog by placing one picture on top of the other); (b) pointing-to-named pictures (when three options are placed on the table in front of the child); and (c) motor imitation (e.g., a teacher would put his or her arms up and say "do this"). The script for each task was divided into 12 teaching trials and 20 DTT components. The scripts for each teaching task were programmed so that the three tasks were taught with a specific level of treatment accuracy, the percentage of correct instructor behaviour during a DTT session. A script was programmed with either (a) low (40%) DTT treatment accuracy, (b) moderate (70%) DTT treatment accuracy, or (c) high (90%) DTT treatment accuracy. The confederate instructor was provided with appropriate stimuli for each of the teaching tasks (described above). Each of the confederates was following a script on how to respond. The confederate instructor's scripts described how to administer each DTT component of each trial in a given session. For example, a script indicated to use an incorrect instruction and incorrect materials in a given trial. The confederate child's with ASD scripts indicated to attend or not attend to the instructor, to respond correctly or incorrectly, and what prompting level was required by the instructor in order to respond.

Across a pair of participants, we evaluated the treatment integrity of the DTT sessions across Baseline (and Post-training) sessions with a programmed level of DTT treatment accuracy of either low (40%), moderate (70%), or high (90%). The pairs of participants were randomly assigned to a DTT treatment accuracy condition. First, a participant received a brief overview of the study and completed the background questionnaire. Next, a participant was given 10 minutes to read a one-page summary of how to score one of three DTT tasks administered to a child with ASD and was given the DTTEF (see Table 3) to review. Once a participant indicated that he/she had finished reading a summary for the teaching task and reviewing the DTTEF, or ten minutes had passed, then he/she attempted to score, using the DTTEF, 12 trials of a live teaching session of the confederate instructor teaching the confederate child. After the first scoring session, a set of abbreviated instructions for a second task was introduced and a participant had 10-minutes to review the instructions and the DTTEF, followed by the attempt to score 12 trials of that session. This was repeated until a participant attempted to score each of three teaching tasks either once (the first participant of a pair) or twice (the second participant of a pair). The order of the tasks that were scored by a participant was random across participants.

Training. A participant received the DTTEF-SIM (available from first author) to study, blank paper, a pen, and a highlighter. The DTTEF-SIM consisted of 18 pages of instruction which included detailed descriptions on correct and incorrect instructor behaviour during HDTT sessions. It also included five training steps corresponding to each of the five parts of the DTTEF (see Table 1), descriptions of the 20 DTT components, 25 study questions, and four practice activities which involved using Parts 1-5 of the DTTEF to evaluate a video demonstration of a confederate instructor using DTT to teach a confederate child with ASD. Throughout the video demonstrations, there were several programmed errors on Components 9, 11, 14a, 14b, 18, and 20. Individuals tend to deliver these components with moderate (e.g., 60%-79%) to poor (e.g., 0%-59%) accuracy after receiving training (Wightman, Yates, Martin, Pear, & Yu, 2013). These components have been noted to be critical elements of DTT (e.g., Carroll et al., 2013; Holcombe, Wolery, & Snyder, 1994). Therefore, it was important to emphasize these components during staff training. When a participant

Table 3. The Discrete-Trials Teaching Evaluation Form (DTTEF) (revised 2012) (Fazzio, Arnal, & Martin, 2012)

DTTEF

SCORING: = ✔ performed correctly; **X** = performed incorrectly; / = did not apply

COMPONENT	**SCORE**
Part I: Prepare to Conduct a Teaching Session	
1. Determine the teaching task(s)	
2. Gather the teaching materials	
3. Select at least 3 reinforcers	
4. Arrange the teaching setting	
5. Determine the prompt-fading procedure and the initial fading step	
6. Invite the child to the table and give a reinforcer choice	

Part II: On Standard Trials, Manage Antecedents	1	2	3	4	5	6	7	8	9	10	11	12
7. Check the data sheet for the arrangement of teaching materials or response to be modelled.												
8. Secure the child's attention												
9. Present the teaching materials and/or model response												
10. Present the correct instruction												
11. Present Prompts												

Part III: On Standard Trials, Manage Consequences & Record Data

	1	2	3	4	5	6	7	8	9	10	11	12
Score 12 or 13, NOT both — 12. Following a **correct response**, praise & present an additional reinforcer												
Score 12 or 13, NOT both — 13. Following an **incorrect response**, block gently if possible, remove materials or stop gesturing & show a neutral expression for 2 or 3 seconds												
14a. Record the response immediately/accurately												
15a. Allow brief intertrial interval of 3-10 seconds												

Part IV: An Error Correction Trial Following An Error

	1	2	3	4	5	6	7	8	9	10	11	12
16. Secure the child's attention												
17. Re-present the materials												
18. Re-present the instruction & prompt immediately to guarantee correct response												
19. Praise only												
14b. Record the response immediately/accurately												
15b. Allow brief intertrial interval of 3-10 seconds												

Part V: Fade Prompts

20. Fade prompts across trials	

had completed studying the manual and each of the practice activities, then the mastery test was administered. It consisted of five questions taken from the DTTEF-SIM, one question from each of Parts 1-5. A participant was required to obtain 100% accuracy on the test in order to proceed to the Post-training phase, If a participant did not receive 100% accuracy, then he/she was required to re-study the material and re-write the question(s) until 100% was achieved.

Post-training. A participant was assessed on the same three tasks, with the same level of DTT treatment accuracy, following the same procedure as during Baseline. A participant was considered to have mastered evaluating the treatment integrity of a task if he/she obtained 80% accuracy or greater on that task. After the post-training assessment, a participant completed a social validity questionnaire.

High-integrity generalization. A participant used the DTTEF to evaluate three videos of an autism tutor teach three tasks, one task per video, to a child with ASD. Participant 5 only scored one generalization video due to scheduling difficulties. The teaching tasks were the same as those used in previous phases. To create the videos used during generalization sessions, the autism tutor was given three data sheets to teach 12 trials of each of the three teaching tasks to the child with ASD. The autism tutor also received a pen, teaching materials (picture flash cards for the pointing-to-named pictures and matching pictures teaching tasks) and edibles to use for reinforcement. The videos showed the autism tutor teaching the child with ASD each of the three tasks with 100% integrity, which was an unplanned result of recruiting an experienced autism tutor, and a high functioning child with ASD.

Follow-up. Approximately seven months after training, four of the six participants were available for a Follow-up assessment. Participants used the DTTEF to evaluate the confederate instructor teach a confederate, who role-played a child with ASD, each of the three teaching tasks, one task per session, as in previous phases. Each teaching task was taught with low (40%), moderate (70%), or high (90%) DTT treatment accuracy. The level of treatment integrity with which a teaching task was taught varied at random across participants.

The dependent variable and inter-observer agreement (IOA). The dependent variable was assessment accuracy using the DTTEF. To calculate the dependent variable, a participant's DTTEF score for a session was compared to the primary researchers DTTEF score for the same session. The scores for each DTT components across trials were compared and session accuracy was calculated by dividing the number of agreements by the number of agreements plus the number of disagreements and then multiplying by 100% (Martin & Pear, 2015).

To obtain IOA, for 30% of the sessions across each phase, a secondary observer who was a trained graduate student also used the DTTEF to score a participant's performance. To calculate IOA, the second observer's DTTEF score was compared to the primary researcher's DTTEF score. IOA was calculated by dividing the number of agreements by the number of agreements plus the number of disagreements and then multiplying by 100% (Martin & Pear, 2015). IOA averaged 92.8% (*SD* = = 5.61%; range: 87.0%–100%).

Procedural integrity (PI). To ensure the procedure was followed correctly, the primary researcher followed a script for each phase of the study. An observer recorded whether the procedure was followed as planned using a checklist that listed the steps that the primary researcher was supposed to follow for a given phase of the study. For a phase, PI was determined by computing the percent of steps that were administered correctly during that session. PI was completed for 78% of the sessions and averaged 100%.

PI was also assessed for the confederate instructor and confederate role-playing a child with ASD for 30% of the sessions. A trained research assistant recorded whether the confederate followed the confederate scripts as intended. Confederate PI was calculated by dividing the number of correct confederate behaviours by the number of correct confederate behaviours plus the number of incorrect confederate behaviours and then multiplying by 100%. Mean confederate instructor PI was 94.2% (*SD* = 3.73%; range: 89.7%–100%) and mean confederate child PI was 97.4% (*SD* = 5.03%; range: 89.2%–100%).

Results

Low DTT Treatment Accuracy Condition

Participant 1 and 2's assessment accuracy improved from Baseline to Post-training (see Figure 1). For Participant 1, improvement was 38.7%. Baseline scores averaged 36.9% (matching pictures, 35.2%; pointing-to-named pictures, 41.4%; motor imitation, 34.1%). Post-training scores averaged 75.6% (matching, 79.1%; pointing, 70.4%; imitation, 77.4%; not meeting mastery on any task. During Generalization assessment accuracy averaged 99.0% (matching, 99.0%; pointing, 99.0%; imitation, 99.0%). During Follow-up, Participant 1 averaged 87.32% (matching, low DTT treatment accuracy, 83.2%; pointing, high DTT treatment accuracy, 88.1%; imitation, moderate DTT treatment accuracy, 91.64%)

During Baseline, Participant 2's assessment accuracy averaged 47.6% across the six tasks (matching, 45.2%, 38.3%; pointing, 44.4%, 36.2%; imitation, 66.4%, 55.1%; see Figure 4). After training, Participant 2's assessment accuracy averaged 83.3% (matching, 80.1%; pointing, 84.9%; imitation, 84.9%; mastery met on all tasks), improving 35.79%. Generalization accuracy was also high (*M* = 99.0%; matching, 99.0%; pointing, 99.0%; imitation, 99.0%). This individual did not participate in the Follow-up sessions due to scheduling conflicts.

Moderate DTT Treatment Accuracy Condition

Participant 5 and 6 also improved from Baseline to Post-training (see Figure 2). For Participant 5, accuracy improved 56.2%. Baseline scores averaged 25.2% (matching, 34.3%; pointing, 22.1%; imitation, 19.3%). At Post-training, accuracy averaged 81.4% (matching, 83.1%; pointing, 76.2%; imitation, 85.1%; mastery met on two tasks). Due to scheduling issues, Participant 5 only evaluated the treatment integrity of one video during Generalization in which he scored 99.0%. During Follow-up accuracy averaged 84.0% (matching, high DTT treatment accuracy, 93.2%; pointing, moderate DTT treatment accuracy, 81.5%; imitation, low DTT treatment accuracy, 77.3%).

Participant 6's Baseline accuracy averaged 50.8% (matching, 57.1%, 41.2%; pointing, 39.3%, 55.1%; imitation, 55.2%, 57%). After training, Participant 6 averaged 89.3% (matching, 88.1%; pointing, 86.3%; imitation, 93.4%; mastery was met on all tasks), improving an average of 38.5%. During Generalization, Participant 6 scored 100% on all tasks. During Follow-up, Participant 6 evaluated the sessions with an average of 87.9% (matching, low DTT treatment accuracy, 83.1%; pointing, moderate DTT treatment accuracy, 88.2%; imitation, high DTT treatment accuracy, 92.4%).

High DTT Treatment Accuracy Condition

Participant 3's assessment of treatment integrity improved from Baseline to Post-training an average of 31.7% (see Figure 3). Participant 3 averaged 66.5% during Baseline (matching, 70.1%; pointing, 73.3%; imitation 56.0%). After training, Participant 3 averaged 98.2% (matching, 99.0%; pointing, 98.3%; imitation, 97.2%), meeting the 80% mastery criterion on all three tasks. During Generalization, Participant 3 averaged of 99% (matching, 99.0%, pointing, 99.0%, imitation, 99.0%). Participant 3 did not participate in the Follow-up Sessions due to scheduling conflicts.

Participant 4's mean assessment accuracy increased from Baseline to Post-training 30.3% (see Figure3). Participant 4 averaged 51.2% during Baseline (matching, 63.4%, 56.0%; pointing, 62.0%, 59.2%; imitation, 32.3%, 34.4%). Post-training scores averaged 81.5% (matching, 74%; pointing, 87.1%; imitation, 83.4%). Thus, Participant 4 met the mastery criterion on two of the three tasks during Post-training. During Generalization Participant 4's assessment accuracy averaged 86.0% (matching; 84.0%, pointing, 85.1%; imitation 88.0%). During Follow-up, assessment accuracy averaged 62.9% (matching, moderate DTT treatment accuracy, 52.0%; pointing, high DTT treatment accuracy, 95.7%; imitation, low DTT treatment accuracy, 41.1%).

Overall, participants averaged 1 hour and 16 minutes to study and master the material in the DTTEF-SIM, ranging from 40 minutes to 1 hour and 55 minutes. For all participants, assessment accuracy during Baseline was low, averaging

47.6% (*SD* = 14.2, range: 19.3%–73.3%), and after training, assessment accuracy increased, averaging 84.7% (*SD* = 8.2, range: 70.4%–99.0%). At Post-training, three of six participants met the mastery criterion of 80% accuracy on all three tasks, two participants met the mastery criterion on two of the three tasks, and one participant did not meet the mastery criterion on any of the three tasks. Accuracy in Generalization was very high for all participants, averaging 96.7% (*SD* = 5.5, range: 84.0%–100%). Finally, in the Follow-up phase, three of the four participants were able to evaluate the treatment integrity of low, moderate, and high DTT treatment conditions with high accuracy (*M* = 77.8%, *SD* = 18.7).

DTTEF Components that Participants Didn't Score

A missed component occurred when a DTT component for a trial was not scored and left blank. As demonstrated in Figure 4, the percentage of missed components was highest during Baseline for all participants except for Participant 5. Following training, the percentage of missed components decreased sizably. During the Generalization phase, the percentage of missed components was very low. During the Follow-up phase, the percentage of missed components also remained low.

The results also demonstrated that some DTTEF Components were missed more frequently than others. Figure 5 demonstrates that Components 14a and 15a were missed a total of 41 times each across participants and phases, followed by Component 20 which was missed 15 times across participants and phases, and then Component 10 which was missed nine times across participants and phases. The remaining DTTEF Components that participants missed occurred at a lower frequency across participants and phases, ranging from 1–6 times.

Social Validity

Participants completed a 7-question social validity form. There were two questions concerning the goals, two questions concerning the procedures, and three questions concerning the effects of the study. The questions were rated on a scale of 1 to 5 (1 = *disagree* and 5 = *agree*). Results suggest that participants found the goals of the study to be important, rating the items an average of 5/5, the procedure to be effective, rating the items an average of 4.88/5, and the training package to be effective, rating the items an average of 4.77/5.

Discussion

Overall, the self-instructional package was found to be effective in improving autism consultants and autism senior tutors accuracy in evaluating the treatment integrity of DTT sessions. The training package took a short amount of time to complete and resulted in an increase in assessment accuracy. Furthermore, positive results were maintained during Generalization sessions one week after training, and also during Follow-up sessions approximately seven months after training.

Several limitations should be noted. During the Generalization phase, the employed autism tutor taught the three DTT tasks with 100% accuracy which may not reflect the range of tutor abilities encountered in the field. The Generalization phase consisted of video files as opposed to live scoring, which would be required in the field. Only four of the six participants were available for the Follow-up sessions and only three of the six participants met the mastery criterion of 80% accuracy on all three tasks at Post-training. This may be due to the fast pace of a live session as there were a large number of missed components in these cases. In comparison to the previous phases, the number of missed components during the Follow-up phase was particularly high for Participant 4 (e.g., up to 48/119). Perhaps a fluency criterion could be used in future studies to improve accuracy at Follow-up and in naturalistic settings. It is also likely that some components are easier to observe and score than others, so further research may want to evaluate which components are more difficult to score and include further training on these components. Finally, participants were scoring DTT between Post-training and Follow-up sessions during regular work hours, so observer drift is possible.

Because this is the first study evaluating the effectiveness of the DTTEF-SIM, future research

needs to replicate and extend the results of the current study, and also evaluate other methods used to train individuals to use DTT checklists. Replications should include additional participants in each condition, participants from other agencies, autism tutors with varying levels of experience in a Generalization phase, and children of varying functioning levels in a Generalization phase. Furthermore, since errors on specific components can differentially affect treatment outcomes (e.g., Carroll et al., 2013), recent research has suggested assessing the treatment integrity of individual components as opposed to a global integrity measure (Cook et al., 2015).

In summary, the current study demonstrated that the DTTEF-SIM can be used to teach autism senior tutors and autism consultants to evaluate the treatment integrity of DTT sessions conducted by a confederate instructor and confederate child with ASD, and an autism tutor and child with ASD. The results of this study are important in the application of EIBI and DTT as this training method was time efficient and effective. The current results may enable agencies providing behavioural services to ensure that DTT is being applied consistently and accurately, resulting in positive gains for their clients.

Key Messages From This Article

People with disabilities. The self-instructional package can train staff to evaluate the treatment integrity of discrete-trials teaching sessions to help ensure that clients are receiving effective treatment.

Professionals. The self-instructional package is an effective tool for teaching individuals to evaluate the treatment integrity of discrete-trials teaching sessions.

Policymakers. It is important that staff are administering discrete-trials teaching with integrity in early intervention programs. This paper provides information on an effective method that can be used to train such staff to evaluate treatment integrity with accuracy.

References

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

Babel, D., Martin, G., Fazzio, D., Arnal, L., & Thomson, K. (2008). Assessment of the reliability and validity of the discrete-trials teaching evaluation form. *Developmental Disabilities Bulletin, 36*, 67–80.

Carroll, R., Kodak, T., & Fisher, W. (2013). An evaluation of programmed treatment-integrity errors during discrete-trials instruction. *Journal of Applied Behavior Analysis, 46*, 379–394.

Cook, J., Subramaniami S., Brunon, L., Larson, N., Poe, S., & St. Peter, C. (2015). Global measures of treatment integrity may mask important errors in discrete-trial teaching. *Behavior Analysis in Practice*, *8*(1), 37–47.

DiGennaro Reed, F. D., & Codding, R. (2014). Advancements in procedural fidelity assessment and intervention: Introduction to the special issue. *Journal of Behavioral Education, 23*, 1–18.

Fazzio, D., Arnal, L., & Martin, G. (2007a). *Discrete-trials teaching evaluation form (DTTEF).* Unpublished manuscript, University of Manitoba,Winnipeg, Canada.

Fazzio, D., Arnal, L., & Martin, G. (2007b). *Discrete-trials teaching evaluation form (DTTEF): Scoring manual.* Unpublished manuscript, University of Manitoba, Winnipeg, Canada.

Fazzio, D., Arnal, L., & Martin, G. (2012). *Discrete-trials teaching evaluation form (DTTEF*) (revised 2012). Unpublished scoring form, University of Manitoba, Winnipeg, Canada.

Gresham, F. M., Gansle, K. A., & Noell, G. H. (1993). Treatment integrity in applied behavior analysis with children. *Journal of Applied Behavior Analysis, 26*, 257–263.

Holcombe, A., Wolery, M., & Snyder, E. (1994). Effects of two levels of procedural fidelity with constant time delay on children's learning. *Journal of Behavioral Education, 4*, 49–73.

Jeanson, B., Thiessen, C., Thomson, K., Vermeulen, R., Martin G., & Yu, C.T. (2010). Field-testing of the discrete-trials teaching evaluation form. *Research in Autism Spectrum Disorders,* 4, 718–723.

Martin, G., & Pear, J. (2015). *Behavior modification: What it is and how to do it?* (10th ed.). Upper Saddle River, NJ: Prentice Hall.

Matson, J. L., & Smith, K. R. (2008). Current status of intensive behavioural interventions for young children with autism and PDD-NOS. *Research in Autism Spectrum Disorder, 2*, 60–74.

Matson, J. L., & Sturmey, P. (2011). *International handbook on autism and pervasive developmental disorders: Autism and child psychopathology series*. New York, NY: Springerlink.

McIntyre, L., Gresham, F., DiGennaro, F., & Reed, D. (2007). Treatment integrity of school-based interventions with children with the Journal of Applied Behavior Analysis 1991–2005. *Journal of Applied Behavior Analysis, 40*, 659–672.

Peterson, L., Homer, A., & Wonderlich, S. (1982). The integrity of independent variables in behavior analysis. *Journal of Applied Behavior Analysis, 15*, 477–492.

Salem, S., Fazzio, D., Arnal L., Fregeau, P., Thomson, K., Martin, G., & 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.

Smith, T. (2001). Discrete trial training in the treatment of autism. *Focus on Autism and Other Developmental Disabilities, 16*, 86–92.

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

Thomson, K., Martin, G., Arnal, L., Fazzio, D., & Yu, C.T. (2009). 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., 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*, 1073–1082.

Wheeler, J. J., Baggett, B. A., Fox, J., & Blevins, L. (2006). Treatment integrity: A review of intervention studies conducted with children with autism. *Focus on Autism and Other Developmental Disabilities, 21*, 45–54.

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 on Developmental Disabilities,18*(3), 33–44.

Wightman, J., Martin, G., Fazzio, D., & Arnal, L. (2014). *The discrete-trials teaching evaluation form: Self-instructional manual*. Unpublished manuscript, University of Manitoba, Winnipeg, Canada.

Wightman, J., Yates, H., Martin, T., Pear, J., & Yu, C.T. (2013, October). Training students to implement discrete-trials teaching: An examination of treatment integrity. Poster presented at Manitoba Association for Behaviour Analysis' 8th Annual Conference, Winnipeg, MB.