EMPIRICAL REPORT

Global Measures of Treatment Integrity May Mask Important Errors in Discrete-Trial Training

James E. Cook • Shrinidhi Subramaniam •
Lashanna Y. Brunson • Nicholas A. Larson •
Susannah G. Poe • Claire C. St. Peter

Published online: 16 January 2015
© Association for Behavior Analysis International 2015

Behavior-analytic interventions must be implemented consistently and correctly to be highly effective (e.g., Allen and Warzak 2000; St. Peter Pipkin et al. 2010; Vollmer et al. 2008). The term *treatment integrity* describes the precision with which interventions are implemented (e.g., Peterson et al. 1982). Treatment integrity is a measure of the correspondence between a behavior-change plan and the execution of that plan.

Treatment integrity is an important measure for several reasons. First, clients have an ethical right to receive the treatment to which they consented. Divergence from a treatment plan, whether intentional or not, may cause harm. Second, ensuring that treatments are implemented as described decreases the possibility of false-negative treatment outcomes (e.g., Hagermoser Sanetti and Fallon 2011; Peterson et al. 1982; Vollmer et al. 2008). Low treatment integrity may lead to erroneous conclusions about the effectiveness of a treatment. Third, ensuring treatment integrity increases the extent to which individuals can be confident that the intended procedures produced behavior change, rather than some other variable (Peterson et al. 1982; Vollmer et al. 2008). Fourth, treatment integrity measures provide an objective, quantifiable metric of therapist performance that can be used as a basis for further training or support.

Treatment integrity has become a topic of study in its own right. Researchers have systematically manipulated levels of

integrity and demonstrated functional relationships between integrity and treatment outcomes (e.g., Carroll et al. 2013; DiGennaro Reed et al. 2011; St. Peter Pipkin et al. 2010; Wilder et al. 2006). Otherwise, effective treatments that are implemented with low treatment integrity may lose their efficacy (Carroll et al. 2013; DiGennaro et al. 2011; Fryling et al. 2012; St. Peter Pipkin et al. 2010; Wilder et al. 2006). Reducing treatment integrity has negative effects on skill-acquisition procedures (e.g., Carroll et al. 2013) and behavior-reduction procedures (e.g., St. Peter Pipkin et al. 2010). These pervasive negative effects highlight the importance of measuring treatment integrity during staff training and using appropriate measures of treatment integrity to guide training procedures and mastery criteria for staff performance.

In recent years, researchers have evaluated methods for teaching parents (Hardy and Sturmey 1994; Lafasakis and Sturmey 2007), teachers (Catania et al. 2009; Lerman et al. 2008; Pence et al. 2013; Sarokoff and Sturmey 2004), and paraprofessionals (LeBlanc et al. 2005) to implement a variety of behavior-analytic interventions. One frequently trained intervention is an effective, evidence-based form of systematic teaching called discrete-trial training (DTT). The “trials” of DTT arrange learning opportunities for the client and include presentations of antecedents (such as materials, instructions, and prompts) and consequences (a reinforcer and praise, or error correction) for responding. DTT is often included in early intensive behavioral interventions for children with autism spectrum disorders (ASD) to teach social and academic skills (Ahearn and Tiger 2013; Green 1996; Lovaas 1987; Smith 2001; Thomson et al. 2009).

Novice instructors can be trained to implement DTT with high integrity using behavioral skills training. Behavioral skills training packages increase the extent to which instructors accurately perform a variety of interventions with

J. E. Cook (✉) • S. Subramaniam • C. C. St. Peter
Psychology Department, West Virginia University, P.O. Box 6040,
Morgantown, WV 26506, USA
e-mail: jamesecook@gmail.com

L. Y. Brunson • N. A. Larson • S. G. Poe
Center for Excellence in Disabilities, West Virginia University, 959
Hartman Run Road, Morgantown, WV 26505, USA

generalization of skills across different programs (Lafasakis and Sturmey 2007) and settings (Nigro-Bruzzi and Sturmey 2010), and maintenance across time (Bolton and Mayer 2008). Koegel et al. (1977) used a behavioral skills training package including written instructions, video modeling, and focused feedback to train educators to implement DTT with high integrity. Although training was effective, it was labor and time intensive. Catania et al. (2009) streamlined training by simply providing a video model, which increased the precision with which untrained staff implemented DTT. Sarokoff and Sturmey (2004) used written and verbal instruction, modeling, rehearsal, and feedback in their training package and trained educators to implement DTT with near-perfect treatment integrity.

In all of the above-cited studies, the experimenters averaged treatment integrity across all components of an intervention. This kind of global integrity measure is frequently reported as the outcome measure of caregiver and staff training protocols (Catania et al. 2009; Hardy and Sturmey 1994; Lafasakis and Sturmey 2007; LeBlanc et al. 2005; Lerman et al. 2008; Sarokoff and Sturmey 2004). Global integrity is calculated by dividing the total number of correct responses made by the implementer across all treatment components by the total number of opportunities to respond and multiplying that value by 100. A global treatment integrity score for multi-component treatment packages is convenient to display visually and communicate verbally. Global scores also provide a “big picture” view of the accuracy with which the intervention was implemented.

Although global treatment integrity scores provide an overall quantification of integrity, such measures may not always be sufficient representations of implementation. Global scores may mask poor performance on individual components of a treatment package and allow consistent and frequent lapses in treatment integrity on particular treatment components to be inadvertently hidden. Thus, trainers may find it helpful to assess integrity within individual treatment components in addition to a single, global summary score. Assessing individual-component integrity allows supervisors to identify improvements in instructors’ integrity within particular critical components, even when global integrity is high. In this way, component integrity could inform training decisions and mastery criteria in staff-training protocols. Requiring staff to meet mastery criteria on each component of an intervention, rather than on a global treatment integrity score, may better ensure treatments are implemented with integrity.

Assessing integrity on individual components may be especially important because integrity failures on particular components can differentially affect treatment outcomes. Carroll et al. (2013) conducted a descriptive assessment of teacher-implemented DTT. The most common integrity error was delivering a consequence: Teachers failed to provide a preferred item following a correct response on 79 % of

opportunities. Other frequent errors included failing to deliver a controlling prompt, presenting inappropriate instructions, and repeating instructions. In their third experiment, Carroll and colleagues manipulated the integrity of these components and measured child outcomes. Across three conditions, the researchers made programmed integrity errors in implementing an individual component (i.e., delivering instructions, prompts, or reinforcers) in 8 of 12 DTT trials, but otherwise implemented DTT with high integrity. Skill acquisition of children with ASD was disrupted when therapists omitted controlling prompts, gave multiple instructions, or omitted reinforcers for correct responses. This study experimentally demonstrated the importance of ensuring individual components of DTT are implemented with high integrity.

Integrity on individual components of DTT has rarely been measured as a primary dependent variable during training procedures despite suggestions to report or monitor global and component integrity of behavioral interventions (Gresham 1989; Power et al. 2005; Vollmer et al. 2008). Further, an analysis of the relationship between individual-component measures and global measures of integrity has yet to be reported. The objectives of this study were to compare global and component treatment integrity scores and systematically assess how training and feedback differentially affect treatment integrity globally and across individual components of DTT.

Method

Participants and Setting

Four undergraduate therapists in a psychology program participated in the study. Therapists received course credit for working in a classroom that served seven children with behavior disorders. Therapists were expected to implement behavior intervention plans and conduct DTT sessions daily. Therapists had completed a class in behavior modification, but did not have previous experience implementing DTT. Michelle was a 21-year-old female who worked in the classroom for 6 h per week. Carrie was a 22-year-old female who worked in the classroom for 6 h per week. Leonard was a 21-year-old male who worked in the classroom for 12 h per week. Phillip was a 25-year-old male who worked in the classroom for 4 h per week. Therapists conducted sessions with an 8-year-old child who had been previously diagnosed with mild intellectual disability, attention deficit hyperactivity disorder, post-traumatic stress disorder, and phonological disorder. The child had experience learning through DTT in the classroom.

Sessions were conducted in the classroom at the child’s desk or at a table in a hallway outside the classroom. Sessions were recorded using a camera placed on a tripod. Data sheets, pens, flashcards, tangible items (high-preference

items we identified previously through preference assessments), and a one-page curriculum sheet detailing how to teach each target skill were present during sessions.

Response Measurement

The dependent measures were global and individual-component treatment integrity for nine components of DTT (shown in Table 1). Some components involved multiple responses (e.g., error correction). For these components, the therapist was required to implement all steps correctly for the component to be scored as correct for that trial. Observers used a modified version of the Discrete-Trials Teaching Evaluation Form (DTTEF; Fazzio et al. 2010) to calculate integrity from videos of sessions. In each session, the therapists conducted a 12-trial DTT program with the child. Sessions began when the therapist was seated at the table with the child and concluded when the therapist completed 12 trials or when the video ended, whichever occurred first. We calculated global integrity by dividing the total number of correctly performed DTT skills in a session by the total number of opportunities to implement each skill in the session and multiplying by 100. We calculated individual-component integrity by dividing the total number of correctly performed responses within a single component by the total number of opportunities for that component response within a session and multiplying by 100.

The mastery criteria were three consecutive sessions with global scores and reinforcer component scores of 80 % or greater. We used dual mastery criteria targeting global and reinforcer-component integrity because performance on the reinforcer component was consistently low across all four therapists during baseline. Furthermore, Carroll et al. (2013) demonstrated that errors in consequence delivery impeded the skill acquisition of children. Therefore, we believed improvements in reinforcer integrity to be clinically indicated.

Interobserver Agreement Two independent observers scored the same videos to assess interobserver agreement (IOA). IOA was calculated by dividing the total number of agreements by the total number of agreements and disagreements and multiplying by 100. An agreement was defined as two independent observers scoring the same component in a given DTT trial as being implemented correctly, incorrectly, or as not applicable. Agreement was calculated for 34 % of sessions for Michelle and averaged 96.5 % (range 89 to 100 %). Agreement was calculated for 32 % of sessions for Carrie and averaged 94.9 % (range 86.7 to 100 %). Agreement was calculated for 40 % of Leonard's sessions and averaged 92.4 % (range 80.3 to 100 %). Agreement was calculated for 32 % of Phillip's sessions and averaged 93 % (range 84.4 to 98.3 %).

Table 1 Description of discrete trial training components

Discrete trial training component	Component description
A. Attention	Therapist praises attentive behavior or provides a vocal prompt and a model every 3 s to secure student's attention
B. Materials	Therapist presents session materials as described in curriculum sheet
C. Instruction	Therapist presents the correct instruction described in curriculum sheet once per trial
D. Prompt	Therapist delivers the prompt described in curriculum sheet that corresponds to that written on the data sheet or delivered in the first trial of the session
E. Praise	Therapist says a praise statement following a correct response
F. Tangible reinforcer	Therapist delivers the item selected by the student following a correct response; student granted access to item for 15–45 s
G. Error correction	Following an incorrect response, therapist: <ul style="list-style-type: none"> • Makes no statement to student • Removes session materials • Removes eye contact for 3 s • Re-secures student attention • Re-presents session materials and instruction • Presents most intrusive prompt • Provides praise only
H. Record data	Therapist records a correct response following reinforcer delivery; therapist records an incorrect response while eye contact is removed during error correction
I. Inter-trial interval	Therapist begins next trial no more than 5 s after the reinforcer is removed or praise is delivered in error correction

Experimental Design We used a multiple-baseline-across-participants design to evaluate the effects of training procedures on global and individual-component treatment integrity scores.

Procedure

We instructed therapists to videotape themselves conducting 12-trial DTT sessions as part of their regularly scheduled classroom duties. The number of sessions conducted each day by therapists varied because circumstances in the classroom (e.g., problem behavior, field trips, other scheduled activities, etc.) affected the number of opportunities a therapist had to conduct sessions ($M=2.5$ sessions per day, range 1 to 11 sessions per day). Classroom teachers provided therapists with the necessary materials to run DTT programs, including flashcards, pens, data sheets, a one-page curriculum sheet, and potential reinforcers. At the end of the day,

we collected session videos and calculated treatment integrity. We conducted all sessions in this manner unless otherwise noted.

All the therapists taught the child expressive letter sounds from the same set of targets using delayed vocal model prompts. The three targets therapists taught the child in a given session changed when the child mastered targets (i.e., three consecutive sessions with greater than 90 % accuracy in the absence of vocal model prompts). In the event the child engaged in disruptive behavior (i.e., noncompliance, loud vocalizations, property destruction, aggression), we instructed therapists to follow the child's behavior intervention plan. The child engaged in disruptive behavior during 14 % of all sessions.

We set mastery criteria for therapist performance at three consecutive sessions with 80 % or greater treatment integrity on global and reinforcer-component scores. If therapist performance did not meet these criteria and was deemed stable by visual analysis, we implemented additional interventions (described below) to improve treatment integrity. We started with interventions that consumed the least amount of trainer time (e.g., watching a video) and progressed to interventions requiring the most amount of trainer time (e.g., live behavioral skills training).

Instructions (Baseline) One day prior to conducting sessions, we emailed a 30-page written instruction manual that described how to implement the individual components of DTT to each therapist. The manual was written at an 8th-grade reading level and included pictures of therapists modeling correct DTT techniques. We did not verify that therapists reviewed the manual before working with the child and did not provide performance feedback to therapists during the baseline phase.

Video We provided therapists with a video model (cf. Catania et al. 2009) to improve performance on the reinforcer component. Therapists viewed a 2-min video that focused on reinforcer identification and delivery. The video consisted of a member of the research team sitting with a child and modeling correct identification of a preferred item, delivering the reinforcer after a correct response, and retrieving the reinforcer after 15–45 s of access. The narrator described the rationale for implementing each portion of the reinforcer component correctly. Written text and still photos highlighted important details such as arranging items evenly in front of the child when identifying a reinforcer to use in the session and the duration of reinforcer access following a correct response. We included this phase to assess whether a video model focused on the reinforcer component would improve performance on that individual component, and to what extent this improvement would generalize to other DTT skills.

Therapists viewed the video individually in the presence of a trainer. They were only allowed to view the video once, and the trainer did not answer any questions about the video. We did not provide feedback on therapist performance in the video phase.

Feedback After reviewing videos from each session, a trainer emailed two to three pages of written performance feedback to the therapist prior to their next session. The feedback included praise for components implemented correctly and constructive feedback for components implemented incorrectly. Constructive feedback included pictures of trainers implementing problematic components of DTT correctly, brief descriptions of correct implementation, and a rationale for performing those components correctly.

Behavioral Skills Training (BST) Leonard and Phillip failed to reach the mastery criteria following written feedback and received behavioral skills training. Leonard participated in a regularly scheduled staff-training session. Phillip was unable to participate in this staff training due to scheduling conflicts, so he received a modified behavioral skills training package.

Leonard Leonard experienced an extended absence from the classroom after viewing the video, and we were unable to collect data on his performance following that component. To rapidly improve his treatment integrity, Leonard participated in behavioral skills training that consisted of modeling, role play, and feedback during mock DTT sessions with a confederate. We conducted the training in the classroom when no children were present. The trainer gave Leonard a one-page curriculum sheet and all the materials necessary to conduct a DTT session, and asked him to conduct a session to the best of his ability with a confederate (i.e., a second trainer acting as a student). The confederate followed a script and responded correctly, incorrectly, or emitted no response in an equal number of trials each session.

The trainer collected in vivo data on Leonard's performance. After each session, the trainer praised Leonard for components implemented correctly and modeled and rehearsed with Leonard components that were implemented incorrectly. Sessions with the confederate continued until Leonard completed three consecutive sessions with global and reinforcer-component integrity scores of 90 % or greater. Following training with the confederate, the trainer asked Leonard to conduct and videotape sessions with the child as previously described.

Phillip We used in situ behavioral skills training to improve Phillip's performance. A researcher observed as Phillip conducted a session with the child. During the session, the researcher provided praise for components implemented

correctly. When Phillip implemented a component incorrectly, the researcher stopped Phillip at the end of the trial and modeled and rehearsed the component with Phillip. The researcher provided feedback at the end of the session and answered any questions Phillip had. Training continued until Phillip completed three consecutive sessions with global and reinforcer-component integrity scores of 90 % or greater. Upon completion of behavioral skills training, Phillip was instructed to conduct and videotape sessions with the child as described above.

Results

Figure 1 shows average individual-component scores for Michelle and Carrie, who met mastery criteria during the feedback condition. Average global scores for the last three sessions of each condition are shown by the dashed horizontal lines. The bars show treatment integrity for each individual component for the last three sessions of each condition. Bars are arranged from the lowest to the highest individual-

component treatment integrity score during the instructions condition. During the instructions condition, global integrity scores were 57.9 % for Michelle and 71.0 % for Carrie. These global integrity scores were not representative of performance on individual components of the treatment. Average individual-component scores ranged from 0 to 100 % for Michelle and from 25.0 to 94.5 % for Carrie. During the video condition, global integrity scores were 65.2 % for Michelle and 55.6 % for Carrie. These global integrity scores remained unrepresentative of performance on individual components of the treatment. Average individual-component scores ranged from 0 to 97.2 % for Michelle and from 0 to 94.5 % for Carrie. During the feedback condition, global integrity scores were 95.8 % for Michelle and 86.4 % for Carrie. These global integrity scores were representative of performance on most individual components of the treatment. Average individual-component scores ranged from 77.8 to 100 % for Michelle and from 50 to 94.4 % for Carrie.

Figure 2 shows average individual-component scores for Leonard and Phillip, who met mastery criteria following behavioral skills training. Average global scores for the last three sessions of each condition are shown by the dashed horizontal

Fig. 1 Average treatment integrity across training conditions for Michelle and Carrie. Bars represent component integrity and dashed line represents global integrity

lines. The bars show treatment integrity for each individual component for the last three sessions of each condition. Bars are arranged from the lowest to the highest individual-component treatment integrity score during the Instructions condition. During the instructions condition, global integrity scores were 70.2 % for Leonard and 63.5 % for Phillip. These global integrity scores were not representative of performance on individual components of the treatment. Average individual-component scores ranged from 0 to 100 % for Leonard and from 0 to 93.1 % for Phillip. During the video condition, global integrity scores were 61.2 % for Phillip. These global integrity scores were not representative of performance on individual components of the treatment. Average individual-component scores ranged from 0 to 75 % for Phillip. We were unable to collect data on Leonard's performance after he watched the video because he experienced an extended absence from the classroom. During the feedback condition, global integrity scores were 86.5 % for Phillip. These global integrity scores remained unrepresentative of performance on individual components of the treatment. Average individual-component scores ranged from 29.9 to

97.2 % for Phillip. During the BST condition, global integrity scores were 98.1 % for Leonard and 94.7 % for Phillip. These global integrity scores were representative of performance on individual components of the treatment. Average individual-component scores ranged from 83.3 to 100 % for Leonard and from 87.5 to 100 % for Phillip.

As can be seen in Figs. 1 and 2, average global integrity scores generally increased across conditions for all therapists, but the range of individual component integrity remained between 0 and 100 % until the last condition for each therapist. Only in the last condition for each therapist did the lowest component integrity score increase above zero. These data show that increases in global integrity did not represent an increase in performance across all individual components, and performance in individual components remained extremely variable.

Even when global integrity scores exceeded the mastery criteria (in the feedback condition for Michelle and Carrie, and in BST for Leonard and Phillip), global integrity was not necessarily predictive of high individual-component integrity for all therapists. Michelle's global integrity scores

Fig. 2 Average treatment integrity across training conditions for Leonard and Phillip. Bars represent component integrity and dashed line represents global integrity

averaged 95.8 % for the last three sessions of the feedback condition, but she only secured attention with 77.8 % accuracy. Likewise, Carrie's global integrity scores averaged 86.4 % for the last three sessions of the feedback condition, but she only implemented the error-correction procedure with 50 % integrity. Phillip's global integrity scores averaged 86.5 % for the last three sessions of the feedback condition, but he secured attention, recorded data, delivered reinforcers, and implemented the error correction procedure with 29.9, 37.5, 37.5, and 58.3 % integrity, respectively. This kind of discrepancy may be particularly important, given that low integrity on error-correction procedures is known to have detrimental effects on DTT outcomes (e.g., Carroll et al. 2014). High global integrity scores did correspond with high integrity scores on individual components for Leonard and Phillip after BST. When Leonard and Phillip met the mastery criterion of three consecutive sessions with at least 80 % global integrity after BST, their obtained global integrity scores averaged 98.1 and 94.7 %, respectively, and integrity on all individual components exceeded 80 %.

Figure 3 depicts session-by-session data for global and reinforcer-component integrity across training conditions for Michelle and Carrie. Though Michelle showed some global integrity improvement during the video condition, neither Michelle nor Carrie consistently and accurately implemented the reinforcer component until the feedback condition.

Figure 4 depicts session-by-session data for global and reinforcer-component integrity across training conditions for Leonard and Phillip. Leonard only consistently implemented the reinforcer component correctly following BST. Phillip's global integrity scores remained consistent across the instructions and video conditions, while reinforcer-component integrity scores remained at zero. During the feedback condition, Phillip's global and reinforcer-component integrity scores became more variable, but changes in one measure did not necessarily correspond with changes in the other. Phillip's performance globally and on the reinforcer component only met the mastery criteria following BST, but his performance dropped in the reinforcer component on the final session.

Discussion

We trained undergraduate therapists to implement DTT in a school setting using instructions and individualized feedback. We assessed the therapists' ability to implement DTT with integrity using global and individual-component treatment integrity scores. Global treatment integrity scores were not necessarily representative of individual-component treatment integrity scores. Specifically, we found that therapists frequently implemented the reinforcer component with low integrity. This result replicates outcomes obtained by Carroll et al. (2013) and extends them to a different population.

Fig. 3 Treatment integrity across training conditions for Michelle and Carrie. Filled data points represent reinforcer component integrity and unfilled data points represent global integrity

Treatment integrity errors negatively affect child outcomes in a number of interventions (Carroll et al. 2013; DiGennaro et al. 2011; Fryling et al. 2012; St. Peter Pipkin et al. 2010; Wilder et al. 2006). In DTT, reduced integrity of the delivery of a reinforcer can result in children failing to acquire targeted skill or requiring more extensive training to acquire those skills (Carroll et al. 2013). Unfortunately, we were unable to directly assess effects of reduced integrity on child performance in our study because each participant worked on the same program with the child, as did other staff who did not participate in this study. Child performance with one participant may have been impacted by the child's experience with other participants and staff. Despite this limitation, the results of the current study, combined with those obtained by Carroll and colleagues, highlight the importance of measuring treatment integrity on individual components in addition to the use of more global measures. Mastery criteria can be applied to integrity on individual components instead of, or in addition to, the global score. Component-based mastery criteria can increase confidence that critical features of the intervention are correctly implemented and allow supervisors to more precisely target skill deficits and provide individualized training

Fig. 4 Treatment integrity across training conditions for Leonard and Phillip. Filled data points represent reinforcer component integrity and unfilled data points represent global integrity

and feedback. Future research should examine the merits of basing mastery criteria on individual components and should identify training procedures that promote rapid acquisition of skills across components.

Our global mastery criteria included a lower global treatment integrity score (i.e., 80 versus 90 %) but more sessions (e.g., 3 versus 2) than other studies assessing effects of behavioral skills training on treatment integrity in DTT (Lafasakis and Sturmey 2007; Sarokoff and Sturmey 2004). An 80 % global treatment integrity score may not be the ideal level of precision to include in mastery criteria. More stringent global mastery criteria (e.g., 95–100 % global scores) may ensure higher individual-component integrity scores. For example, basing mastery criteria on 100 % global treatment integrity would ensure that therapists were implementing each component with perfect integrity. Then again, such stringent mastery criteria may be difficult for staff to meet for consecutive sessions and thereby dramatically increase training time. Future research should be conducted to determine how to balance decisions about acceptable treatment integrity, the acceptable minimum number of sessions that score is observed, and the efficiency of training.

Treatment integrity checklists that clinicians use to assess teacher behavior should also be carefully considered. The published literature on training staff to conduct DTT have used a variety of checklists (e.g., Fazzio et al. 2010; Severtson and Carr 2012). These checklists include slightly different definitions of components and, in some cases, scored components of the intervention that we excluded (e.g., dealing with problem behavior; Carroll et al. 2013). The extent to which staff implement an intervention with high integrity globally and on each individual component of an intervention with high integrity may be influenced by the complexity of the component scored. We based our components on an empirically evaluated checklist (DTTEF; Fazzio et al. 2010). This checklist (see Table 1) includes some single-step (e.g., instruction) and some multi-step (e.g., error correction) components. For example, the reinforcer component required that the instructor give a choice between reinforcers at the beginning of the session, deliver the selected reinforcer immediately after a correct child response, and remove the reinforcer after 15–45 s of access. All of these steps must be implemented correctly for the component to be scored as correct on a given trial. The error-correction component required that, following the recognition of an error, the instructor engage in an eight-step behavior chain. It is possible that using multiple steps to define a single component is partially responsible for the low integrity observed for some components.

Researchers and clinicians should agree on the core components of DTT before undertaking component-based performance criteria. Carroll et al. (2013) suggest correct delivery of instructions, prompts, and reinforcers are critical components. Additional research analyzing effects of skill complexity on treatment integrity would increase understanding of how to define DTT components. Moreover, research directly comparing scoring checklists and differing DTT procedures may inform which components are necessary and sufficient for DTT to produce significant changes in child behavior.

Measuring integrity on individual intervention components may have the additional advantage of permitting supervisors to specifically target the components critical for effective behavior change. For example, the delivery of a reinforcer following a correct response seems to be a critical component of DTT (Carroll et al. 2013). Thus, it may be important for supervisors to know when integrity is low on this component and to conduct immediate and targeted trainings to improve performance. Our attempt to use a video model (cf. Catania et al. 2009) designed to improve performance on the reinforcer component was not successful for any of the three participants that viewed it, but written feedback and behavioral skills training that praised components implemented correctly and provided feedback on components implemented incorrectly were effective at improving treatment integrity. Future research should examine the extent to which training on individual

components of an intervention generalizes to untrained components or results in increases in child skill acquisition.

The reliability of the independent variables was not assessed across the study. We provided participants with written instructions and feedback, but did not take measures to ensure that the participants read them. Trainers ensured that participants viewed the video, but did not require that participants answer any questions about the video to assess whether they attended to its content. Additionally, we did not conduct reliability checks on components of BST. However, participant performance only improved following written feedback and BST, despite teachers and other therapists being available to model or provide feedback on implementing DTT throughout the study.

Implications for Practitioners

When training and supervising staff, we suggest the following strategy for monitoring and improving treatment integrity. This strategy is displayed in Fig. 5 and can be applied to a variety of multi-component behavioral interventions and different levels of staff experience. First, we suggest calculating global integrity scores by collapsing across components of the intervention and examining the total percentage of opportunities implemented correctly. Global scores are useful because they provide an overall summary of how well an intervention is being implemented. Low global integrity scores would clearly indicate a need for further training or staff support. On the other hand, global scores do not always indicate how

well individual components of the intervention are performed. Therefore, we also recommend calculating individual-component integrity scores by examining the percentage of opportunities each component is implemented correctly. Calculating and graphing individual-component integrity scores allows for a more fine-grained analysis of how treatment integrity might affect clients than the use of global scores alone. Calculating individual-component scores may not be necessary if trainers are using an exceptionally high global score mastery criterion (e.g., 95–100 %), as individual-component scores would also need to be high to meet this mastery criterion.

Supervisors should use the data from individual-component integrity scores to inform decisions about providing feedback or additional skills training. More time and effort during training should be allocated to components with persistent integrity errors. Individual-component integrity graphs can be used as specific performance feedback when providing supervision to therapists. Time and effort during supervision should also be allocated to components that have been empirically demonstrated to increase desired behavior (e.g., delivering the reinforcer, using controlling prompts, giving the correct instruction; Carroll et al. 2013).

Supervisors should conduct regular checks of treatment integrity on all components of an intervention and individualize feedback to trainees' specific integrity lapses. As recommended in previous literature (McIntyre et al. 2007; Vollmer et al. 2008), we suggest scheduling brief observations during a small percentage of sessions (e.g., 15–25 %) or during a scheduled period (e.g., once or twice per week). The duration of these observations will differ depending on factors such as session duration, supervisor availability, and number of clients present.

During observations, we recommend the use of component-by-component treatment-integrity monitoring sheets. Live or in-vivo observations and feedback are ideal; however, for supervisors who cannot be present for sessions, remote forms of observation (e.g., recorded video) and feedback (e.g., written feedback or videos) may be required. Results of the present experiment suggest that if persistent lapses in integrity occur during observations, individualized feedback, even in the form of delayed written feedback, may be sufficient to improve performance. When additional training is needed and using confederates is not an option, in situ training involving modeling, role play, and feedback may improve performance.

As an agency or practitioner, it is important to develop training procedures that fit the available resources while producing quality therapists. Taking a fine-grained look at treatment integrity and addressing specific lapses will help build this quality. Data on staff progress can also be used to identify whether the training procedures adopted by the agency or practice are appropriate and to make improvements if necessary.

Fig. 5 Strategy for integrity monitoring and improvement consists of analyzing integrity globally and for individual intervention components, followed by either training or feedback to improve performance or periodic checks for maintenance of integrity

Ultimately, examining treatment integrity across all components of an intervention can provide important information about the effectiveness of an intervention in bringing about positive behavior change. A high level of treatment integrity in all intervention components would ensure clients are receiving the services to which they consented and would allow practitioners to make informed decisions about whether to continue or modify an intervention for a client (Hagermoser et al. 2011). Constant monitoring of treatment integrity is thus beneficial on a micro (therapist and client) and macro (agency or practice) level.

Acknowledgments The authors would like to thank Aimee Giles and Keegan Kowcheck for their assistance in developing the materials. This study was supported in part by grant R40 MC 20444 from the Maternal and Child Health Bureau (Combating Autism Act of 2006), Health Resources and Services Administration, Department of Health and Human Services.

References

- Ahearn, W. H., & Tiger, J. H. (2013). Behavioral approaches to the treatment of autism. In G. J. Madden (Ed.), *APA handbook of behavior analysis* (Vol. 2, pp. 301–327). Washington: American Psychological Association.
- Allen, K. D., & Warzak, W. J. (2000). The problem of parental nonadherence in clinical behavior analysis: effective treatment is not enough. *Journal of Applied Behavior Analysis*, 33(3), 373–391. doi:10.1901/jaba.2000.33-373.
- Bolton, J., & Mayer, M. D. (2008). Promoting the generalization of paraprofessional discrete trial teaching skills. *Focus on Autism and Other Developmental Disabilities*, 23(2), 103–111. doi:10.1177/1088357608316269.
- 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. doi:10.1002/jaba.49.
- Carroll, R. A., Joachim, B., Robinson, N., & St. Peter, C. C. (2014). A comparison of different error-correction approaches on skill acquisition during discrete-trial instruction. Poster presented at the 40th annual convention of the Association for Behavior Analysis International, Chicago, IL.
- 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. doi:10.1901/jaba.2009.42-387.
- DiGennaro Reed, F. D., Reed, D. D., Baez, C. N., & Maguire, H. (2011). A parametric analysis of errors of commission during discrete-trial training. *Journal of Applied Behavior Analysis*, 44(3), 611–615. doi:10.1901/jaba.2011.44-611.
- Fazio, D., Arnal, L., & Martin, G. (2010). Discrete-trials teaching evaluation form (DTTEF) Scoring Manual. Unpublished manuscript. Retrieved from: <http://www.dtteaching.com/>.
- Fryling, M. J., Wallace, M. D., & Yassine, J. N. (2012). Impact of treatment integrity on intervention effectiveness. *Journal of Applied Behavior Analysis*, 45(2), 449–453. doi:10.1901/jaba.2012.45-449.
- Green, G. (1996). Early behavioral intervention for autism. In C. Maurice, G. Green, & R. Foxx (Eds.), *Behavioral intervention for young children with autism: a manual for parents and professionals* (pp. 29–44). Austin: Pro-Ed.
- Gresham, F. M. (1989). Assessment of treatment integrity in school consultation and prereferral intervention. *School Psychology Review*, 18, 37–50.
- Hagermoser Sanetti, L. M., & Fallon, L. M. (2011). Treatment integrity assessment: how estimates of adherence, quality, and exposure influence interpretation of implementation. *Journal of Educational and Psychological Consultation*, 21(3), 209–232. doi:10.1080/10474412.2011.595163.
- Hardy, N., & Sturmey, P. (1994). Portage guide to early education, III: a rapid training and feedback system to teach and maintain mothers' teaching skills. *Educational Psychology*, 14(3), 345–357. doi:10.1080/0144341940140308.
- 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. doi:10.1901/jaba.1977.10-197.
- Lafasakis, M., & Sturmey, 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. doi:10.1901/jaba.2007.685-689.
- 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.
- Lerman, D. C., Tetreault, A., Hovanetz, A., Strobel, M., & Garro, J. (2008). Further evaluation of a brief, intensive teacher-training model. *Journal of Applied Behavior Analysis*, 41(2), 243–248. doi:10.1901/jaba.2008.41-243.
- Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual function in young autistic children. *Journal of Consulting and Clinical Psychology*, 55(1), 3–9. doi:10.1037/0022-006X.55.1.3.
- McIntyre, L. L., Gresham, F. M., DiGennaro, F. D., & Reed, D. D. (2007). Treatment integrity of school-based interventions with children in the *Journal of Applied Behavior Analysis* 1991–2005. *Journal of Applied Behavior Analysis*, 40(4), 659–672. doi:10.1901/jaba.2007.659-672.
- Nigro-Bruzzi, D., & Sturmey, P. (2010). The effects of behavioral skills training on mand training by staff and unprompted vocal mands by children. *Journal of Applied Behavior Analysis*, 43(4), 757–761. doi:10.1901/jaba.2010.43-757.
- Pence, S. T., Peter, S., Claire, C., & Giles, A. F. (2013). Teacher acquisition of functional analysis methods using pyramidal training. *Journal of Behavioral Education*. doi:10.1007/s10864-013-9812-4.
- Peterson, L., Homer, A. L., & Wonderlich, S. A. (1982). The integrity of independent variables in behavior analysis. *Journal of Applied Behavior Analysis*, 15(4), 477–492. doi:10.1901/jaba.1982.15-477.
- Power, T. J., Blom-Hoffman, J., Clarke, A. T., Riley-Tillman, T. C., Kelleher, C., & Manz, P. H. (2005). Reconceptualizing intervention integrity: a partnership-based framework for linking research with practice. *Psychology in the Schools*, 42(5), 495–507. doi:10.1002/pits.20087.
- 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(4), 535–538.
- 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.
- Smith, T. (2001). Discrete trial training in the treatment of autism. *Focus on Autism and Other Developmental Disabilities*, 16(2), 86–92.
- St. Peter Pipkin, C. C., Vollmer, T. R., & Sloman, K. N. (2010). Effects of treatment integrity failures during differential reinforcement of

- alternative behavior: a translational model. *Journal of Applied Behavior Analysis*, 43(1), 47–70.
- Thomson, K., Martin, G. L., 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(3), 590–606.
- Vollmer, T. R., Sloman, K. N., & St. Peter Pipkin, C. C. (2008). Practical implications of data reliability and treatment integrity monitoring. *Behavior Analysis in Practice*, 1(2), 4–11.
- Wilder, D. A., Atwell, J., & Wine, B. (2006). The effects of varying levels of treatment integrity on child compliance during treatment with a three-step prompting procedure. *Journal of Applied Behavior Analysis*, 39(3), 369–373.