

# The Effects of Discrete-Trial Training Commission Errors on Learner Outcomes: An Extension

Sarah R. Jenkins · Jason M. Hirst ·  
Florence D. DiGennaro Reed

Published online: 25 October 2014  
© Springer Science+Business Media New York 2014

**Abstract** We conducted a parametric analysis of treatment integrity errors during discrete-trial training and investigated the effects of three integrity conditions (0, 50, or 100 % errors of commission) on performance in the presence and absence of programmed errors. The presence of commission errors impaired acquisition for three of four participants. Following removal of commission errors from instruction, we observed a carry-over effect for only two participants suggesting that adverse effects are reversible with relatively small amounts of high-quality instruction.

**Keywords** Treatment integrity · Errors of commission · Discrete-trial training

## Introduction

Behavioral treatment of autism represents an established literature base with over 200 studies documenting its effectiveness (National Standards Report 2009). As a result, a number of entities endorse behavioral approaches to autism intervention (e.g., National Research Council 2001). Because published studies commonly involve researchers who ensure sufficient adherence to instructional protocols, we know little about the effectiveness of behavioral treatment, particularly discrete-trial training, when educators implement treatment poorly. Unfortunately, previous research has shown that educators do not always implement instructional protocols accurately and consistently (e.g., Dib and Sturmey 2007).

The degree to which a treatment is implemented consistently and accurately is known as *treatment integrity* (Gresham 1989). A handful of studies have demonstrated

---

S. R. Jenkins · J. M. Hirst · F. D. DiGennaro Reed (✉)  
Department of Applied Behavioral Science, 4056 Dole Human Development Center, University of  
Kansas, 1000 Sunnyside Avenue, Lawrence, KS 66045-7555, USA  
e-mail: fdreed@ku.edu

that degradations in the integrity of teaching procedures impair performance (e.g., Carroll et al. 2013; Groskreutz et al. 2011; Grow et al. 2009). In a recent translational study, Hirst et al. (2013) demonstrated that a history of performance feedback errors was associated with delayed acquisition even when errors were corrected, suggesting that integrity errors may produce long-term negative influences on learning for some participants. In the experiment, undergraduate students completed a computerized arbitrary match-to-sample task. Initially, participants encountered errors in the feedback following responses, with errors defined as receiving feedback that the response was correct following an incorrect response, or vice versa. Hirst et al. programmed feedback errors to occur during a proportion of trials to simulate 25, 50, 75, and 100 % integrity. The four levels of integrity were compared in a between-groups design. Within the groups, participants experienced one of the three imperfect levels of integrity followed by a high integrity phase. Analysis of acquisition revealed a weak linear relation with integrity, for which participants exposed to lower levels of integrity showed lower acquisition than participants exposed to higher levels of integrity. When a phase of high integrity feedback followed imperfect integrity, many participants achieved mastery, but only after a substantial delay. Several participants did not reach mastery prior to the end of the study.

In a subsequent study, Hirst and DiGennaro Reed (2014) replicated these effects with a larger sample of undergraduate students and then extended the findings to typically developing preschool-age children. In the extension, children were taught four receptive identification skills using an interspersed, discrete-trial training procedure with varying levels of feedback errors. In a multielement design, each task was associated with either 0, 25, 50, or 75 % feedback errors. The errors were combined errors of commission and omission in that participants encountered errors both in the form of positive feedback for incorrect responses and in the omission of positive feedback for correct responses. Results indicated that the participants only mastered the task associated with 0 % error in the feedback procedure. In a second condition, errors were removed and participants met mastery for the tasks previously associated with feedback errors, but this typically occurred after a delay.

Leon et al. (2014) conducted two translational studies involving children to evaluate the effects of integrity errors on child compliance. Their experimental preparation adopted a reversal design that allowed an evaluation of the influence of previous integrity errors on subsequent performance. The first study incorporated a parametric analysis involving experimenter omission of a preferred edible contingent upon compliance with relinquishing a toy. Programmed omission errors occurred for a proportion of trials to achieve 100, 60, and 20 % integrity conditions. Leon et al. showed that child compliance varied according to the integrity level of the differential reinforcement procedure. Compliance was highest during 100 % integrity and lowest during 20 % integrity and baseline during which the experimenters did not provide the edible. An interesting outcome occurred during the 60 % integrity condition: the level of integrity during the *previous* condition influenced compliance (i.e., demonstration of a sequence effect). When the preceding condition was baseline, compliance was relatively higher during the 60 % condition. However, when the preceding condition consisted of 100 % integrity, performance during the 60 % integrity was relatively lower, more variable, or both.

Their second study involved a similar experimental preparation, but the parametric analysis involved commission errors during which the experimenter delivered an edible and praise during a proportion of trials for noncompliance with relinquishing the toy (100 and 0 % integrity). Compliance also produced an edible and praise. Participants emitted low compliance during 0 % integrity and baseline (during which there was no reinforcement for compliance) and high compliance during 100 % integrity. The researchers conducted a supplemental analysis with one participant who showed low compliance during 50 % integrity regardless of the previous condition. Across both studies, Leon et al. replicated previous research by documenting that compliance was influenced by the integrity of the differential reinforcement procedure in place for both omission and commission errors. In addition, they showed that compliance might be influenced by preceding conditions (i.e., a sequence effect for omission errors). The latter finding suggests that one's treatment integrity history may influence subsequent performance, but this finding is not universal (e.g., Northup et al. 1997; St. Peter Pipkin et al. 2010).

We were only able to find two studies that evaluated the degree to which differing levels of treatment integrity of discrete-trial training methods influence performance of children with autism (Carroll et al. 2013; DiGennaro Reed et al. 2011). DiGennaro Reed et al. (2011) conducted a parametric analysis with three children with autism to evaluate the effects of treatment integrity level on the acquisition of nonsense shapes. The parametric analysis incorporated commission errors, defined as providing reinforcement in the form of praise and tokens following a proportion of incorrect responses to the task. Three integrity levels were evaluated (100, 50, and 0 % errors of commission). Participant performance was highest in the 0 % errors condition and lower when instruction included commission errors. These findings are consistent with previous studies and support the conclusion that researchers generally achieve the best outcomes under the highest level of integrity examined. Given that educators sometimes may make errors during implementation of treatment procedures (Carroll et al. 2013; Dib and Sturmey 2007), an important area of research involves identifying the potential long-term impact of integrity errors on performance. Unfortunately, there are few studies examining this issue. In perhaps one of the most elegant investigations of treatment integrity errors to date, Carroll et al. (2013) conducted a series of studies to evaluate the effects of integrity errors on skill acquisition during discrete-trial training. Their first study consisted of a descriptive assessment of integrity errors made by teachers or paraprofessionals during discrete-trial training in one-on-one or small group teaching arrangements. They found that the highest percentage of errors occurred during contingent delivery of tangible items for correct responses followed by errors during delivery of the controlling prompt and one-time presentation of the instruction. In a second and third study, they manipulated the levels of integrity for the three most common integrity errors from the descriptive assessment. Their findings supported previous research and documented that integrity errors influence acquisition. Acquisition was greatest during high integrity instruction. Interestingly, participants occasionally met the mastery criterion when instruction contained errors, but required many more sessions to reach criterion compared to acquisition rates during high integrity instruction. To explain the latter finding, the authors

speculated that learning might be more substantially and negatively impacted when errors occur in combination (Study 2) versus in isolation (Study 3) or that the stimuli and responses targeted for training were more difficult for some participants resulting in less learning. All of the participants exposed to a phase of high integrity following a training phase containing integrity errors showed improved performance during high integrity; however, two of these participants demonstrated slight delays to acquisition.

Taken together, these collective findings suggest that performance is generally better when treatment integrity is high (Carroll et al. 2013; DiGennaro Reed et al. 2011; Leon et al. 2014), but that learning may occur despite instructional errors (Carroll et al. 2013). Findings from the small number of studies that exist also document that performance is influenced by *previous* treatment integrity of a procedure (Hirst and DiGennaro Reed 2014; Hirst et al. 2013; Leon et al. 2014), but this finding is idiosyncratic across participants (Carroll et al. 2013), skills (Carroll et al. 2013), and type of treatment integrity error (Leon et al. 2014). Thus, the purpose of the present study was to contribute to this small, but emerging literature. Specifically, we evaluated the short- and long-term effects of imperfect integrity during discrete-trial training on learning of students with autism.

## Methods













### Participants and Setting

We recruited four boys (Caleb, Donovan, Elin, and Felix) with autism receiving educational services from a private school in the Midwest to participate in the study. All participants had received a diagnosis of autism from independent professionals in their community prior to the study. Participants were 3–7 years old, had vocal communication skills, intact visual skills, and were familiar with discrete-trial training and delayed reinforcement. The school provided the most recent annual scores from the Kaufman Assessment Battery for Children, Second Edition (KABC; Kaufman and Kaufman 2004) as well as the total scores from the Autism Diagnostic Observation Schedule (ADOS; Lord et al. 1999), which was administered at the time of diagnosis. Caleb's KABC score was 71 (below average), and ADOS score was 21. Donovan's KABC score was 125 (above average), and ADOS score was 13. Elin and Felix both scored in the average range for the KABC (110 and 109, respectively), and their ADOS scores were 19 and 20, respectively. Caleb and Donovan had prior experience with a token economy. All sessions took place in the participants' classrooms, which contained up to four children and four teachers.

### Materials

We created a match-to-sample task using modified Hiragana characters (i.e., Japanese alphabet). The study included three target shapes each associated with a

**Table 1** Array of target shapes

	0% Integrity: Smuzy	50% Integrity: Rapple	100% Integrity: Telo
Caleb			
Donovan			
Elin			
Felix			

nonsense name (i.e., rapple, telo, and smuzy), which were pseudo-counterbalanced across shapes and participants for three of four participants. Table 1 displays the Hiragana characters used as target stimuli for each participant. The stimuli were presented in a horizontal array consisting of three shapes—one target shape and two distractors (also modified Hiragana characters), each measuring  $3 \times 3$  cm. The array of stimuli was printed on  $21.5 \times 27.9$  cm colored paper, placed within sheet protectors, and stored within a binder, which served as the primary experimental interface. The colored paper was associated with an experimental condition (white: 0 % commission errors; yellow: 50 % commission errors; pink: 100 % commission

errors), but was used across all phases of the study. The location of the target and distractor shapes varied on each trial; they were not in the same location on two consecutive trials. Twenty-one different distractors were created with seven assigned to each experimental condition. We created four versions of the binders for each participant. In each version, the order of the colored pages, distractor pairings, and position of the target shapes differed.

### Dependent Variable, Interobserver Agreement (IOA), and Procedural Fidelity

Instruction consisted of 30 interspersed trials; we presented 10 trials of each target shape in a pseudorandom fashion determined a priori by computer randomization. The same target shape was not presented on two consecutive trials. We summarized the data by session and condition, which consisted of 10 trials of each target shape associated with that condition. The primary dependent variable was the percentage of correct responses, defined as pointing to the correct target shape within 5 s of presentation of the discriminative stimulus. We scored incorrect responses when a participant did not select a shape (no response) or when a participant selected the wrong shape. The percentage of correct responses was calculated by dividing the number of correct responses by the total number of trials in the session and multiplying by 100. The second dependent variable was the number of sessions to reach the mastery criterion, which was responding at least 80 % correct for two consecutive sessions.

Throughout each phase, the experimenters collected data via paper and pencil during school hours for approximately 10 min (range 5–10 min), 2–3 days per week. An independent observer recorded IOA and procedural fidelity by viewing videos of sessions for 31, 38, 38.7, and 38.7 % of sessions for Caleb, Donovan, Elin, and Felix, respectively. We used an adaptation of the trial-by-trial method to calculate IOA by dividing the number of trials with agreement on the occurrence or nonoccurrence of a correct response by the total number of trials, multiplied by 100 (Reed and Azulay 2011). Mean percentage agreement for Caleb was 99.7 % (range 97–100 %), for Donovan was 99 % (range 97–100 %), for Elin was 99 % (range 93–100 %), and for Felix was 98.9 % (range 97–100 %). Procedural fidelity was calculated by dividing the number of steps the experimenter implemented correctly (including making planned commission errors) by the total number of steps, multiplied by 100. Procedural fidelity averaged 99.7 % (range 97–100 %) for Caleb, 99 % (range 97–100 %) for Donovan, 99 % (range 93–100 %) for Elin, and 99.4 % (range 96–100 %) for Felix.

### Experimental Design and Procedure

We used a multielement design to examine the short- and long-term effects of various levels of treatment integrity (i.e., errors of commission) on acquisition. The analysis included three phases: (a) baseline, (b) consequence manipulation, and (c) high integrity.

### *Baseline*

At the start of each trial, the experimenter presented the experimental stimuli by placing the binder on a table centered directly in front of the participant and turning the page. Once the participant displayed appropriate attending (i.e., hands placed on the table directly in front of himself or in his lap, sitting silently, and eyes directed to the task or the experimenter), the experimenter delivered the discriminative stimulus (“Where’s [shape]?”). Participant selections did not contact reinforcement; however, we provided reinforcement for appropriate attending based on the schedules and items used during typical classroom instruction (informed by preference assessments and clinical plans completed by teachers outside of research sessions). Caleb and Donovan received a conditioned reinforcer (one token and behavior-specific praise) at the start of each trial for appropriate attending (fixed ratio [FR] 1) and exchanged these for a preferred edible or a toy of their choice after obtaining 10 tokens. Elin and Felix received praise for appropriate attending on a variable interval [VI] 5-min schedule and received access to a preferred toy of their choice at the end of the session.

### *Consequence Manipulation*

The purpose of this phase was to conduct a parametric analysis of treatment integrity on acquisition. The experimental arrangement was similar to baseline except that the three target shapes were associated with varying levels of treatment integrity, which we counterbalanced across participants and shapes. Three integrity levels were examined (0, 50, and 100 % commission errors). A commission error consisted of reinforcement for an incorrect participant response. For the shape associated with 0 % errors, the experimenter did not make errors of commission. For the shape associated with 50 % commission errors, the experimenter provided reinforcement following *every other* incorrect response (FR 2 schedule for participant errors). For the shape associated with 100 % errors, the experimenter provided reinforcement for *every* incorrect response (FR 1 schedule for participant errors). Correct responses, as well as responses during which commission errors were made, contacted reinforcement in the form of generic praise (e.g., “Way to go!”) for Elin and Felix and generic praise and a token for Caleb and Donovan. For every set of 10 tokens earned, Caleb and Donovan exchanged their tokens for 2-min access to a preferred toy of their choice or one edible. The exchange periods occurred during the instructional period (up to three times). Elin and Felix had 2-min access to preferred items twice during instruction (midway and at the end of instruction). We also implemented error correction contingent on incorrect participant responses occurring on trials during which we did not program a commission error. Error correction consisted of least-to-most prompting, which consisted of a gestural prompt followed by a physical prompt if participants did not respond correctly with the gestural prompt. Physical prompting was necessary on fewer than 10 trials throughout the entire study. A neutral statement (“That’s [shape]”) was provided after participants responded correctly. The experimenter did not implement error correction on trials containing experimenter commission errors

to simulate teacher behaviors during actual instruction. That is, teachers are unlikely to provide error correction after delivering reinforcement.

### *High Integrity*

The purpose of this phase was to evaluate the effects of prior exposure to errors of commission on acquisition after we discontinued programmed commission errors. Instruction was similar to previous phases except that we delivered reinforcement for correct responses only and implemented error correction following every incorrect response. Participants did not receive any other programmed consequences during this phase.

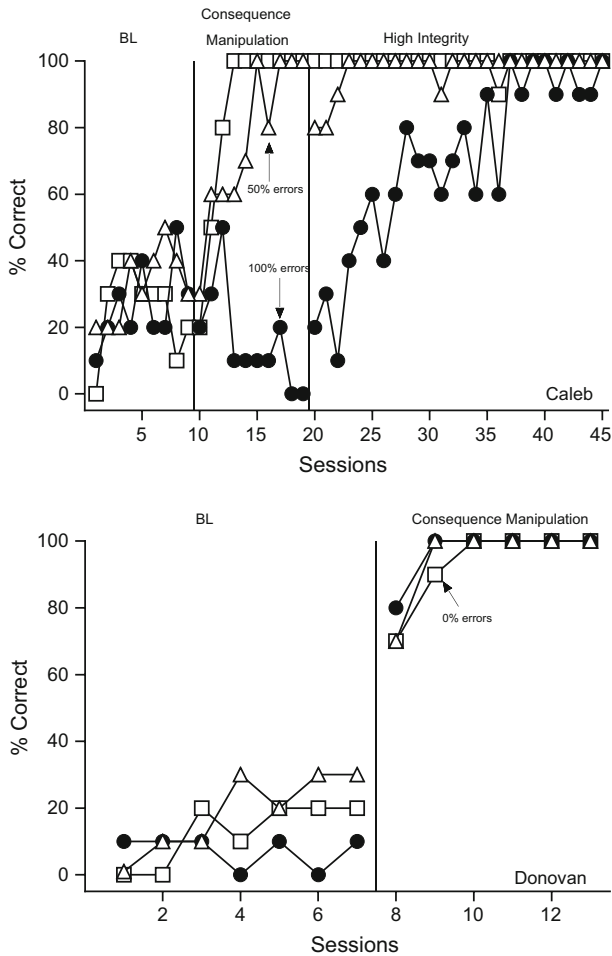
## **Results**

Figures 1 (Caleb and Donovan) and 2 (Elin and Felix) depict the participants' percentage of correct responding across all 10-trial sessions. Table 2 presents the number of sessions to mastery criterion by phase and condition for each participant. Caleb's performance during baseline was variable and low. During consequence manipulation, Caleb displayed differentiated performance within three sessions in each condition. Caleb showed a higher percentage correct during the 0 % ( $M = 85\%$ ) and 50 % ( $M = 76\%$ ) errors conditions and met mastery criterion for both of these conditions. He reached criterion in 7 sessions in the 50 % errors condition and required fewer sessions to reach mastery when instruction did not feature errors (0 % errors: 4 sessions). Caleb did not reach mastery criterion for the shape associated with the 100 % errors condition during consequence manipulation. During high integrity, Caleb required 19 sessions to meet mastery criterion for the shape previously associated with 100 % errors, which is substantially greater than the number of sessions necessary to meet mastery for the shapes associated with 0 and 50 % errors. This finding indicates that Caleb experienced a delay to acquisition for the shape previously associated with 100 % errors.

During baseline, Donovan demonstrated low but slightly variable performance across all target shapes. He showed immediate increases in the percentage of correct responses upon introduction of consequence manipulation in all conditions. He obtained 100 % correct within two to three sessions across all integrity levels and maintained this level of performance throughout this phase. Because Donovan's performance met mastery criterion during consequence manipulation for all three target shapes (100 %: 2 sessions; 50 %: 3 sessions; 0 %: 3 sessions), he was not exposed to the high integrity condition.

Elin's baseline performance was generally low for all target shapes. During consequence manipulation, his percentage correct was variable and undifferentiated for approximately 15 sessions in each condition. Although some differentiation is evident toward the end of consequence manipulation, Elin's data pattern indicates acquisition across 0 % ( $M = 59\%$ ), 50 % ( $M = 34\%$ ), and possibly 100 % ( $M = 41\%$ ) errors. Elin achieved mastery during consequence manipulation after 21 sessions for only the 0 % errors condition. During the high integrity phase, Elin

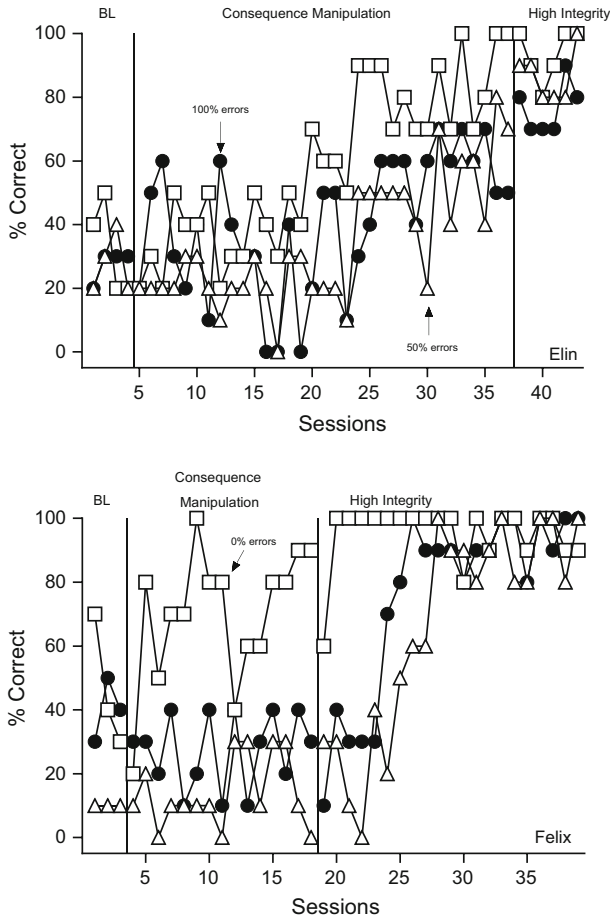




**Fig. 1** Percentage of correct responses on a match-to-sample task during discrete-trial training across baseline, consequence manipulation, and high integrity (one participant) phases for Caleb and Donovan

demonstrated mastery after only 2 sessions for the 50 % errors condition and 6 sessions for the 100 % errors condition.

Felix's performance during baseline was also variable and low. His performance was differentiated within two sessions in each condition during consequence manipulation. Felix showed a higher percentage correct during 0 % errors only ( $M = 70 %$ ), meeting the mastery criterion after only 7 sessions in the consequence manipulation phase. His performance during 50 % ( $M = 14 %$ ) and 100 % ( $M = 27 %$ ) errors was low and variable. He did not meet mastery criterion for the shapes associated with these conditions during consequence manipulation. Felix required 8 and 11 sessions to meet mastery criterion for the shape associated with the 100 and 50 % errors conditions, respectively, during the high integrity phase. Although a slight delay to acquisition may be evident for the shape previously



**Fig. 2** Percentage of correct responses on a match-to-sample task during discrete-trial training across baseline, consequence manipulation, and high integrity phases for Elin and Felix

**Table 2** Sessions to mastery criterion by phase and condition for each participant

	Consequence manipulation			High integrity		
	100 %	50 %	0 %	100 %	50 %	0 %
Caleb	– (10)	7 (10)	4 (10)	19 (26)	– (26)	– (26)
Donovan	2 (6)	3 (6)	3 (6)	– (0)	– (0)	– (0)
Elin	– (33)	– (33)	21 (33)	6 (6)	2 (6)	– (6)
Felix	– (15)	– (15)	7 (15)	8 (21)	11 (21)	– (21)

The number outside of the parentheses denotes the number of trials participants experienced before reaching mastery criterion. The number in parentheses denotes the total number of trials participants experienced in the consequence manipulation and high integrity phases

associated with 50 % errors, his acquisition period for the shape previously associated with the 100 % errors is consistent with the number of sessions required to master the shape in the 0 % error condition during consequence manipulation.

## Discussion

The purposes of the present study were to (a) systematically replicate and extend previous studies, and (b) evaluate the short- and long-term effects of imperfect integrity during discrete-trial training on learning in students with autism. The integrity manipulation included programmed errors of commission as well as error correction omissions, both of which must be considered when interpreting the findings. Portions of these results replicate DiGennaro Reed et al. (2011), which documented the immediate effects of degradations in treatment integrity on acquisition. Integrity errors impaired acquisition for three of four participants; however, interesting data patterns emerged in the present study. Caleb's performance during 50 % errors was similar to his performance during the condition containing no commission errors, which differs from Felix's performance and that of participants in DiGennaro Reed et al. Elin's extended undifferentiated pattern suggests other variables may have influenced his performance. For example, the schedule-correlated stimuli may have been insufficient to differentiate performance across tasks such that the amount and type of feedback delivered during other tasks influenced responding. Integrity level did not influence acquisition for Donovan. An analysis of within-session data revealed that he made a correct response on the first trial of each condition during session 2 of consequence manipulation, after which he contacted accurate feedback, which may have been sufficient to influence his future responses. Both Elin's and Donovan's data patterns indicate that integrity level does not influence acquisition similarly for all learners, which is an important area for future research.

This study also evaluated the long-term impact of treatment integrity errors on skill acquisition after the errors were no longer committed. Only two of the three participants who experienced the high integrity condition displayed a delay to acquisition when treatment integrity errors were no longer committed. Caleb required a substantially higher number of sessions to reach mastery criterion in the high integrity phase for the shape previously associated with 100 % errors. Felix displayed a slight delay to acquisition for the shape previously associated with 50 % errors. This finding supports previous research that the long-term effects of exposure to treatment integrity errors may be idiosyncratic. Although participants in Hirst et al. (2013) showed a substantial delay to acquisition after errors were removed from instruction, Hirst and DiGennaro Reed (2014), Leon et al. (2014), and Carroll et al. (2013) did not consistently demonstrate delays to acquisition with their participants. These collective findings suggest that individual performance is not predictably sensitive to degradations in treatment integrity. Moreover, the delays were not lengthy in the present study, particularly for Felix, suggesting that teachers may be able to reverse the adverse effects of treatment integrity errors with relatively small amounts of high-quality instruction. Carroll et al. reported similar

findings; however, the results should be interpreted with caution because participant exposure to treatment integrity errors was brief in both the present study and Carroll et al. As a result, students who receive poor integrity teaching for months or years may not show the same pattern once teaching is improved.

These findings have implications for educators who teach children with autism using discrete-trial training techniques. First, it appears that for some learners, even a moderate number of integrity errors (i.e., 50 % errors) may negatively affect acquisition, both immediately and after errors no longer occur. This effect does not equally apply to all learners and appears idiosyncratic, which may be a welcome relief to educators who occasionally make instructional errors. Moreover, this finding differs from the results of DiGennaro Reed et al. who showed that the 50 and 100 % error conditions consistently and negatively influenced performance. Although the finding that treatment integrity errors do not affect all learners in the same way (and for some learners, not at all) is an important contribution to the literature, we are unaware of the variables that are responsible for modulating the effects of integrity on learning. Future research might also investigate how idiosyncratic differences (e.g., discrimination skills, level of functioning) influence the sensitivity of responding to treatment integrity degradations. For example, Elin could tact features of the stimuli (e.g., straight lines, curvy lines) and make comparisons with the stimuli names (“Rapple looks like apple”). The participants in DiGennaro Reed et al. did not have these skills, which may account for differences in results across these studies. Interestingly, Donovan had the highest score on the KABC, which was in the above average range, and was the only participant whose responding was not influenced by integrity level. It may be the case that level of functioning interacts with integrity level to influence patterns of responding. Firm conclusions must be tempered given the small sample size, but this observation suggests an important next step for research.

The results of the present study, in combination with other published research, also have implications from a training perspective and underscore the need for initial and ongoing staff training. For example, the findings of Leon et al. (2014) and St. Peter Pipkin et al. (2010) suggest that treatment integrity errors are less detrimental to learner outcomes when they follow a period of high integrity. As a result, we encourage service providers to provide high-quality training to educators before educators implement a treatment or instructional protocol (e.g., Catania et al. 2009; Dib and Sturmey 2007). Carroll et al. and the present study demonstrated that the adverse effects of integrity errors on learner outcomes could be remediated quickly once instruction no longer featured errors. Thus, service providers should develop processes and practices for ongoing educator support and follow-up (e.g., supervisor observations of plan implementation with feedback or other techniques to address integrity errors once they occur; e.g., Coddling et al. 2005).

Despite these important implications, a number of limitations warrant discussion. This study does not represent a perfect analog to errors that occur during instruction. It is unlikely that teachers would make errors similar to the conditions we evaluated, which is supported by the findings of Carroll et al. It is not clear from their methodological descriptions, however, if their scoring procedures included errors of omission *and* commission. Thus, their findings may

only reflect errors of omission and tell us very little about the extent to which natural contexts feature commission errors. The prevalence of errors of commission, then, is still unclear. It may be beneficial for descriptive research highlighting the types and frequencies of errors that are likely to occur in applied settings to guide future inquiries into the influence of commission errors in the context of discrete-trial training. Future research could also examine different types of integrity errors as well as errors committed during other behavioral procedures to improve the generalizability of these findings. The external validity may also be limited due to the nonsense task, which we used to limit any negative affect on learning outside of the research sessions. Future research may evaluate educational tasks that resemble content covered by and teaching strategies implemented by classroom teachers, similar to Carroll et al. Another limitation involves the stimuli we used as potential reinforcers, which participants' teachers identified outside of research sessions. A lack of stimulus preference and reinforcer assessments prevented us from demonstrating that the consequences actually functioned as reinforcers; however, we observed differentiation across most participants from baseline to consequence manipulation, suggesting the programmed consequences (e.g., tokens, toys) functioned as reinforcers. Although we did not conduct formal stimulus preference and reinforcer assessments for each participant in the current study, the programmed consequences (i.e., token economy, receipt of tangibles after work sessions) mirror the application of reinforcement in applied settings and procedures that have been described in the literature for several decades (e.g., Martin et al. 1968; Matson and Boisjoli 2009; Tarbox et al. 2006). That is, individuals engage in work tasks and receive preferred items after the session is complete. Finally, to address the ecological validity limitations of DiGennaro Reed et al. (2011), we did not implement error correction for incorrect responses equally across all conditions, which may have influenced responding during the consequence manipulation phase. Error correction did not occur during the 100 % error condition because every incorrect response was followed by an experimenter commission error. Thus, error correction was present in only two of the three integrity conditions (i.e., 0 and 50 % errors). To assess the unique contributions made by the reinforcement errors of commission and error correction omissions, researchers could evaluate the effects of commission errors with and without error correction. Although this manipulation may prove to be a useful area of investigation, it would necessarily sacrifice some degree of ecological validity as it would not directly mirror instructional practices in natural settings.

**Acknowledgments** This investigation was supported by the University of Kansas General Research Fund allocation #2301523. The authors wish to thank Tanya Baynham for her assistance with this project.

## References

- 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*, 379–394. doi:10.1002/jaba.49.

- 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*, 387–392. doi:[10.1901/jaba.2009.42-387](https://doi.org/10.1901/jaba.2009.42-387).
- Codding, R. S., Feinberg, A. B., Dunn, E. K., & Pace, G. M. (2005). Effects of immediate performance feedback on implementation of behavior support plans. *Journal of Applied Behavior Analysis, 38*, 205–219. doi:[10.1901/jaba.2005.98-04](https://doi.org/10.1901/jaba.2005.98-04).
- Dib, N., & Sturmey, P. (2007). Reducing student stereotypy by improving teachers' implementation of discrete-trial teaching. *Journal of Applied Behavior Analysis, 40*, 339–343. doi:[10.1901/jaba.2007.52-06](https://doi.org/10.1901/jaba.2007.52-06).
- 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*, 611–615. doi:[10.1901/jaba.2011.44-611](https://doi.org/10.1901/jaba.2011.44-611).
- Gresham, F. M. (1989). Assessment of treatment integrity in school consultation and prereferral intervention. *School Psychology Review, 18*, 37–50.
- Groskreutz, M. P., Groskreutz, N. C., & Higbee, T. S. (2011). Response competition and stimulus preference in the treatment of automatically reinforced behavior: A comparison. *Journal of Applied Behavior Analysis, 44*, 211–215. doi:[10.1901/jaba.2011.44-211](https://doi.org/10.1901/jaba.2011.44-211).
- Grow, L., Carr, J. E., Gunby, K. V., Charania, S. M., Gonsalves, L., Ktaech, I. A., et al. (2009). Deviations from prescribed prompting procedures: Implications for treatment integrity. *Journal of Behavioral Education, 18*, 142–156. doi:[10.1007/s10864-009-9085-6](https://doi.org/10.1007/s10864-009-9085-6).
- Hirst, J. M., & DiGennaro Reed, F. D. (2014). An examination of the effects of feedback accuracy on academic task acquisition in analogue settings. *The Psychological Record, .* doi:[10.1007/s40732-014-0087-y](https://doi.org/10.1007/s40732-014-0087-y).
- Hirst, J. M., DiGennaro Reed, F. D., & Reed, D. D. (2013). Effects of varying feedback accuracy on task acquisition: A computerized translational study. *Journal of Behavioral Education, 22*, 1–15. doi:[10.1007/s10864-012-9162-0](https://doi.org/10.1007/s10864-012-9162-0).
- Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman assessment battery for children* (2nd ed.). Circle Pines, MN: American Guidance Service.
- Leon, Y., Wilder, D. A., Majdalany, L., Myers, K., & Saini, V. (2014). Errors of omission and commission during differential reinforcement of compliance: The effects of varying levels of treatment integrity. *Journal of Behavioral Education, 23*, 19–33. doi:[10.1007/s10864-013-9181-5](https://doi.org/10.1007/s10864-013-9181-5).
- Lord, C., Rutter, M., DiLavore, P., & Risi, S. (1999). *Autism diagnostic observation schedule: Manual*. Los Angeles: Western Psychological Services.
- Martin, G. L., England, G., Kaprowy, E., Kilgour, K., & Pilek, V. (1968). Operant conditioning of kindergarten-class behavior in autistic children. *Behavior Research and Therapy, 6*(3), 281–294. doi:[10.1016/0005-7967\(68\)90062-4](https://doi.org/10.1016/0005-7967(68)90062-4).
- Matson, J. L., & Boisjoli, J. A. (2009). The token economy for children with intellectual disability and/or autism: A review. *Research in Developmental Disabilities, 30*, 240–248. doi:[10.1016/j.ridd.2008.04.001](https://doi.org/10.1016/j.ridd.2008.04.001).
- National Research Council. (2001). Educating children with autism. In C. Lord, & J. P. McGee (Eds.), *Committee on educational interventions*. Washington, DC: National Academy Press.
- National Standards Report. (2009). *Evidence-based practice and autism in the schools: A guide to providing appropriate interventions to students with autism spectrum disorders*. Randolph, MA: National Autism Center.
- Northup, J., Fisher, W., Kahng, S. W., Harrell, R., & Kurtz, P. (1997). An assessment of the necessary strength of behavioral treatments for severe problem behavior problems. *Journal of Developmental and Physical Disabilities, 9*, 1–16. doi:[10.1023/A:1024984526008](https://doi.org/10.1023/A:1024984526008).
- Reed, D. D., & Azulay, R. L. (2011). A Microsoft Excel<sup>®</sup> 2010 based tool for calculating interobserver agreement (IOA). *Behavior Analysis in Practice, 4*, 45–52.
- St. Peter Pipkin, 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*, 47–70. doi:[10.1901/jaba.2010.43-47](https://doi.org/10.1901/jaba.2010.43-47).
- Tarbox, R. S., Ghezzi, P. M., & Wilson, G. (2006). The effects of token reinforcement on attending in a young child with autism. *Behavioral Interventions, 21*, 155–164. doi:[10.1002/bin.213](https://doi.org/10.1002/bin.213).