RESEARCH ARTICLE

Treatment Integrity Reporting in *Behavior Analysis in Practice* 2008–2019

Galan Falakfarsa[1] · Denys Brand[1] · Lea Jones[1] · Erik S. Godinez[1] · Deborah C. Richardson[1] · Robbie J. Hanson[1] · Savannah D. Velazquez[1] · Colin Wills[1]

Abstract
Treatment integrity is the extent to which procedures are implemented in a manner consistent with their prescribed protocols and is necessary for reaching accurate conclusions regarding functional relations between dependent (i.e., behavior) and independent (i.e., the environment) variables. Several studies assessing the frequency that studies report treatment integrity have been conducted. However, no review has included articles from *Behavior Analysis in Practice.* Thus, the current study reviewed *Behavior Analysis in Practice* between 2008 and 2019 to assess the frequency of studies reporting treatment integrity data. A total of 193 articles consisting of 205 studies met the inclusionary criteria for this review. Ninety-six studies (46.83%) reported treatment integrity data, compared to 193 (94.15%) that provided interobserver agreement data. Additionally, 98 studies (47.80%) were considered high risk for treatment implementation inaccuracies. Recommendations and implications for research and practice are discussed.

Keywords Internal validity · Intervention evaluation · Procedural fidelity · Procedural integrity · Treatment integrity · Treatment fidelity

One of the main goals of applied behavior analysis is to produce socially significant behavior change (Baer et al., 1968, 1987). In order to achieve this goal, practitioners within behavior analysis rely on empirically validated research to guide the interventions they use based on individual client needs. Therefore, research articles published in behavior-analytic journals have a great responsibility to demonstrate behavior change with strong internal validity. More specifically, published articles must (a) show that changes in behavior (i.e., dependent variables [DVs]) are due to the application of prescribed interventions (i.e., independent variables [IVs]) and (b) describe the methods in a way that lends itself to replicability (Baer et al., 1968, 1987).

Treatment integrity, or the consistency with which IVs are administered according to their prescribed protocol, has been documented within and outside of behavior-analytic literature and position statements as one important factor that impacts

the internal validity of a study (e.g., Association for Behavior Analysis International, 1989; Bellg et al., 2004; Brand et al., 2019; DiGennaro Reed & Codding, 2014; Hagermoser Sanetti & Kratochwill, 2014; Van Houten et al., 1988). Without this measure, it is difficult, if not impossible, to ensure that changes in DVs (or lack thereof) were not due to deviations in IV manipulation, limiting the conclusions that can be made about the results (Cook & Campbell, 1979; Kazdin & Tuma, 1982; Peterson et al., 1982).

Additionally, previous research has indicated that intervention effectiveness may be impacted by varying levels of treatment integrity (e.g., Fryling et al., 2012). Thus, studies that do not report treatment integrity data may lead to erroneous conclusions regarding the efficacy and effectiveness of interventions, indicating that ineffective treatments are effective and vice versa (Gresham et al., 1993a; Peterson et al., 1982; Sanetti et al., 2012). These inaccurate conclusions are concerning because they have the potential to misinform other areas of research and cause direct harm to clinically relevant populations when practitioners reference research to design interventions. For example, if a procedure is implemented with the goal of increasing learner compliance, and no change in learner behavior is observed, it can lead to a conclusion that

✉ Denys Brand
denys.brand@csus.edu

[1] Department of Psychology, California State University, 6000 J Street, Sacramento, CA 95819-6007, USA

the procedure is ineffective for teaching compliance and is less likely to be adopted as part of regular clinical practice. However, without measuring the integrity with which the procedure was administered, conclusions regarding its effectiveness may be premature or incorrect (Peterson et al., 1982).

Another area for practitioners that could be directly impacted by inaccurate conclusions regarding intervention effectiveness is the client discharge process. Practitioners have an ethical obligation to appropriately plan for the termination of services, and health care funders rely on practice guidelines regarding appropriate client discharge planning that mirror these guidelines (e.g., Council of Autism Service Providers, 2020; Kaiser Permanente, 2018). For example, the *Professional and Ethical Compliance Code for Behavior Analysts* Section 2.15 (d) states,

> Discontinuation only occurs after efforts to transition have been made. Behavior analysts discontinue a professional relationship in a timely manner when the client: (1) no longer needs the service, (2) is not benefiting from the service, (3) is being harmed by continued service, or (4) when the client requests discontinuation. (Behavior Analyst Certification Board, 2014, p. 10)

Thus, drawing inaccurate conclusions regarding an intervention's effectiveness may lead a practitioner to use an ineffective procedure with a client, resulting in a lack of progress toward individualized goals. This in turn may lead both the practitioner and the funder to believe that the client is no longer benefiting from the service, which may initiate the discharge process prematurely.

Moreover, inaccurate conclusions may have a negative impact on funding for behavior-analytic services on a larger scale. Funders of behavior-analytic interventions rely heavily on published research to guide funding requirements. Faulty conclusions in published research could lead health care funders to deem components of applied behavior analysis ineffective, resulting in a lack of funding for services in the future. Conversely, health care funders may determine that a published intervention is effective, although it lacks treatment integrity, resulting in the reallocation of funding for ineffective treatments. This point should be especially troubling, given that as of August 2019, all 50 states within the United States passed reform laws mandating that applied behavior analysis is covered by health care insurance as a medically necessary treatment for autism (Autism Speaks, 2019). In other words, misinformation on the health care level could have a profound impact on continued funding and in turn has the potential to have a damaging impact on the public's perception of behavior analysis. Finally, treatment integrity errors may also adversely affect the external validity of a procedure, which

reduces the likelihood that the effects of the intervention will generalize across behaviors, settings, and clients (Baer et al., 1968).

Despite the importance of treatment integrity, it is not necessarily a requirement when submitting research to a journal to be considered for publication (Behavior Analysis in Practice, 2020). Thus, this measure may be entirely absent from a published study. When this occurs, readers may incorrectly assume that treatment integrity data were recorded but simply not reported in the article. Several studies have been conducted to examine the frequency with which treatment integrity data are reported within behavior-analytic journals (e.g., Gresham et al., 1993a; Lee et al., 2007; McIntyre et al., 2007; Peterson et al., 1982). Table 1 displays a summary of the main findings from these studies.[1] Studies in Table 1 include articles published between 1968 and 2012 (44 years), cover 22 unique journals, and include at least 1,717 articles across a variety of interventions and populations (e.g., school based, treatments involving individuals with autism spectrum disorders).

The *Journal of Applied Behavior Analysis* was included in seven review studies (Gresham et al., 1993a; Gresham et al., 1993b; McIntyre et al., 2007; Peterson et al., 1982; Progar et al., 2001; Sanetti et al., 2012; Wheeler et al., 2006), making it the most frequently reviewed behavior-analytic journal on this topic. Collectively, the results from Table 1 indicate that it is not typical for studies published in behavior-analytic journals to report treatment integrity data and that the percentage of studies reporting such data varied between 7.4% (Gresham et al., 2000) and 56% (Lee et al., 2007). Although the results indicate the need for improvement, reporting treatment integrity data has slowly increased since the initial Peterson et al. (1982) review. However, it appears that despite its documented importance, reporting treatment integrity data is (still) often neglected as part of the research process (Sanetti et al., 2012). Additionally, data not reported in Table 1 show the frequency that operational definitions of IVs are provided within journal articles have also increased.

The reviews by Peterson et al. (1982), McIntyre et al. (2007), and Sanetti et al. (2012) also investigated the number of studies that are considered high risk for treatment implementation inaccuracies. High-risk studies were operationally defined as those "including person-implemented interventions that included multiple behavioral components (e.g., contingent reinforcement with response cost)" (McIntyre et al., 2007, p. 663). Additionally, these studies did not report any treatment integrity data, nor did they indicate that this was monitored at any point in the study. The results showed that 43% (Sanetti et al., 2012), 45% (McIntyre et al., 2007), and

---

[1] Articles included in Table 1 were identified via PsycINFO and Google Scholar searches. More information regarding the exact parameters of the searches can be provided upon request.

Table 1 Summary of Studies That Review Treatment Integrity Reporting in Peer-Reviewed Behavior Analytic Journals

Authors	Years reviewed	Journal reviewed	No. of articles/studies reviewed	% of studies reporting TI	Mean high-risk %	Purpose
Peterson et al. (1982)	1968–1980	Journal of Applied Behavior Analysis	539	20%	55%[a]	Behavioral interventions
Monchar and Prinz (1991)	1980–1988	Behavior Therapy	125	16%		Behavioral interventions
Gresham et al. (1993a)	1980–1990	Journal of Applied Behavior Analysis	158	16%		All studies involving children under the age of 19
Gresham et al. (1993)	1980–1990	Behavior Modification	32			Behavioral interventions involving children under the age of 19
		Behavior Therapy	19			
		Journal of Applied Behavior Analysis	84			
		Journal of Abnormal Child Psychology	9			
		Journal of Consulting and Clinical Psychology	9			
		Journal of Behavior Therapy and Experimental Psychiatry	9			
		Behavioral Disorders	19			
		Total articles/studies reviewed in Gresham et al. (1993b)	181	14.9%		
Armstrong et al. (1997)	1991–1995	Journal of Developmental and Physical Disabilities	39	23%		Behavioral interventions
Gresham et al. (2000)	1995–1999	Journal of Learning Disabilities	27	7.4%		Treatment interventions
		Learning Disability Quarterly	22	22.7%		
		Learning Disabilities Research & Practice	16	31.3%		
Progar et al. (2001)	1968–1972 1996–2000	Journal of Applied Behavior Analysis	Exact number not reported[b]	25%		Behavioral intervention
Wheeler et al. (2006)	1993–2003	Journal of Applied Behavior Analysis	36			Behavioral interventions
		Focus on Autism and Other Developmental Disabilities	6			
		Journal of Positive Behavior Interventions	5			
		Research in Developmental Disabilities	4			
		Journal of Autism and Developmental Disorders	3			
		Education and Treatment of Children	2			
		Education and Training in Mental Retardation and Developmental Disabilities	2			
		Journal of Early Intervention	1			
		Journal of Developmental and Physical Disabilities	1			
		Total articles/studies reviewed in Wheeler et al. (2006)	60	18%		
Lee et al. (2007)	1991–2005	Journal of Behavioral Education	375	19% (College) 56% (P-12)		School-based interventions involving P-12 and College
McIntyre et al. (2007)	1991–2005	Journal of Applied Behavior Analysis	142	30%	45%	School-based interventions
Sanetti et al. (2012)	1999–2009	Journal of Positive Behavior Interventions	72	41.7%	43%	Experimental interventions
		Journal of Applied Behavior Analysis				
		Focus on Autism and Other Developmental Disabilities				
		Journal of Autism and Developmental Disorders				
		Research in Autism Spectrum Disorder				
		Education and Training in Autism and Developmental Disabilities				
Neely et al. (2015)	1992, 2002, 2012	Total articles/studies reviewed in Neely et al. (2015)	119	48.7%		Behavioral interventions involving participants with an ASD diagnosis

Note. P–12 = preschool through 12th grade; ASD = autism spectrum disorder.

[a] Mean obtained via WebPlotDigitizer Version 4.3.   [b] Authors reviewed studies published in issues *1*(1)–5(4) and *29*(4)–33(4).

55% (Peterson et al., 1982) of the studies reviewed were considered high risk for treatment implementation inaccuracies. These numbers are alarming and should be concerning to behavior analysts and readers of these journals. Given the importance of measuring treatment integrity data and the role of these data in accurately determining functional relations between DVs and IVs (one of the foundations of a science of behavior; Skinner, 1938), it is crucial for these data to be included as part of our publication practices.

To the best of our knowledge, no treatment integrity review has included articles published in *Behavior Analysis in Practice*. *Behavior Analysis in Practice* is a relatively new behavior-analytic journal that published its first issue in spring 2008. The journal has since increased in popularity among researchers and clinicians. It seems that a review of its publication practices with regard to treatment integrity is appropriate at this time, given the relative age of the journal and its current standing in the field. Such a review can help identify potential gaps in the publication practices of the journal that need to be addressed (Lee et al., 2007), as self-correction is an essential part of the scientific process (Martin & Clarke, 2017). Thus, the purpose of this study was to review articles published in *Behavior Analysis in Practice* between 2008 and 2019 to assess the frequency that treatment integrity data are reported.

Method

Criteria for Review

We reviewed 468 articles to determine potential inclusion. The inclusionary criteria consisted of two features: The study had to involve manipulations of one or more IVs to create a change in the DV, and it had to contain a Method section describing the procedures of the study. Discussion articles, meta-analyses, literature reviews, technical and tutorial articles, surveys, corrections, errata, and various nonexperimental special issue articles were excluded. A total of 193 articles consisting of 205 studies met the inclusionary criteria and were reviewed for analysis.

Coding

If an article consisted of multiple studies, each was coded separately. This review focused on the following variables for analysis: (a) whether operational definitions of the DV were provided, (b) whether authors provided operational definitions of the IVs that were manipulated in the study, (c) whether treatment integrity and interobserver agreement (IOA) data were measured and reported, (d) the treatment agent, (e) the assessments and interventions used as part of the study, (f) the number of studies considered high risk for IV

implementation inaccuracies, (g) the publication year, and (h) the article type (e.g., brief report/practice, research article). The following section provides definitions and descriptions of how each of the aforementioned variables was coded.

Operational Definition of the DV

Studies were coded "yes" or "no" to assess whether the DV was operationally defined in a way that would allow for replication.

Operational Definition of the IV

Each study was coded "yes" or "no" to determine whether the IV was operationally defined. In order to do so, each coder was given the following criterion first proposed by Baer et al. (1968) and used by Gresham et al. (1993a) and McIntyre et al. (2007): "If you could replicate the intervention with the information provided, the intervention is considered operationally defined" (McIntyre et al., 2007, p. 662).

Monitoring Treatment Integrity

Studies were coded "yes" when treatment integrity was monitored and reported for at least one IV. Studies that did not report or did not mention treatment integrity were coded "no." Studies that monitored treatment integrity but did not report any corresponding data were coded "monitored, but not reported." This coding system was adapted from Gresham et al. (1993a) because it allows for a distinction between "yes" and "monitored, but not reported." Studies were also coded with respect to how treatment integrity data were reported (e.g., percentage) and how much treatment integrity data were recorded (i.e., the percentage of sessions for which treatment integrity data were recorded).

Monitoring IOA

Studies that reported IOA data were coded "yes." Studies with no mention of IOA data were coded "no," and those that reported monitoring IOA without reporting any corresponding data were coded as "monitored, but not reported." Studies were also coded with respect to how IOA data were reported (e.g., percentage).

Treatment Agent

The implementer of the intervention was classified into one of seven mutually exclusive categories: (a) teacher (e.g., classroom teachers, early childhood educators), (b) researcher, (c) parent/guardian, (d) direct support professional/teaching assistant/paraprofessional, (e) nurse/other medical staff, (f) licensed/certified practitioner (e.g., psychologist/school

psychologist, Board Certified Behavior Analyst, speech-language pathologist), or (g) not specified/other (e.g., coaches, peers, roommates, self-administered interventions).

Assessments and Interventions

Studies were coded to determine whether they consisted of an assessment (e.g., preference assessments, functional assessments) or an intervention (e.g., reinforcement, antecedent, response cost, extinction). If a study consisted of both an assessment and an intervention, we coded for which part of the study treatment integrity data were reported (assessment only, intervention only, or both). The information was obtained by reading the Method section of each study.

Risk for Procedural Inaccuracies

All included studies were coded as "no risk," "low risk," or "high risk" for IV implementation inaccuracies based on the guidelines used in past reviews (i.e., McIntyre et al., 2007; Peterson et al., 1982). Procedures were coded as "no risk" if treatment integrity was monitored and reported. Procedures were coded as "low risk" if the intervention included single behavioral interventions (e.g., feedback contingent on target skill), mechanically defined treatment (e.g., machine or computer mediated), permanent products (e.g., a decision tree), or continuous application of the IV (e.g., noncontingent activation of toys). Procedures were coded "high risk" if treatment integrity data were not reported but were judged to be necessary. Treatment integrity data were judged to be necessary when the study included "person-implemented interventions that included multiple behavioral components" (McIntyre et al., 2007, p. 663).

Publication Year

The publication year of the article was recorded.

Type of Article

Each article was coded according to its categorization in *Behavior Analysis in Practice* (e.g., brief report, research article).

## Rater Training and Intercoder Agreement

The following section describes how raters were trained and how intercoder agreement scores were calculated. Graduate and undergraduate students served as raters. Prior to coding, all raters received training to discuss the coding scheme and to revise any ambiguous terminology. Each rater received 5 to 10 practice articles prior to independent coding until all raters reached 100% agreement via consensus. A random sample of 89 studies (43.20% of total studies) was assessed for intercoder agreement using the point-by-point method. Percentage agreement scores were calculated by dividing the number of agreements by the total number of agreements and disagreements and multiplying by 100. Intercoder agreement scores averaged 93.15% across eight codes (100% for the operational definition of the IV, 83.15% for treatment integrity reporting, 88.68% for how treatment integrity data were reported, 100% for operational definitions of the DV, 93.26% for IOA reporting, 93.10% for how IOA data were reported, 99.00% for the publication year, and 88.00% for the type of article). We coded two variables (treatment agent; assessments and interventions) across all studies ($n = 205$) via consensus. In addition, we also coded 100% of the included studies regarding the risk of IV implementation inaccuracies, and intercoder agreement scores were 86.89%. In the case of disagreements, the raters met to discuss the disagreement and reach a consensus.

Results

Figure 1 shows the percentage of studies published in *Behavior Analysis in Practice* that reported treatment integrity and IOA data. All studies ($n = 205$) provided operational definitions of the DVs and IVs. Of the studies included in this review, 46.83% ($n = 96$) reported treatment integrity data, 50.73% ($n = 104$) did not report these data, and 2.44% ($n = 5$) monitored but did not report these data. Of the 96 studies that reported treatment integrity data, 7.32% ($n = 15$) reported IOA for the treatment integrity data collected. Treatment integrity data were always reported as a percentage. The percentage of studies per year reporting treatment integrity data ranged from 0% ($n = 0$) in 2011 to 71.43% ($n = 5$) in 2014. Treatment integrity data were recorded for an average of 50.92% (range 2%–100%) of sessions. Of the 96 studies for which treatment integrity data were reported, 94 studies also reported IOA. IOA data were recorded for an average of 47.54% (range 5%–100%) of sessions. Twenty studies (20.83%) explicitly stated that treatment integrity data were recorded during baseline. Treatment integrity data were recorded for an average of 56.96% (range 20%–100%) of baseline sessions. Moreover, 31.58% ($n = 12$) of studies included in this review that were published between 2008 and 2013 ($n = 38$) reported treatment integrity data, whereas 50.29% ($n = 84$) of studies included in this review ($n = 167$) that were published between 2014 and 2019 reported such data—an increase of 18.71%. Thus, treatment integrity data were more likely to be reported in studies published more recently compared to the initial volumes of the journal. We also found that of the 84 studies that were published as brief reports, 40 (47.62%) reported treatment integrity data, and 39 (54.93%) of the 71 studies published as research articles reported

Fig. 1 Percentage of Studies Reporting Treatment Integrity and Interobserver Agreement (IOA) Data by Year (2008–2019)
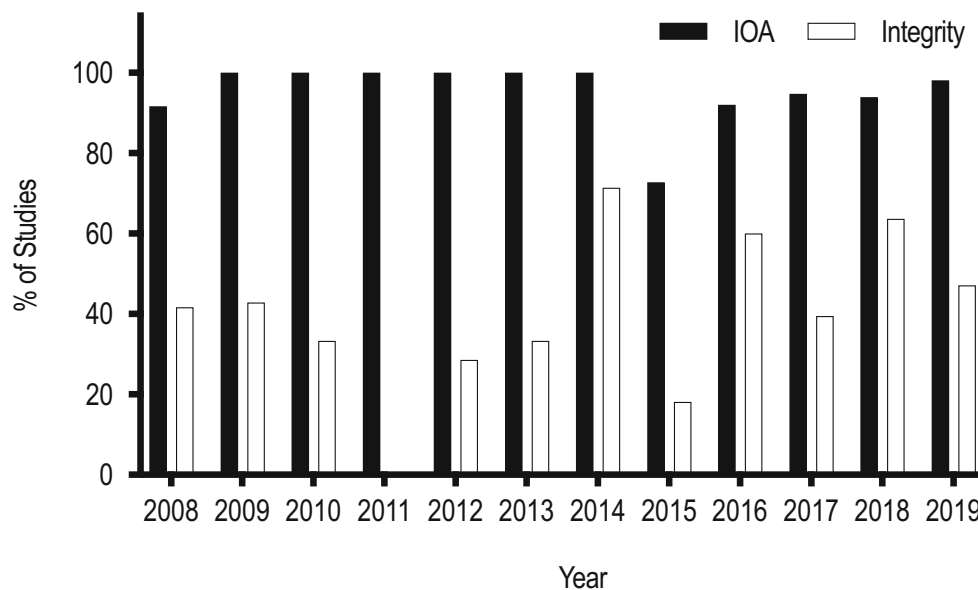
treatment integrity data. The remaining 50 studies did not fall within either of those categories.

Regarding IOA reporting, 94.15% (*n* = 193) of studies reported data, and 5.83% (*n* = 12) did not. The percentage of studies reporting IOA data per year ranged from 72.73% (2015) to 100% (2009–2014). IOA scores were most often reported as a percentage (98.97%; *n* = 192). One study (0.52%) reported IOA as an (unweighted) kappa coefficient, and one study (0.52%) used both a percentage score and a kappa coefficient. The results showed that IOA data were reported over twice as often as treatment integrity data.

Figure 2 displays the percentage of studies that were coded as high, low, or no risk of IV implementation inaccuracies per year. Overall, there is not a stable trend across time, apart from the low-risk category, which remained low. A notable spike for the high-risk category can be seen in 2011, with all studies failing to report treatment integrity data when deemed necessary. Nearly half of the studies included in this review (47.80%; *n* = 98) were considered to be at high risk for IV implementation inaccuracies. The percentage of studies considered high risk per year ranged from 32.00% (2018) to 100% (2011). A slightly smaller percentage of studies (47.32%; *n* = 97) were coded as no risk for treatment inaccuracies per year and ranged from 0% (2011) to 63.64% (2018) of studies. The remaining studies were considered low risk for treatment inaccuracies at 4.88% (*n* = 10). Moreover, 60.52% (*n* = 23) of studies included in this review that were published between 2008 and 2013 (*n* = 38) were coded as high risk, whereas 44.91% (*n* = 75) of studies included in this review that were published between 2014 and 2019 (*n* = 167) were coded as high risk—a decrease of 15.34%. We also found that the percentage of studies coded as no risk increased by

16.08% for studies published from 2008 to 2014 (13/38 = 34.21%) and 2014 to 2019 (84/167 = 50.29%).

Of the 205 studies reviewed, 6.34% (*n* = 13) were assessment studies, 44.39% (*n* = 91) were intervention studies, and 49.26% (*n* = 101) were studies that consisted of both an assessment and an intervention. Treatment integrity data were reported for 38.46% (*n* = 5) of assessment studies, 38.46% (*n* = 35) of intervention studies, and 55.44% (*n* = 56) of studies consisting of both an assessment and an intervention. Of the 56 studies consisting of both an assessment and an intervention (and for which treatment integrity data were reported), 10.71% (*n* = 6) reported treatment integrity data for both the assessment and intervention portions of the study. The remaining 89.28% (*n* = 50) of studies reported treatment integrity data only for the intervention.

Table 2 shows the results for the analyses involving treatment integrity reporting by different treatment agents. The most commonly reported treatment agents were researchers (*n* = 87), not specified (*n* = 38), and multiple (*n* = 37). Treatment integrity data were most likely not to be reported when there were multiple treatment agents (*n* = 24; 64.86%), teachers (*n* = 10; 62.50%), and researchers (*n* = 44; 50.57%). Treatment integrity data were most likely to be reported when treatment agents were licensed/certified practitioners (*n* = 3; 75.00%), direct support professionals/teaching assistants/ paraprofessionals (*n* = 12; 66.66%), and parents/guardians (*n* = 3; 60.00%). Of the 87 studies that involved researchers as treatment agents, 48.28% (*n* = 42) of studies reported treatment integrity data, and one (1.14%) reported that treatment integrity was only monitored.

Fig. 2 Percentage of Studies Coded as High, Low, or No Risk of Independent Variable Inaccuracies per Year (2008–2019)
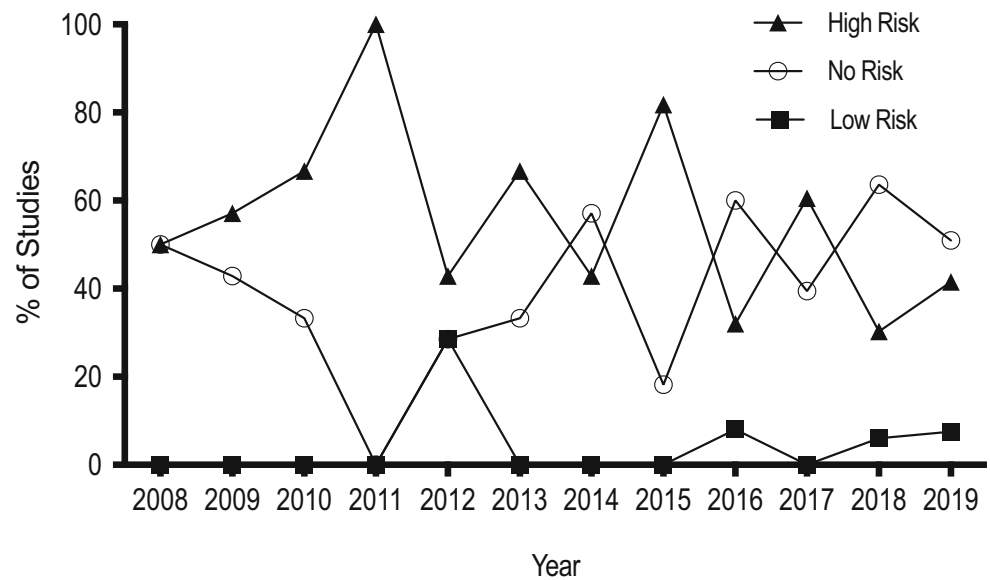
Discussion

The purpose of this study was to review studies published in *Behavior Analysis in Practice* between 2008 and 2019 to assess its publication practices concerning treatment integrity data reporting. The results showed that less than half (46.83%) of the studies included in this review reported treatment integrity data, despite all (100%) including operational definitions of the IV. Thus, in behavior-analytic journals, the rates of treatment integrity reporting remain low. However, the results from this study were slightly higher than those reported in previous reviews on this topic (see Table 1), and the percentages of studies reporting treatment integrity data appear to be increasing over time.

Twenty studies explicitly stated that treatment integrity data were recorded during baseline conditions. Recording such data during baseline may be an important (but often

overlooked) part of a study. Consider an experiment in which researchers are assessing the effectiveness of a particular type of prompting procedure during the acquisition of a specific skill. In this case, to properly assess the effectiveness of the prompting procedure, the researchers may consider recording whether any unplanned or additional prompts were delivered across all parts of the study, including baseline conditions. If unplanned prompts were delivered during baseline (when they should be absent), it could lead to erroneous conclusions regarding the effectiveness of the procedure (Ledford & Gast, 2014).

It was also found that nearly half of the studies included in this review were considered high risk for treatment implementation inaccuracies. This number is consistent with the results reported by previous reviews (McIntyre et al., 2007; Peterson et al., 1982; Sanetti et al., 2012). Furthermore, it was found that treatment integrity data were most frequently reported

Table 2 Treatment Integrity Monitoring by Treatment Agent for Studies Published in *Behavior Analysis in Practice*

Treatment agent	Yes + data *n* (%)	No *n* (%)	Monitored *n* (%)	Total no. of agents
Teacher	5 (31.25)	10 (62.50)	1 (6.25)	16
Researcher	42 (48.28)	44 (50.57)	1 (1.14)	87
Parent/guardian	3 (60.00)	2 (40.00)	0 (0.00)	5
DSP/TA/paraprofessionals	12 (66.66)	4 (22.23)	2 (11.11)	18
Nurse/other medical staff	0 (0.00)	0 (0.00)	0 (0.00)	0
Licensed/certified practitioner	3 (75.00)	1 (25.00)	0 (0.00)	4
Not specified[a]	19 (50.00)	19 (50.00)	0 (0.00)	38
Multiple	12 (32.44)	24 (64.86)	1 (2.70)	37
Total	96 (46.83)	104 (50.73)	5 (2.44)	205

Note. DSP/TA/paraprofessionals = direct support professionals, teaching assistants, and other paraprofessionals.

[a] Includes agents such as dance and fitness instructors, playground facilitators, peer buddies, siblings, job coaches, and unspecified coaches, trainers, and interventionists.

when treatment agents were identified as licensed/certified practitioners and least likely to be reported when multiple treatment agents were listed. As was the case with McIntyre et al. (2007), a high proportion of studies involving researchers as treatment agents did not report treatment integrity data.

Additionally, the current study found that IOA data were reported in over twice as many studies compared to treatment integrity data. It appears that the "curious double standard" first described by Peterson et al. (1982) remains. That is, in the field of behavior analysis, there is a much greater emphasis on ensuring clear specifications of DVs than on IVs. The results from this review do little to dispel this notion. It remains unclear why this double standard exists as part of our research and publication practices. McIntyre et al. (2007) and Sanetti et al. (2012) suggested that it may be a function of the editorial processes of a journal. Some journals (or manuscript types) may have space limitations and require authors to submit manuscripts consistent with a word or page limit, which may result in the eventual omission of treatment integrity data. For example, manuscripts submitted as brief reports in *Behavior Analysis in Practice* have a 3,000-word limit, which includes the abstract, references, and figure legends, whereas a research article allows for up to 40 double-spaced pages. However, the results of the current study found that despite the discrepancy in the allotted space between these two types of manuscript submissions, the numbers of articles reporting treatment integrity data for each were similar. Given these numbers, it appears unlikely that space limitations can fully account for the lack of treatment integrity data reported in *Behavior Analysis in Practice*.

McIntyre et al. (2007) also stated that it might be the case that researchers do not view the collection of treatment integrity data as important, especially if desired changes in behavior are observed. However, empirical investigations need to be conducted to determine the exact reasons why treatment integrity data are not reported with greater frequency. The current state of the literature is purely descriptive and has not yet attempted to systematically determine the cause of this phenomenon. One such investigation could include the dissemination of a survey asking authors in the field of behavior analysis to indicate the reasons why they did not measure or report treatment integrity data. Another method to further investigate this issue is to work with journal editors and associate editors to investigate the frequency that reviewers mentioned the lack of treatment integrity data when no such data were reported. Such an investigation may help determine the importance that journals place on treatment integrity data as part of the review process, and may suggest some areas of improvement. At present, we recommend that the general submission information/guidelines typically included on journal websites explicitly state the requirement for studies to report data regarding the integrity of the IV (if applicable).

The results also showed that when a study consisted of both an assessment and an intervention, it was rare for treatment integrity data to be reported for the assessment portion. The reason why this is the case is not immediately obvious. Assessments are conducted to assist in the selection of appropriate treatments to be used as part of individualized treatment plans, and represent a fundamental part of the clinical process (Behavior Analyst Certification Board, 2014). The efficacy of such an approach has been documented in the literature (e.g., Heinicke et al., 2019; Saini et al., 2020). Thus, given the importance of assessments in clinical practice and research, it may be important to record and report the integrity with which assessments are administered to ensure that they are conducted in a manner consistent with their prescribed protocols. However, the extent to which errors affect the outcomes of such assessments (and the types of errors that are likely to be committed when administering them) remains relatively unknown, and more research is needed (Brand et al., 2019). For example, it would be interesting to assess whether errors committed during stimulus preference assessments result in inaccurate preference rankings. Such an outcome has the potential to adversely affect the effectiveness of the intervention given that preferred items are often provided as reinforcers during skill acquisition programs. Another example involves functional assessments used to identify the function of problem behavior. If treatment integrity errors contributed to the incorrect function of problem behavior being identified, it can have severely detrimental (and potentially dangerous) outcomes for the client. Given the procedural complexity of some of these procedures (e.g., functional analysis), it seems plausible that errors could be committed when administering them. It is recommended here that if a study consists of both an assessment and an intervention, that treatment integrity data be recorded and reported for both.

Further, the results may help identify some of the barriers around measuring and reporting treatment integrity, which may lead to possible solutions and recommendations for future researchers. For example, researchers can take a proactive approach by ensuring that discussions around measuring and reporting treatment integrity data are included when designing a study. Specifically, the factors that need to be considered include (a) the method by which treatment integrity data will be recorded (direct vs. indirect measures), (b) the frequency with which such data will be recorded (e.g., the percentage/proportion of sessions that data will be recorded), and (c) the procedural components to be monitored for treatment integrity purposes. Furthermore, Bellg et al. (2004) recommended that other factors of the experimental design should also be considered when planning to record treatment integrity data, such as the study setting and the burden that such observations will place on both treatment agents and research participants. All the aforementioned factors must be considered in order to

design a plan that is practical, achievable, and effective for monitoring treatment integrity (Bellg et al., 2004).

When conducting research in applied settings where experimental procedures may be performed by individuals not part of the research team (e.g., teachers, direct support personnel, parents), it may be prudent to have discussions from the outset of the study to plan for the frequency with which (e.g., every second or third session) and method by which (in vivo, video recordings) treatment integrity checks will be conducted. Such discussions may identify some barriers that need to be addressed to ensure such data can be obtained. We strongly advocate that investigators establish the importance of measuring and reporting treatment integrity data as part of their regular research practices.

Additionally, given the potential for harm that can be caused when using interventions with clients based on inaccurate conclusions in publications, it seems especially relevant to conduct research on how often this takes place. For example, future research might survey practitioners for common interventions they implement with clients, and then examine the percentage of those interventions that stem from empirical research that (a) exhibits strong internal validity and (b) reports high levels of treatment integrity. This type of research may start to shed light on how the issue of treatment integrity transcends the research–practitioner bridge in behavior analysis.

Further, another question that remains to be answered is the extent that the results from treatment integrity reviews within research carry over to applied settings with relevant populations. It is unclear how often practitioners monitor and report treatment integrity data within everyday practice or how often health care funders require this type of report (if at all). Although practice guidelines and the Behavior Analyst Certification Board Task List items include treatment integrity as a requirement (e.g., Behavior Analyst Certification Board, 2012; Council of Autism Service Providers, 2020; Kaiser Permanente, 2018), the frequency with which this is to occur is unspecified. Given that the lack of treatment integrity data in clinical practice may negatively impact services for clients (e.g., premature termination of services based on false conclusions regarding intervention effectiveness), future studies should examine the correspondence between the lack of treatment integrity reporting in research and the frequency of treatment integrity checks within applied contexts.

Several strategies exist to maximize the likelihood that treatment agents administer procedures with high levels of integrity. Parsons et al. (2012) advocated for the use of evidence-based training procedures consisting of both performance- and competency-based strategies. That is, training needs to continue until the treatment agent can administer the procedure at a predetermined criterion that demonstrates skill mastery (as opposed to providing training for a certain duration of time; McIntyre et al., 2007; Parsons et al., 2012). Behavioral skills training (BST; Miltenberger, 2003) is a frequently cited evidence-based training procedure that consists of instructions, modeling, rehearsal, and feedback. The training literature consists of several resources describing best practices when using BST to train treatment agents across a variety of settings (e.g., DiGennaro Reed et al., 2018; Parsons et al., 2012). Moreover, several resources outline best practices and modifications that can be applied to enhance individual BST components, such as enhanced written instructions (Graff & Karsten, 2012), video modeling (e.g., Catania et al., 2009; DiGennaro Reed et al., 2018), and performance feedback (e.g., Hagermoser Sanetti & Kratochwill, 2014).

Following initial training, treatment integrity checks need to be conducted periodically to detect any procedural inaccuracies (e.g., Brand et al., 2017) or instances of treatment drift (e.g., Hansford et al., 2010) so that interventionists can quickly be retrained to conduct procedures consistent with their prescribed protocols. The staff training literature contains several examples of indirect (e.g., permanent product; e.g., Noell et al., 2005) and direct methods (e.g., checklists, data sheets; e.g., Clayton & Headley, 2019) that can be used to conduct quick treatment integrity spot checks (e.g., DiGennaro Reed et al., 2018; McIntyre et al., 2007) throughout the course of treatment. Identifying and correcting procedural inaccuracies once they have been detected will help avoid exposing treatment recipients to prolonged periods of compromised treatment integrity, which may adversely affect the outcomes of a procedure (Brand et al., 2019). Many of the resources listed here also include examples of data sheets and checklists that can be adapted for use when conducting integrity checks.

Although the current study highlights continued reporting issues with treatment integrity, there are some limitations that warrant further discussion. First, we used the same operational definition for identifying studies as "no risk," "low risk," or "high risk" for treatment implementation inaccuracies as McIntyre et al. (2007), Peterson et al. (1982), and Sanetti et al. (2012). McIntyre et al. pointed out that coding risk categories did not include other potentially important factors, such as the setting, the treatment agent, or the experience level of the treatment agent—all of which could affect the risk of treatment implementation inaccuracies. Research is needed to assess the extent to which these features affect the risk of treatment inaccuracies. For example, descriptive studies (e.g., Brand et al., 2017, Brand et al., 2018; Carroll et al., 2013; Cook et al., 2015) can be conducted to assess the frequency and likelihood of treatment inaccuracies when implementing procedures of varying complexity (McIntyre et al., 2007). Additionally, monitoring and reporting treatment integrity were treated as equivalent when coding studies for risk. However, reporting treatment integrity has greater value to both researchers and practitioners than simply stating that it was monitored (McIntyre et al., 2007). A reconceptualization

of this coding category might be required for future research. Another limitation is how we coded the operational definition of the IV (i.e., "if you could replicate the intervention with the information provided, the intervention is considered operationally defined"). This definition is consistent with previous reviews of this kind, but it is somewhat subjective with respect to how replicable the procedure was (Sanetti et al., 2012). However, despite these limitations, the current study makes an important contribution to the treatment integrity literature and provides several recommendations for ensuring procedures are implemented with integrity in both research and practice settings. It is our assertation that the importance of measuring and reporting treatment integrity data becomes amplified when providing examples of seemingly effective behavioral interventions. It becomes important at that point to conclude with confidence that it was indeed manipulations of the IVs that caused the changes in behavior, as well as to rule out (to the greatest extent possible) the influence of extraneous variables (e.g., treatment drift; Hansford et al., 2010). Without conducting regular treatment integrity checks, the researcher is left to assume that the accuracy with which the IVs were administered remained relatively consistent throughout various parts of a study. Such an assumption does not seem reasonable, especially as procedures become more complex to administer. Without measuring treatment integrity data, there exists the possibility that researchers may inadvertently disseminate results regarding treatment effectiveness that cannot be replicated by other professionals in the field.

Declarations

Ethical approval   This article does not contain any studies with human participants performed by any of the authors.

Informed consent   No informed consent was required to conduct this research.

Conflict of interest   The authors declare that they have no conflict of interest.

References

Armstrong, K. J., Ehrhardt, K. E., Cool, R. T., & Poling, A. (1997). Social validity and treatment integrity data: Reporting in articles in the *Journal of Developmental and Physical Disabilities*, 1991–1995. *Journal of Developmental and Physical Disabilities*, *9*(4), 359–367. 10.1023/A:1024982112859

Association for Behavior Analysis International. (1989). *Statement on the right to effective behavioral treatment, 1989*. https://www.abainternational.org/about-us/policies-and-positions/right-toeffective-behavioral-treatment,-1989.aspx

Autism Speaks. (2019). *Autism Speaks commends Tennessee as it becomes 50th state requiring that insurance plans cover autism*. https://www.autismspeaks.org/press-release/autism-speaks-commends-tennessee-it-becomes-50th-state-requiring-insurance-plans

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1*(1), 91–97. https://doi.org/10.1901/jaba.1968.1-91.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1987). Some still-current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 20*(4), 313–327. https://doi.org/10.1901/jaba.1987.20-313.

Behavior Analysis in Practice. (2020). Behavior Analysis in Practice *submission guidelines*. Springer. https://www.springer.com/journal/40617/submission-guidelines#Instructions%20for%20Authors_Manuscript%20Submission

Behavior Analyst Certification Board. (2012). *Fourth edition task list*.

Behavior Analyst Certification Board. (2014). *Professional and ethical compliance code for behavior analysts*.

Bellg, A. J., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D. S., Ory, M., Ogedegbe, G., Orwig, D., Ernst, D., & Czajkowski, S. (2004). Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychology, 23*(5), 443–451. https://doi.org/10.1037/0278-6133.23.5.443.

Brand, D., Elliffe, D., & DiGennaro Reed, F. D. (2018). Using sequential analysis to assess component integrity of discrete-trial teaching programs. *European Journal of Behavior Analysis, 19*(1), 30–47. https://doi.org/10.1080/15021149.2017.1404392.

Brand, D., Henley, A. J., DiGennaro Reed, F. D., Gray, E., & Crabbs, B. (2019). A review of published studies involving parametric manipulations of treatment integrity. *Journal of Behavioral Education, 28*(1), 1–26. https://doi.org/10.1007/s10864-018-09311-8.

Brand, D., Mudford, O. C., Arnold-Saritepe, A., & Elliffe, D. (2017). Assessing the within-trial treatment integrity of discrete-trial teaching programs using sequential analysis. *Behavioral Interventions, 32*(1), 54–69. https://doi.org/10.1002/bin.1455.

Carroll, R. A., Kodak, T., & Fisher, W. W. (2013). An evaluation of programmed treatment-integrity errors during discrete-trial instruction. *Journal of Applied Behavior Analysis, 46*(2), 379–394. https://doi.org/10.1002/jaba.49.

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. https://doi.org/10.1901/jaba.2009.42-387.

Clayton, M., & Headley, A. (2019). The use of behavioral skills training to improve staff performance of discrete trial training. *Behavioral Interventions, 34*(1), 136–143. https://doi.org/10.1002/bin.1656.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimental design and analysis issues for field settings*. Houghton Mifflin.

Cook, J. E., Subramaniam, S. S., Brunson, L. Y., Larson, N. A., Poe, S. G., & St. Peter, C. C. (2015). Global measures of treatment integrity may mask important errors in discrete-trial training. *Behavior Analysis in Practice, 8*(1), 37–47. https://doi.org/10.1007/s40617-014-0039-7.

Council of Autism Service Providers. (2020). *Applied behavior analysis treatment of autism spectrum disorder: Practice guidelines for healthcare funders and managers* (2nd ed.). https://casproviders.org/asd-guidelines/

DiGennaro Reed, F., & Codding, R. S. (2014). Advancements in procedural fidelity assessment and intervention: Introduction to the special issue. *Journal of Behavioral Education, 23*, 1–18. https://doi.org/10.1007/s10864-013-9191-3.

DiGennaro Reed, F. D., Erath, T. G., Brand, D., & Novak, M. D. (2018). Video modeling during coaching and performance feedback. In A. J. Fisher, T. A. Collins, E. H. Dart, & K. C. Radley (Eds.), *Technology applications in school psychology consultation, supervision, and training* (pp. 46–61). Routledge.

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. https://doi.org/10.1901/jaba.2012.45-449.

Graff, R. B., & Karsten, A. M. (2012). Evaluation of a self-instruction package for conducting stimulus preference assessments. *Journal of Applied Behavior Analysis, 45*(1), 69–82. https://doi.org/10.1901/jaba.2012.45-69.

Gresham, F. M., Gansle, K. A., & Noel, G. H. (1993a). Treatment integrity in applied behavior analysis with children. *Journal of Applied Behavior Analysis, 26*(2), 257–263. https://doi.org/10.1901/jaba.1993.26-257.

Gresham, F. M., Gansle, K. A., Noell, G. H., Cohen, S., & Rosenblum, S. (1993b). Treatment integrity of school-based intervention studies: 1980-1990. *School Psychology Review, 22*(2), 254-272. https://doi.org/10.1080/02796015.1993.12085651.

Gresham, F. M., MacMillan, D., Beebe-Frankenberger, M. E., & Bocian, K. M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented? *Learning Disabilities Research and Practice, 15*(4), 198–205. https://doi.org/10.1207/sldrp1504_4.

Hagermoser Sanetti, L. M., & Kratochwill, T. R. (2014). *Treatment integrity: A foundation for evidence-based practice in applied psychology*. American Psychological Association.

Hansford, A., Zilber, I., LaRue, R. H., & Weiss, M. J. (2010). Practical issues in effective behavioral intervention development. *Behavior Analyst Today, 11*(1), 59–76. https://doi.org/10.1037/h0100689.

Heinicke, M. R., Carr, J. E., & Copsey, C. J. (2019). Assessing preferences of individuals with developmental disabilities using alternative stimulus modalities: A systematic review. *Journal of Applied Behavior Analysis, 52*(3), 847–869. https://doi.org/10.1002/jaba.565.

Kaiser Permanente. (2018). *Applied behavioral analysis treatment for individuals with autism spectrum disorder*. https://wa-provider.kaiserpermanente.org/provider-manual/clinical-review/aba-therapy

Kazdin, A. E., & Tuma, A. H. (1982). *Single-case research designs*. Oxford University Press.

Ledford, J. R., & Gast, D. L (2014). Measuring procedural fidelity in behavioral research. *Neuropsychological Rehabilitation, 24*(3–4), 332–348. https://doi.org/10.1080/09602011.2013.861352

Lee, D., Hamlin, D., Hildebrand, K., Carranza, M., Wannarka, R., & Hua, Y. (2007). A descriptive analysis of publication trends in the *Journal of Behavioral Education*: 1991–2005. *Journal of Behavioral Education, 16*(4), 327–332. https://doi.org/10.1007/s10864-007-9044-z.

Martin, G. N., & Clarke, R. M. (2017). Are psychology journals anti-replication? A snapshot of editorial practices. *Frontiers in Psychology, 8*, 1–6. https://doi.org/10.3389/fpsyg.2017.00523.

McIntyre, L. L., Gresham, F. M., DiGennaro Reed, 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. https://doi.org/10.1901/jaba.2007.659-672.

Miltenberger, R. G. (2003). *Behavior modification: Principles and procedures* (3rd ed.). Wadsworth.

Monchar, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review, 11*(3), 247–266. https://doi.org/10.1016/0272-7358(91)90103-2.

Neely, L., Davis, H., Davis, J., & Rispoli, M. (2015). Review of reliability and treatment trends in autism focused research. *Research in Autism Spectrum Disorders, 9*, 1–12. https://doi.org/10.1016/j.rasd.

Noell, G. H., Witt, J. C., & Slider, N. J. (2005). Treatment implementation following behavioral consultation in schools: A comparison of three follow-up strategies. *School Psychology Review, 34*(1), 87–106. https://doi.org/10.1080/02796015.2005.12086277.

Parsons, M. B., Rollyson, J. H., & Reid, D. H. (2012). Evidence-based staff training: A guide for practitioners. *Behavior Analysis in Practice, 5*(2), 2–11. https://doi.org/10.1007/BF03391819.

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. https://doi.org/10.1901/jaba.1982.15-477.

Progar, P. R., Perrin, F. A., DiNovi, B. J., & Bruce, S. S. (2001). Treatment integrity: Some persistent concerns and some new perspectives. *Behavior Analyst Today, 2*(1), 28–32. https://doi.org/10.1037/h0099912.

Saini, V., Fisher, W. W., Retzlaff, B. J., & Keevy, M. (2020). Efficiency in functional analysis of problem behavior: A quantitative and qualitative review. *Journal of Applied Behavior Analysis, 53*(1), 44–66. https://doi.org/10.1002/jaba.583.

Sanetti, L. M. H., Dobey, L. M., & Gritter, K. L. (2012). Treatment integrity of interventions with children in the *Journal of Positive Behavior Interventions* from 1999 to 2009. *Journal of Positive Behavior Interventions, 14*(1), 29–46. https://doi.org/10.1177/1098300711405853.

Skinner, B. F. (1938). *The behavior of organisms*. Prentice-Hall.

Van Houten, R., Axelrod, S., Bailey, J. S., Favell, J. E., Foxx, R. M., Iwata, B. A., & Lovaas, O. I. (1988). The right to effective behavioral treatment. *Journal of Applied Behavior Analysis, 21*(4), 381–384. https://doi.org/10.1901/jaba.1988.21-381.

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*(1), 45–54. https://doi.org/10.1177/10883576060210010601.