

# Journal of Positive Behavior Interventions

<http://pbi.sagepub.com/>

---

## Treatment Integrity of Interventions With Children in the *Journal of Positive Behavior Interventions* From 1999 to 2009

Lisa M. Hagermoser Sanetti, Lisa M. Dobej and Katie L. Gritter  
*Journal of Positive Behavior Interventions* 2012 14: 29 originally published online 26 April 2011  
DOI: 10.1177/1098300711405853

The online version of this article can be found at:  
<http://pbi.sagepub.com/content/14/1/29>

---

Published by:  
Hammill Institute on Disabilities



and  


<http://www.sagepublications.com>

Additional services and information for *Journal of Positive Behavior Interventions* can be found at:

**Email Alerts:** <http://pbi.sagepub.com/cgi/alerts>

**Subscriptions:** <http://pbi.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

>> [Version of Record](#) - Dec 19, 2011

[OnlineFirst Version of Record](#) - Apr 26, 2011

[What is This?](#)

# Treatment Integrity of Interventions With Children in the *Journal of Positive Behavior Interventions* From 1999 to 2009

Lisa M. Hagermoser Sanetti<sup>1</sup>, Lisa M. Dobey<sup>2</sup>, and Katie L. Gritter<sup>1</sup>

## Abstract

For more than 10 years, the *Journal of Positive Behavior Interventions* has published, among other types of articles, behavioral intervention outcome studies related to positive behavior support. Operationally defining interventions is important to facilitating replication studies and adoption of intervention in applied settings. Furthermore, treatment integrity data are necessary to make valid claims that changes in outcomes resulted from intervention implementation and are thus essential to the internal validity of intervention outcome research. Reviews of treatment outcome research in related fields (e.g., applied behavior analysis) indicate that although many researchers operationally define interventions, a majority of researchers fail to report treatment integrity data. The purpose of this study was to review the treatment integrity data reported in all experimental intervention studies published in the *Journal of Positive Behavior Interventions* between 1999 and 2009. Results indicate that in recent years, a majority of published studies include a definition of the independent variable but do not provide quantitative treatment integrity data.

## Keywords

treatment integrity, intervention evaluation, behavior intervention

Just more than 10 years ago, the publication of the first issue of the *Journal of Positive Behavior Interventions* helped to substantiate the fledgling field of positive behavior support (PBS). Having evolved, in part, out of the discipline of applied behavior analysis (ABA), it shares with its predecessor an emphasis on a plethora of behavioral concepts (e.g., stimulus control, generalization), empirically based assessment techniques (e.g., functional assessment), and educational strategies (e.g., shaping, fading; Carr et al., 2002; Dunlap, Carr, Horner, Zarccone, & Schwartz, 2008). Furthermore, with roots also in the normalization/inclusion movement and person-centered values, the manipulation of environmental variables that lead to an improved quality of life at both the individual and the systemic (e.g., school) level is a key feature of PBS (Carr et al., 2002). Thus, a defining feature of PBS is its explicit commitment to translating the evidence-based principles of behavior change to the complex systems in which individuals live and learn (see Carr et al., 2002, and Dunlap et al., 2008, for full discussions of this and other defining features of PBS).

An essential element of PBS planning, as in ABA, is the operational definition of the variables of concern (e.g., Carr et al., 2002; Johnston, Foxx, Jacobson, Green, & Mulick, 2006). In research using single-subject methodology, which has a

rich history in the operant literature (Cooper, 1982; Morgan & Morgan, 2001), operational descriptions of the independent (i.e., intervention) and dependent (i.e., outcome) variables are two of a variety of indicators of study quality (Horner et al., 2005). Independent variable description is also one of several quality indicators in group experimental research (Gersten et al., 2005). If provided in sufficient detail, a description of the intervention allows for replication as well as accurate conclusions regarding outcomes (Gersten, Baker, & Lloyd, 2000; Horner et al., 2005). Unfortunately, behavioral researchers in years past have appeared to adhere to a “curious double standard” whereby they operationally define and assess reliability for the dependent variable but virtually ignore the independent variable (Peterson, Homer, & Wonderlich, 1982, p. 478). The degree to which the

<sup>1</sup>University of Connecticut, Storrs, CT, USA

<sup>2</sup>Belmont Public School District, MA, USA

## Corresponding Author:

Lisa M. Hagermoser Sanetti, University of Connecticut,  
Department of Educational Psychology, 249 Glenbrook Road,  
Unit 2064, Storrs, CT 06269-2064, USA  
Email: [lisa.sanetti@uconn.edu](mailto:lisa.sanetti@uconn.edu)

Action Editor: Kathleen Lane

**Table 1.** Results of Reviews of Treatment Integrity Data in Treatment Outcome Literatures

Author	Intervention area reviewed	Percentage of studies including an operational definition of the intervention	Percentage of studies including quantitative treatment integrity data
Dane & Schneider (1998)	Prevention	57 <sup>a</sup>	11 <sup>b</sup>
Gansle (2005)	Anger management	NA	18
Gresham, MacMillan, Beebe-Frankenberger, & Bocian (2000)	Learning disabilities	NA	18.5
Gresham, Gansle, & Noell (1993)	Applied behavior analysis	34	16
Gresham, Gansle, Noell, & Cohen (1993)	School-based interventions	35	14.9
McIntyre, Gresham, DiGennaro, & Reed (2007)	Applied behavior analysis	95	30
Perepletchikova, Treat, & Kazdin (2007)	Psychotherapy	65 <sup>a</sup>	3.5
Peterson, Homer, & Wonderlich (1982)	Applied behavior analysis	NA	16
Snell, Chen, & Hoover (2006)	Augmentative/alternative communication	NA	30
Wheeler, Baggett, Fox, & Blevins (2006)	Autism	92	18

a. Percentage of studies in which a detailed treatment manual was referred to; other forms of intervention description not included.

b. Adherence data only reported for Dane and Schneider (1998) to be consistent with data coded in other studies.

independent variable (i.e., intervention) is implemented as planned has traditionally been referred to as treatment integrity or fidelity (e.g., Gresham, 1989; Gresham, Gansle, Noell, & Cohen, 1993).

Documentation of the integrity of an intervention, like its operational definition, is an indicator of study quality (Gersten et al., 2005; Horner et al., 2005). Treatment integrity data are essential to the internal validity of intervention outcome research (Moncher & Prinz, 1991); that is, they are necessary to make valid claims about the relationship between the independent and dependent variables (Gersten et al., 2005; Peterson et al., 1982). Treatment integrity is also related to external validity (Moncher & Prinz, 1991). Replication is compromised if previous evaluations of an intervention have not adequately documented intervention components or how they were applied (Peterson et al., 1982). Reviews of the treatment outcome literature in several fields (e.g., ABA: Gresham, Gansle, & Noell, 1993; McIntyre, Gresham, DiGennaro, & Reed, 2007; Peterson et al., 1982; school-based intervention: Gresham, Gansle, Noell, & Cohen, 1993; anger management: Gansle, 2005; learning disabilities: Gresham, MacMillan, Beebe-Frankenberger, & Bocian, 2000; autism and alternative communication: Snell, Chen, & Hoover, 2006; Wheeler, Baggett, Fox, & Blevins, 2006; prevention science: Dane & Schneider, 1998; psychotherapy: Perepletchikova, Treat, & Kazdin, 2007) have provided validation for concerns over researchers' lack of attention to definition and assessment of the independent

variable (see Table 1). Across all of these reviews, less than one third (range: 4%–30%) of treatment outcome studies included quantitative treatment integrity data (see Sanetti & Kratochwill, 2008, for a review).

Reviews of the ABA literature (Gresham, Gansle, & Noell, 1993; McIntyre et al., 2007; Peterson et al., 1982) are of particular interest for three reasons. First, the fields of PBS and ABA have much in common with regard to fundamental assessment and intervention approaches (e.g., Carr et al., 2002; Dunlap et al., 2008). Second, together these reviews provide nearly 40 years worth of longitudinal data with respect to treatment integrity assessment in behavioral intervention research. Third, the coding criteria used in these reviews served as the basis of the criteria in the current study.

The first review, conducted by Peterson and colleagues (1982), included all experimental articles published in the *Journal of Applied Behavior Analysis (JABA)* between 1968 and 1980. For each article, the authors coded both the operational definition and the assessment of the independent variable in the same way (i.e., reported, not reported but considered unnecessary, not reported and considered necessary). Assessment was coded as reported as long as some form of assessment (e.g., interobserver reliability, calibration) was reported. Results indicated that articles were far more likely to provide operational definitions of the independent variable than to report assessment of them, even when assessment was deemed necessary (i.e., there was a high risk of implementation inaccuracy). On average, only 16% (range 3%–34%) of the articles with operational definitions

of the independent variable also reported a check on treatment integrity. Moreover, the percentage of articles reporting treatment integrity assessment showed barely any increase over the 12-year span.

The second review (Gresham, Gansle, & Noell, 1993) included studies conducted with children less than 19 years of age published in *JABA* between 1980 and 1990. If the intervention could be replicated based on the definition provided, the study was considered to include an operational definition of the independent variable. Three categories (i.e., yes, monitored, not mentioned) were used to rate the level of treatment integrity assessment reported in a study. If quantitative data in the form of a percentage were provided, the study was rated “yes.” If treatment integrity assessment was mentioned but no quantitative data were provided, the study was rated as “monitored.” If no reference was made to treatment integrity, the study was rated as “not mentioned.” The results of the review indicated that only 34% of the reviewed studies provided an operational definition of the intervention. Furthermore, only 16% provided quantitative treatment integrity assessment data, 9% mentioned assessment but provided no data, and a full 75% did not report having assessed treatment integrity at all.

Finally, McIntyre and colleagues (2007) reviewed school-based experimental studies published in *JABA* between 1991 and 2005. The authors used the same criteria as Gresham, Gansle, and Noell (1993) to code operational definition and assessment of the independent variable, except that studies not providing an operational definition were coded as “footnote” if there was a reference to a more extensive source (e.g., book chapter, treatment manual). In addition, the criteria applied by Peterson et al. (1982) were used to rate the risk of treatment implementation inaccuracy. The results of this latest review indicated that 95% of the studies supplied an operational definition of the intervention within the article, and an additional 3% referenced an outside source containing more information. Only 39%, however, provided any mention of treatment integrity assessment, with 30% of all studies including quantitative data and 9% reporting having monitored integrity but providing no data. Thus, 61% of all the studies reviewed did not report treatment integrity assessment at all.

Taken together, the results of these three reviews indicate that during the past 40 years, the field of ABA has shown improvement with respect to defining and assessing the independent variable in treatment outcome research. More specifically, the rate of reporting an operational definition of a treatment increased more than 2.5-fold (34% in 1980–1990 to 95% in 1991–2005) whereas the rate of reporting treatment integrity assessment data increased nearly twofold (16% in 1980–1990 to 30% in 1991–2005). There is clearly room for further improvement, however, as even within the past decade, less than half of the articles provided any reference to treatment integrity whatsoever. Largely equivalent

results have been found in reviews of research in other fields (see Table 1).

In an era of increasing scrutiny with regard to accountability, particularly in the field of education where the PBS movement is gaining force, it is disconcerting that a substantial proportion of researchers still appear to ignore measurement of the independent variable. Such measurement is critical in both efficacy and effectiveness trials. In efficacy studies, treatment integrity data permit the researcher to ensure that the interventions were implemented exactly as planned. In effectiveness studies, when initially successful interventions are studied in more diverse applied settings, treatment integrity data are essential to understanding what (a) elements of the interventions are essential to their effectiveness and (b) adaptations can be made to the interventions without sacrificing effectiveness (Brown & Rahn-Blakeslee, 2009). Conducting and publishing the results of high-quality efficacy and effectiveness trials, which include treatment integrity assessment, are necessary to provide practitioners with the evidence-based interventions statutes such as the No Child Left Behind Act of 2001 and the Individuals with Disabilities Education Improvement Act of 2004 mandate them to implement.

As a relatively young journal quickly gaining momentum in the field of behavioral intervention, there is no better time for the *Journal of Positive Behavior Interventions (JPBI)* to be given the same critical review regarding treatment integrity assessment as has been done in other fields. As such, the primary purpose of the current study was to systematically code the treatment integrity data reported in studies conducted with children and published in *JPBI* between 1999 and 2009. Given that studies published in this journal include a diverse array of treatment agents, settings, target concerns, and student ages, a secondary purpose of this study was to summarize these characteristics of intervention outcome research in the field of PBS, as well as treatment integrity trends across them. Investigation of these variables may provide a better understanding of the foci of current research in the field and any related treatment integrity assessment trends.

## Method

### Criteria for Review

A total of 324 articles from *JPBI* were reviewed by the second and third authors, advanced graduate students in school psychology, to determine potential for inclusion. Articles for review were located through a serial search of the table of contents of each *JPBI* issue published between 1999 and 2009. There were three inclusion criteria, which are similar to those used by Sanetti, Gritter, and Dobey (2011). First, the article had to have been published between the journal's inception in 1999 and 2009. Second, as we were

interested in interventions implemented with youth, the participants in the study had to be younger than 19 years of age. Third, the study had to be experimental (i.e., an independent variable was manipulated to effect change in a dependent variable) and employ either a recognized between- or within-group design or a single-case design (i.e., reversal, multiple baseline, alternating treatments, changing criterion). To be considered experimental, single-case design studies had to include a baseline phase (except in cases in which such a phase was implied or unnecessary; e.g., student does not demonstrate the skill at all prior to intervention, alternating treatments design) and the opportunity for three demonstrations of effect (i.e., multiple-baseline design with at least three baselines). Demonstration studies, quasi-experimental studies, case studies, and single-case design studies (a) without a baseline phase when such a phase was necessary and was not implied or (b) that did not provide an opportunity for three demonstrations of effect were considered nonexperimental. The focus of this study is on intervention evaluation; therefore, studies that only included an evaluation of an assessment approach (e.g., functional assessment or analysis) were not included. Studies that presented an initial assessment followed by an experimental study evaluating an assessment-driven intervention were included.

After applying these three criteria during the serial search of each issue of *JPBI* between 1999 and 2009, 72 articles (22% of all reviewed articles) were included for further review. A full list of the included articles is presented in the appendix.

### Article Selection Reliability

To determine the reliability of article selection, the first author conducted independent rating of the inclusion criteria for 65 (20%) of the articles reviewed for inclusion in the study. A percentage agreement statistic was calculated (i.e., number of agreements divided by the number of agreements and disagreements, multiplied by 100%), yielding 100% agreement. In addition, the authors met weekly to discuss and resolve any questions about potential articles for inclusion.

### Coding

**Operational definition of the independent variable.** Studies were coded “yes,” “reference,” or “no” in answer to the question “Is the independent variable (intervention) operationally defined?” If one would be able to implement the intervention based on the description provided, the intervention was considered operationally defined. Articles in which authors referred to a source that provided more extensive and detailed information regarding the intervention (e.g., intervention manual, book chapter) were coded as “reference.” Articles that neither included an adequate

definition for replication nor a reference to another source were coded as “no.” These coding criteria are based on those used in previous reviews (Gresham, Gansle, & Noell, 1993; McIntyre et al., 2007).

**Assessment of treatment integrity.** Studies were coded “yes,” “monitored,” or “no” with regard to their inclusion of treatment integrity data. Articles that included quantitative treatment integrity data were coded as “yes.” More specifically, to be coded as such, studies had to report quantitative treatment integrity data. Studies for which authors mentioned treatment integrity in the narrative but did not provide quantitative data (e.g., “no deviations from intervention protocol were observed”) were coded as “monitored.” Studies that did not mention treatment integrity were coded as “no.” These coding criteria are based on those used in previous reviews (Gresham, Gansle, & Noell, 1993; McIntyre et al., 2007).

**Risk for implementation inaccuracies.** Interventions were coded as either “no,” “low,” or “high” risk for implementation inaccuracies. Interventions were coded as “no risk” if assessment of treatment integrity was coded as “yes” or “monitored.” Interventions were coded as “low risk” if treatment integrity assessment was coded as “no” but was judged to be at low risk for inaccuracies (e.g., computer-delivered interventions, permanent products such as posting classroom rules). Interventions were coded as “high risk” if treatment integrity assessment was coded as “no,” did not meet criteria for the “low risk” rating, and were judged to be at high risk for inaccuracies. These coding criteria are based on those used in previous reviews (McIntyre et al., 2007; Peterson et al., 1982).

**Publication year.** The publication year of each article was coded.

**Treatment agent.** Each individual responsible for implementing the intervention was coded into one of the following categories, based on those used by McIntyre et al. (2007): (a) teacher (e.g., general education, special education, early childhood), (b) professional (e.g., psychologists, speech language pathologists, school principals), (c) paraprofessional (e.g., teacher’s aides, members of the support staff), (d) parent, (e) sibling, (f) researcher (e.g., member of a research team collecting data for the purpose of a study), (g) peer tutors (i.e., students who were not the focus of the intervention), (h) self (i.e., self-administered or self-mediated interventions), or (i) other (i.e., individuals who did not fit into any of the above categories, such as volunteers). When more than one person was responsible for implementing the intervention, each treatment agent was coded.

**Location of intervention.** The location where the intervention took place was categorized into one of the following nine categories: (a) public school, (b) charter school, (c) private school (nonparochial), (d) parochial school, (e) residential school, (f) hospital school, (g) home, (h) not specified, or (i) “other,” for locations that did not fit any of these categories (e.g., Head Start, university-based lab school). When

the intervention was implemented in more than one location, each location was coded.

**Design.** Studies were coded with regard to experimental design: single-case or group design.

**Dependent variable.** In addition to information about the independent variable, all reported dependent variables were coded for each study. Each dependent variable was coded as belonging to one of eight categories. Academic achievement included behaviors related to the acquisition or performance of academic skills (e.g., reading, math, spelling). Stereotypic behaviors included constant repetitive behaviors such as vocalizations, self-injurious behaviors, and hand flapping. Disruptive behaviors included behaviors that led to disruptions of social environments, such as a student being noncompliant with teacher directions or talking without permission. Social skills included, for example, social initiations, play behaviors, and sharing behaviors. Academic-related behaviors were defined as those that were related to academic outcomes, such as academic engagement or having appropriate materials. Behaviors that were coded as daily living skills included, for example, those related to safety, independent functioning, mobility, and eating. Psychological well-being included, for example, self-esteem, self-concept, or psychological adjustment. “Other” dependent variables included, for example, sleep behaviors and social validity.

**Student age.** Participants’ ages were coded as belonging to one of the following categories: (a) preschool age (0–4 years), (b) elementary age (5–12 years), or (c) high-school age (13–18 years). When participants’ ages spanned across these categories, all relevant categories were coded.

### **Rater Training and Interrater Agreement**

The rater training procedures are the same as those used by Sanetti et al. (2011). Across three sessions (one 1-hr and two 2-hr), the first author, a faculty member in school psychology, trained two advanced graduate students in school psychology how to apply the coding criteria. The first training session included (a) providing each rater with a coding handbook that defined each category and code; (b) reviewing the handbook; (c) discussing the categories, codes, and definitions with the raters; and (d) providing the raters with four experimental treatment outcome studies (published in school psychology journals prior to 1995) to code before the second training session. The second session included (a) discussing the coding results, difficulties with coding, and confusing codes; (b) modifying the coding scheme to increase clarity; and (c) providing the raters with four additional experimental treatment outcome articles to code using the modified coding scheme before the third session. The third session included (a) discussing the results of applying the results of the modified coding scheme, (b) reaching 100% agreement via consensus after discussing disagreements in

ratings, and (c) assigning each rater approximately half of the articles in the database.

All studies were initially coded by the second and third authors. To determine interrater agreement, the first author independently coded a random sample of 26.4% studies meeting the inclusion criteria. Studies were coded on nine categories. Each category and the number of possible categories on which it could have been coded (e.g., the two categories for experimental design were single-case and group design) are as follows: (a) experimental design (2 categories), (b) year of publication (11 categories), (c) student age (3 categories), (d) location of the intervention (9 categories), (e) treatment agent (9 categories), (f) operational definition of the independent variable (3 categories), (g) treatment integrity assessment (3 categories), (h) risk for inaccuracies (3 categories), and (i) dependent variable (8 categories). There was perfect agreement for seven categories: experimental design, year of publication, student age, location of intervention, operational definition of the independent variable, treatment integrity assessment, and risk for inaccuracies. Interrater reliability for the remaining categories in the coding scheme was estimated using percentage agreement (i.e., number of agreements/number of agreements + number of disagreements  $\times$  100). Percentage agreement was 99.4% across all codes, 99.5% for treatment agent, and 96.8% for dependent variable.

### **Data Analysis**

Chi-square analyses were conducted to analyze the presence of operational definitions of the independent variable and treatment integrity data in the reviewed studies across the following variables: year of publication, student age, treatment agent, location, and dependent variable. Some studies included more than one student age, treatment agent, location, and/or dependent variable. In these cases, the variables were coded into mutually exclusive categories. For example, student age was represented by three categories: preschool, elementary, and high school. If all of the study participants fell within one of these categories, the relevant category was coded. If study participants were from more than one age category, “multiple ages” was coded. Furthermore, due to small cell sizes (i.e., fewer than 5 observations per cell), we collapsed categories for both operational definition of the intervention and treatment integrity assessment. For the former, the categories “operational definition,” and “referenced” were collapsed together (both allow readers to obtain a description of the intervention) and compared to the category “not defined.” For the latter, “assessed” and “monitored” were collapsed together (both represent a decreased risk for inaccurate implementation of the intervention) and compared to “not assessed.” Study characteristics also required collapsing to ensure appropriate cell size. More specifically, (a) years of publication were collapsed

into four 2- to 3-year blocks (i.e., 1999–2001, 2002–2004, 2005–2007, 2008–2009); (b) student ages were collapsed to elementary or other age(s); (c) treatment agents were collapsed into five categories: teachers, researchers, parents, multiple treatment agents, or others (i.e., professionals, paraprofessionals, peers, self, and “other” combined); (d) location of intervention was collapsed to public school or other location(s); and (e) dependent variables were collapsed into five categories: disruptive behavior, social skills, daily living skills, multiple dependent variables, and “other” (i.e., academic achievement, stereotypic behaviors, academic-related behaviors, psychological well-being, and “other” combined). Chi-square analyses were conducted for all coded categories once the above-mentioned categories were collapsed. Unless otherwise stated, the .05 level of statistical significance was used in all analyses.

## Results

A majority of the coded studies were conducted in elementary schools ( $n = 59$ , 81.9%). Teachers ( $n = 31$ , 43.1%), researchers ( $n = 26$ , 36.1%), and parents ( $n = 20$ , 27.8%) were the most common treatment agents. Most interventions were implemented in a public school ( $n = 35$ , 48.6%), home ( $n = 20$ , 27.8%), or other location ( $n = 25$ , 34.7%). All but one of the studies were carried out using a single-case design methodology ( $n = 71$ , 98.6%). Disruptive behaviors ( $n = 33$ , 45.8%), social skills ( $n = 23$ , 31.9%), and other behaviors ( $n = 19$ , 26.4%) were the most frequently targeted dependent variables. Additional data for each study characteristic are displayed in Table 2.

Of the 72 studies coded, 59.7% ( $n = 43$ ) included an operational definition of the independent variable, 26.4% ( $n = 19$ ) provided a reference to a more detailed definition of the independent variable, and 13.9% ( $n = 10$ ) did not include an operational definition of the independent variable. As illustrated in Figure 1, there was a decreasing trend in the percentage of articles that included an operational definition of the independent variable across years studied. Furthermore, there was a slight increasing trend in the percentage of articles that included a reference to an operational definition of the independent variable. The prevalence of operational definitions of the independent variable by student age, year, treatment agent, location, and dependent variable are presented in Table 3. Chi-square analyses revealed no significant differences in the provision of an operational definition of the intervention related to these variables.

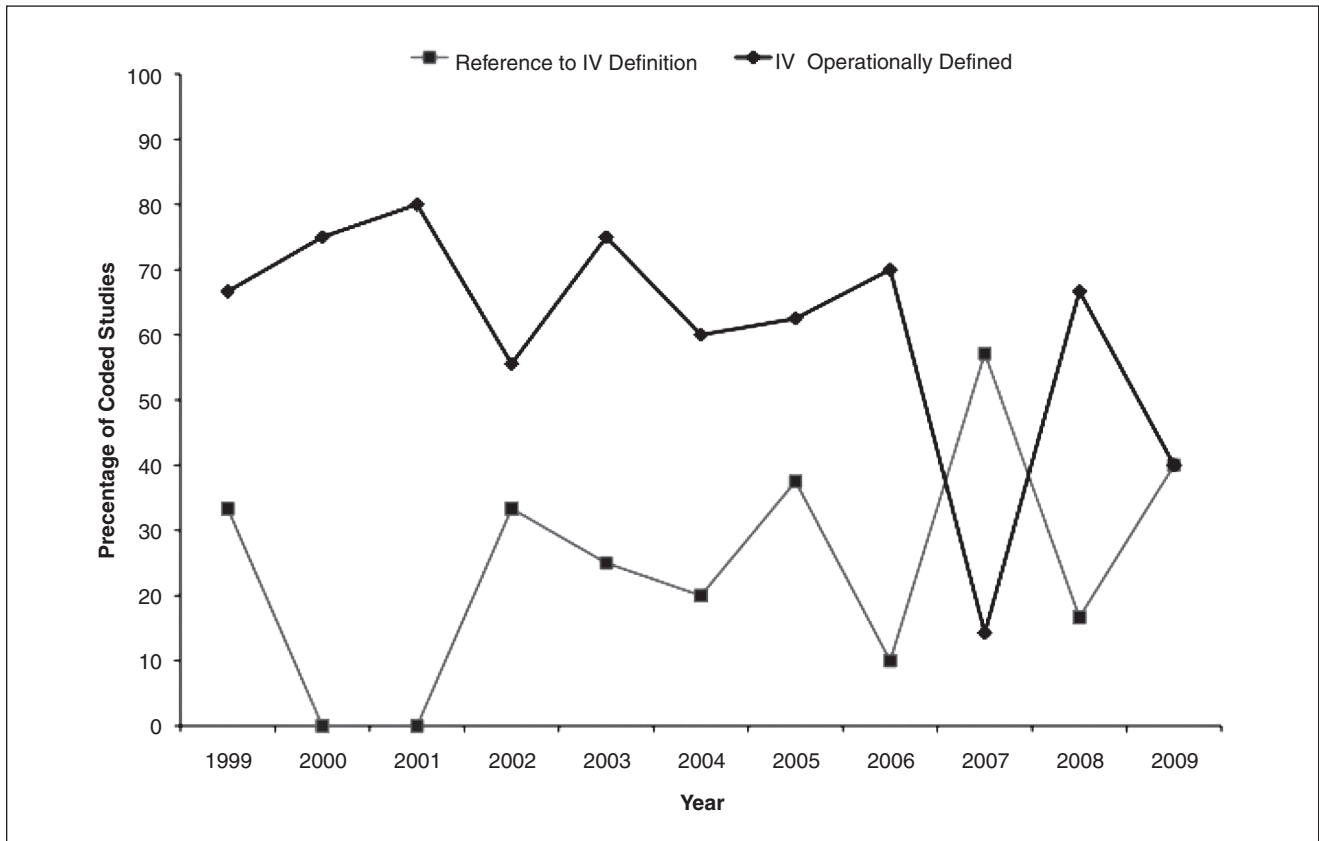
With regard to the prevalence of treatment integrity assessment data in the coded studies, 41.7% ( $n = 30$ ) of the studies included quantitative treatment integrity data, 6.9% ( $n = 5$ ) included a statement about treatment integrity assessment but did not provide quantitative data, and 51.4% ( $n = 37$ ) neither monitored nor assessed treatment integrity. Figure 2 illustrates the rates of treatment integrity assessment

**Table 2.** General Characteristics of Coded Studies

Study characteristic	<i>n</i>	%
<b>Age</b>		
Preschool	16	22.2
Elementary	59	81.9
High school	14	19.4
Total	89	
<b>Treatment agent</b>		
Teacher	31	43.1
Professional	9	12.5
Paraprofessional	12	16.7
Parent	20	27.8
Sibling	2	2.8
Researcher	26	36.1
Peer tutor	9	12.5
Self	10	13.9
Other	13	18.1
Total	132	
<b>Location</b>		
Public school	35	48.6
Charter school	1	1.4
Private school	5	6.9
Parochial school	0	0.0
Hospital school	0	0.0
Residential school	0	0.0
Home	20	27.8
Other	25	34.7
Not specified	0	0.0
Total	86	
<b>Experimental design</b>		
Group	1	1.4
Single case	71	98.6
Total	72	
<b>Dependent variable</b>		
Academic achievement	8	11.1
Stereotypic behaviors	10	13.9
Disruptive behaviors	33	45.8
Social skills	23	31.9
Academic-related behaviors	16	22.2
Daily living skills	10	13.9
Psychological well-being	4	5.6
Other	19	26.4
Total	123	

Note. Totals for each category may not sum to 72 as some studies included multiple age ranges, treatment agents, locations, and dependent variables.

across studied years. A steep increasing trend in reporting quantitative treatment integrity data is evident from 2003 to 2006, with a decline in 2007–2008. The reporting of quantitative treatment integrity data increased again in 2009. In 4 of the 11 years of coded studies, a relatively small percentage (10%–33%) of authors mentioned treatment integrity



**Figure 1.** Definition of the independent variable by year

but failed to provide quantitative data. Chi-square analyses indicated a significant relationship between year of publication and treatment integrity assessment,  $\chi^2(3) = 16.6, p = .001$ . Follow-up analyses were conducted for all possible  $2 \times 2$  relationships, and a Bonferroni correction was applied ( $p < .008$ ). Results indicated that treatment integrity was significantly more likely to be assessed during 2005–2007,  $\chi^2(1) = 13.7, p < .001$ , and 2008–2009,  $\chi^2(1) = 11.5, p = .001$ , than during 1999–2001. The prevalence of treatment integrity assessment by student age, year, treatment agent, location, and dependent variable is presented in Table 4. Chi-square analyses revealed no significant differences in the reporting of treatment integrity data related to these variables.

## Discussion

The primary purpose of this study was to examine studies conducted with children and published in *JPBI* between 1999 and 2009 to determine the extent to which researchers provided an operational definition of the independent variable and reported quantitative treatment integrity data. A secondary purpose of this study was to examine treatment integrity trends across various characteristics of these

published studies, such as treatment agent, setting, target concern, and student age.

With respect to providing an operational definition of the independent variable, the data revealed that a large majority (86.11%) of researchers included an operational definition of the independent variable or a reference to a more detailed source. This is a slightly lower rate of providing an operational definition than McIntyre et al. (2007) and Wheeler et al. (2006) found in recent reviews of the ABA and autism literatures. It is, however, significantly higher than rates found in a majority of previous reviews (Dane & Schneider, 1998; Gresham, Gansle, & Noell, 1993; Gresham, Gansle, Noell, & Cohen, 1993; Perepletchikova et al., 2007). These results are promising, yet there is room for improvement as access to an operational definition of the intervention being evaluated is an important component of every treatment outcome study report. There are negative consequences for both researchers and practitioners when operational definitions of interventions are not provided in treatment outcome studies. With regard to research, conducting a replication study to independently validate an intervention's effectiveness is much more difficult without access to the operational definition of the independent variable in the original study. With regard to practice, lack of access to research is



**Table 3.** Operational Definition of the Independent Variable by Study Characteristics

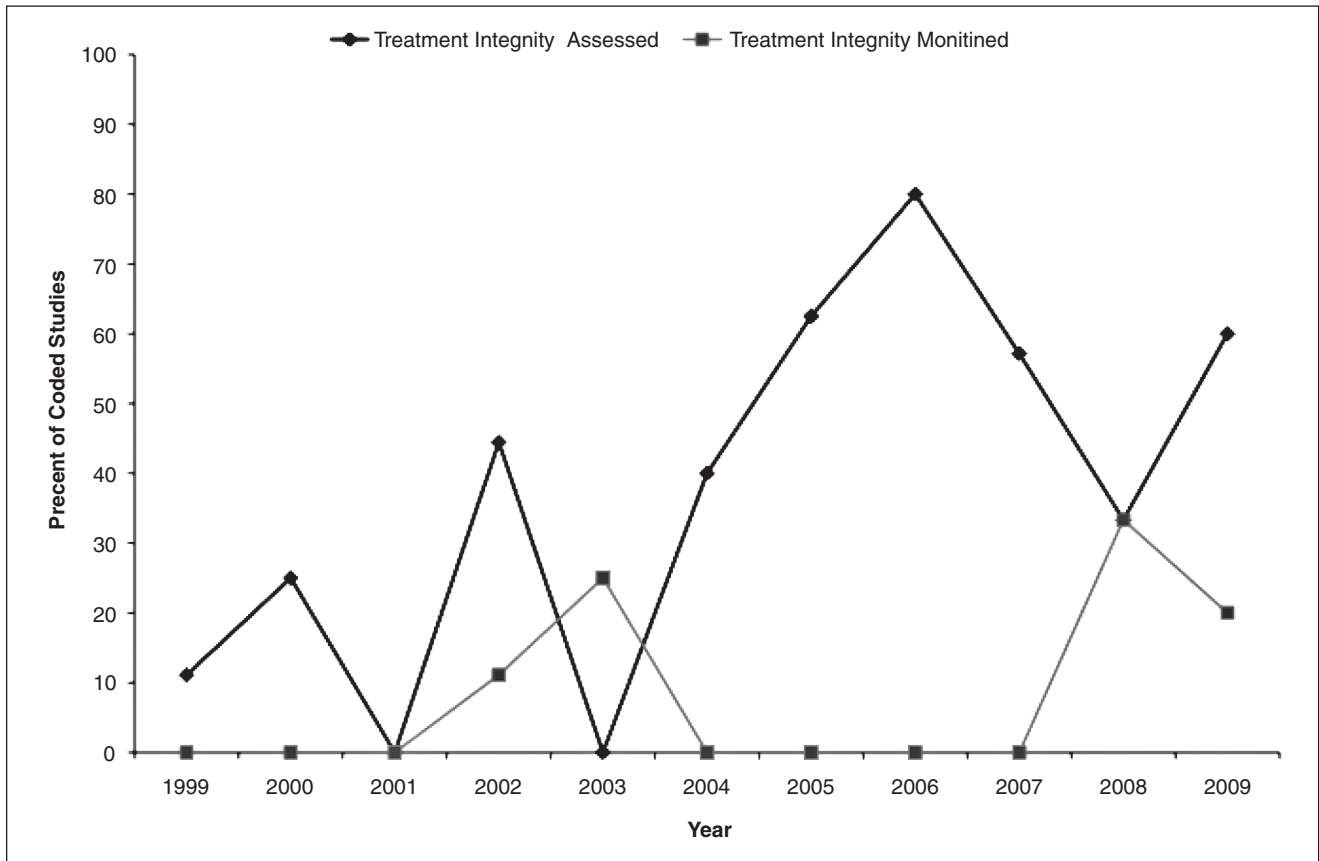
Study characteristic	Operational definition of the independent variable				Total <i>n</i>
	Defined or referenced		Not defined		
	<i>n</i>	%	<i>n</i>	%	
<b>Student age</b>					
Elementary	37	51.39	5	6.94	42
Not elementary/multiple	25	34.72	5	6.94	30
Total <i>n</i>	62	86.11	10	13.88	72
<b>Year</b>					
1999–2001	16	22.22	2	2.78	18
2002–2004	16	22.22	2	2.78	18
2005–2007	21	29.17	4	5.55	25
2008–2009	9	12.50	2	2.78	11
Total <i>n</i>	62	86.11	10	13.89	72
<b>Treatment agent</b>					
Teacher	7	9.72	0	0.0	7
Researcher	5	6.94	1	1.39	6
Parents	5	6.94	1	1.39	6
Other	10	13.89	1	1.39	11
Multiple	35	48.61	7	9.72	42
Total <i>n</i>	62	86.10	10	13.89	72
<b>Location</b>					
Public	29	40.28	2	2.78	31
Other	33	45.83	8	11.11	41
Total <i>n</i>	62	86.11	10	13.89	72
<b>Dependent variable</b>					
Disruptive behaviors	8	11.11	2	2.78	10
Social skills	7	9.72	1	1.39	8
Daily living skills	6	8.33	0	0.00	6
Other	10	13.89	1	1.39	11
Multiple	31	43.05	6	8.33	37
Total <i>n</i>	62	86.10	10	13.89	72

Note. Study characteristic and operational definition of the independent variable categories align with those used in the chi-square analyses.

a commonly cited speculation as to why the research-to-practice gap exists (e.g., Burns & Ysseldyke, 2009). Interventions with positive research support may be less accessible to clinicians, educators, or others who want to apply the intervention in a practical setting. If a clinician cannot, relatively easily, determine precisely what an intervention was and how it was implemented, he or she may be less likely to pursue implementing that intervention with a client. Thus, when researchers provide an operational definition of the independent variable, it allows researchers and clinicians access to the information they need to replicate and use an intervention and thus potentially further lines of research and help close the research-to-practice gap.

Visual analysis of the trend of providing an operational definition for the independent variable provides interesting information about access to operational definitions of interventions. The data from this study revealed a decreasing

trend in the percentage of articles that included an operational definition of the independent variable across the years studied. In contrast, there was a slight increasing trend in the percentage of articles that included a reference to an operational definition of the independent variable. These results differ from those found by Gresham, Gansle, and Noell (1993), which indicated a highly variable pattern of inclusion of an operational definition of the independent variable in *JABA* studies between 1980 and 1990. These authors did not code “referenced” definitions, so no comparison is able to be made. Likewise, McIntyre et al. (2007) did not report data regarding operational definition of the independent variable by year, so comparisons of trends are not possible. However, overall results from McIntyre et al. suggest that articles in *JABA* are more likely to include an operational definition of the independent variable than a reference to another source.



**Figure 2.** Treatment integrity assessment by year

Although the trend in *JPBI* of providing a reference to an operational definition is preferable to not providing any information about the independent variable, there is a problem with this situation as well; unless the intervention as described in the referenced material is identical to the one evaluated in the publication, it is not useful to the consumer for the purposes of replication. Additionally, unless the referenced material is accessible to all readers, and readers can access the referenced material easily and quickly, the effect of providing a reference to the operational definition may be similar to not providing an operational definition at all.

In addition to providing the reader with an operational definition of the independent variable, the other component in treatment integrity reporting is providing the reader with quantitative treatment integrity data. With respect to this element of treatment integrity, the percentage of studies reporting treatment integrity in *JPBI* is higher than all previous reviews (Dane & Schneider, 1998; Gansle, 2005; Gresham et al, 2000; Gresham, Gansle, & Noell, 1993; Gresham, Gansle, Noell, & Cohen, 1993; McIntyre et al., 2007; Perepletchikova et al., 2007; Peterson et al., 1982; Snell et al., 2006; Wheeler et al., 2006). That said, still fewer than half of the studies reviewed reported quantitative

treatment integrity data, indicating there is clearly room for improvement.

It is essential that researchers report quantitative data on treatment integrity because without this information, it is difficult to truly determine the efficacy or the effectiveness of an intervention. It is often assumed that treatments are or will be implemented as planned and outcomes are a result of high levels intervention implementation (Gresham, 1989). Results of multiple school-based intervention studies, however, indicate that a vast majority of teachers displayed decreasing or low treatment integrity levels (range: 0%–65% intervention steps implemented as planned) within 1 to 10 days after training (Mortenson & Witt, 1998; Noell, Witt, Gilbertson, Ranier, & Freeland, 1997; Sanetti & Kratochwill, in press, 2009b; Witt, Noell, LaFleur, & Mortenson, 1997). These data make clear that assuming high levels of implementation is inappropriate. Furthermore, results of a review of the impact of implementation on outcomes suggest that intervention adaptations may be more likely to improve intervention effectiveness than decrease it (Durlack & DuPre, 2008). Together these results further underscore the need for researchers to provide quantitative treatment integrity data. Doing so will enable us to learn

**Table 4.** Treatment Integrity Assessment by Study Characteristics

Study characteristic	Treatment integrity assessment				Total <i>n</i>
	Assessed or monitored		Not assessed		
	<i>n</i>	%	<i>n</i>	%	
<b>Student age</b>					
Elementary	21	29.17	21	29.17	42
Not elementary/multiple	14	19.44	16	22.22	30
Total <i>n</i>	35	48.61	37	51.39	72
<b>Year</b>					
1999–2001	2	2.78	16	22.22	18
2002–2004	8	11.11	10	13.89	18
2005–2007	17	23.61	8	11.11	25
2008–2009	8	11.11	3	4.17	11
Total <i>n</i>	35	48.61	37	51.39	72
<b>Treatment agent</b>					
Teacher	3	4.17	4	5.55	7
Researcher	1	1.39	5	6.94	6
Parents	4	5.55	2	2.78	6
Other	6	8.33	5	6.94	11
Multiple	21	29.17	21	29.17	42
Total <i>n</i>	35	48.61	37	51.38	72
<b>Location</b>					
Public	19	26.39	12	16.67	31
Other	16	22.22	25	34.72	41
Total <i>n</i>	35	48.61	37	51.39	72
<b>Dependent variable</b>					
Disruptive behaviors	4	5.55	6	8.33	10
Social skills	5	6.94	3	4.17	8
Daily living skills	2	2.78	4	5.55	6
Other	6	8.33	5	6.94	11
Multiple	18	25.00	19	26.39	37
Total <i>n</i>	35	48.60	37	51.38	72

Note. Study characteristic and treatment integrity assessment categories align with those used in the chi-square analyses.

more about which intervention components need to be implemented as planned and how or how much an intervention can be adapted while still achieving effective results (Gansle & McMahon, 1997; Gresham, 1989).

This has particular relevance for the field of PBS. Given the focus of PBS on ecological validity and naturalistic contexts (Carr et al., 2002; Sugai et al., 2000) it is essential that interventionists be able to adapt an intervention based on the needs and contexts of the student, client, child, classroom, or system with which they are working. Without quantitative treatment integrity data, there is no way to empirically determine the essential components of an intervention, and therefore no data-based way to know whether or how best to adapt an intervention.

In contrast to the decreasing trend of providing an operational definition of the independent variable, the trend of

reporting quantitative treatment integrity data increased considerably from 2003 to 2006, declined in 2007 and 2008, and increased again in 2009. The reason(s) for the variability during the last 3 years reviewed is unclear. Regardless, the data trend suggests that in recent years treatment integrity data are still not reported as frequently as is ideal, but more researchers are cognizant of the need to report treatment integrity data.

### Limitations

Several limitations need to be considered when interpreting our results. First, we applied fairly stringent criteria for inclusion as an experimental study, especially for single-case studies (i.e., opportunity for three demonstrations of effect, appropriate baseline). This may have resulted in an elevated

level of treatment integrity reporting, with studies with stronger methodology being more likely to address treatment integrity. Second, to conduct chi-square analyses, we had to collapse coding categories to have adequate cell sizes. As a result, we might have missed differences in more discrete categories (e.g., operationally defined and referenced) that we were unable to analyze because of cell size. Third, when an author provided a reference for a definition of an intervention, we did not obtain and review the referenced publication to determine the adequacy or the accessibility of the definition. Fourth, similar to other reviews, an operational definition was considered adequate if we believed we could replicate the intervention solely based on the description provided. As we are very familiar with the intervention literature base, it is possible that our decisions would not generalize to others (e.g., clinicians, parents) who are less familiar.

### *Implications for Research and Practice*

The operational definition of the intervention and assessment of treatment integrity are essential, but often overlooked, components of the research process. It makes intuitive sense that one cannot draw conclusions about the impact of an intervention on a behavior if one is not certain what the intervention is or if, or to what extent, the intervention was implemented as planned. The reporting of treatment integrity data, then, should be an essential component of any intervention study. The inclusion of intervention definitions can only improve the research base by explicitly providing researchers with the steps necessary to replicate an intervention. Likewise, inclusion of quantitative treatment integrity data will allow researchers to determine which parts of the intervention may be essential to enact behavior change. Presently, however, both types of treatment integrity are often left out of research publications both in this and other journals (Sanetti et al., 2011).

There are several potential reasons why treatment integrity data are not as prevalent in the treatment outcome literature as, for example, dependent variable data. Three such reasons are presented below. First, treatment integrity is a methodological concern that has been largely overlooked for decades (Sanetti & Kratochwill, 2009a). In recent years, however, there has been a flurry of scholarly activity in this area and multiple conceptual models of the construct have been proposed. These models have advanced our thinking about treatment integrity and suggest treatment integrity is a more complex construct than initially thought (e.g., Dane & Schneider, 1998; Fixsen, Naoom, Blasé, Friedman, & Wallace, 2005; Jones, Clarke, & Power, 2008; Noell, 2008; Power et al., 2005; Waltz, Addis, Koerner, & Jacobson, 1993). Despite these advances, however, there is still a lack of consensus on issues as central as the definition of treatment

integrity (Sanetti & Kratochwill, 2009a). As a result, researchers may rightfully be confused about what assessing treatment integrity really means for their research. Second, there are very few psychometrically sound treatment integrity assessments available to researchers, forcing researchers to create checklists or other methods for use in their research (Sanetti & Kratochwill, 2009a). Third, strict page limits in journals may be affecting the types of data researchers report or are asked to report by journal editors. As most consumers of research have access to the Internet, editors may consider allowing authors to publish some data on-line only so that page limits can be adhered to while still providing consumers with all necessary data.

The omission of treatment integrity data in research publications has implications not only for the researcher, but also for the practitioner. A practitioner seeking the best outcomes for a client should have access not only to a research base but to a research base that provides the information necessary for that practitioner to replicate an effective intervention (i.e., treatment integrity data). With regard to intervention definition, although a practitioner could contact the author of a study for a more detailed description of the intervention, it is unlikely that most would have the time or the inclination to do so. Considering the increasing trend in *JPBI* for authors to reference a definition of an intervention, it seems imperative that during the manuscript review process, the referenced material is reviewed. For example, reviewers and editors might ensure that (a) the reference provides an operational definition of the intervention, (b) the intervention was implemented in the study exactly as defined in the referenced source, and (c) the referenced source is accessible to readers. With regard to quantitative treatment integrity data, interventions in applied settings are often adapted to the needs of the client or of the setting (Larsen & Samdal, 2007; Ringwalt et al., 2003). When treatment integrity assessment data are available, practitioners may be able to make data-based decisions about which intervention components are essential in order to obtain the best outcomes for their clients, and therefore how to adapt an intervention appropriately based on their needs.

PBS is a field dedicated to achieving positive behavior change through the application of research-validated interventions in relevant, real-world contexts (Carr et al., 2002; Sugai et al., 2000). In order to remain true to these principles, researchers must ensure that interventions are replicated and validated across different settings and populations, and that their essential components are determined. Advancing the field of PBS, then, requires a focus on treatment integrity data. Both researchers and practitioners can use the tools available to them to contribute to this effort by developing treatment integrity assessments, keeping data on treatment integrity, and reporting these data as appropriate.

## Appendix

Table of Reviewed Studies

Article	Operational definition of treatment	Treatment integrity assessment	Risk for treatment inaccuracy	Treatment agent (all that apply)	Location of intervention	Design (single case/group)	Dependent variable (all that apply)	Student age (all that apply)
Todd, Horner, & Sugai (1999)	Yes	No	High	Teacher, professional	Public school	Single case	Disruptive, academic-related behaviors	Elementary
Dunlap & Fox (1999)	Reference	No	High	Teacher, parent, other	Home, other	Single case	Disruptive behaviors	Preschool
Shukla, Kennedy, & Cushing (1999)	Reference	No	High	Peer tutor	Public school	Single case	Social skills, academic-related behaviors	Elementary, high school
Smith & Camarata (1999)	Reference	Yes	No	Teacher	Public school	Single case	Social skills, daily living skills	Preschool, elementary
Stiebel (1999)	Yes	No	High	Parent	Home, community	Single case	Daily living skills	Preschool, elementary
Lee, Sugai, & Horner (1999)	Yes	No	High	Researcher	Public school	Single case	Academic achievement, disruptive, academic-related behaviors	Elementary
Buggey, Toombs, Gardner, & Cervetti (1999)	Yes	No	Low	Researcher, self	Home	Single case	Social skills	Elementary
Hinton & Kern (1999)	Yes	No	Low	Teacher, researcher	Public school	Single case	Academic-related behaviors	Elementary
Clarke, Dunlap, & Vaughn (1999)	Yes	No	High	Researcher, parent, sibling, other	Home	Single case	Disruptive behaviors, daily living skills	Elementary
Schreibman, Whalen, & Stahmer (2000)	Yes	No	High	Parent, researcher	Home, other	Single case	Disruptive behaviors	Preschool, elementary
Baker (2000)	Yes	No	High	Other	Public school	Single case	Social skills, other	Elementary
Charlop-Christy & Carpenter (2000)	Yes	Yes	No	Parent	Home	Single case	Social skills	Elementary
Smith & Sugai (2000)	No	No	High	Researcher, self	Other	Single case	Disruptive, academic-related behaviors	High school
Kern, Mantegna, Vordran, Bailin, & Hilt (2001)	Yes	No	High	Teacher, professional	Private school, other	Single case	Disruptive, stereotypic, academic-related behaviors	Elementary, high school
Lohrmann-O'Rourke & Yurman (2001)	Yes	No	Low	Teacher, paraprofessional, researcher	Public school	Single case	Stereotypic behaviors	Elementary
Barry & Singer (2001)	Yes	No	High	Parent, sibling, researcher, self	Home	Single case	Social skills, stereotypic behaviors	Elementary
Carter (2001)	No	No	High	Researcher	Other	Single case	Social skills, disruptive behaviors, other	Elementary

(continued)

## Appendix (continued)

Article	Operational definition of treatment	Treatment integrity assessment	Risk for treatment inaccuracy	Treatment agent (all that apply)	Location of intervention	Design (single case/group)	Dependent variable (all that apply)	Student age (all that apply)
Morrison, Kamps, Garcia, & Parker (2001)	Yes	No	High	Researcher	Public school	Single case	Social skills, disruptive, stereotypic behaviors	Elementary, high school
Cole & Levinson (2002)	Yes	Yes	No	Paraprofessional	Other	Single case	Disruptive behaviors	Elementary
Vaughn, Wilson, & Dunlap (2002)	Yes	Yes	No	Parent	Other	Single case	Disruptive, stereotypic behaviors, daily living skills	Elementary
Lorimer, Simpson, Myles, & Ganz (2002)	Yes	No	High	Professional, parent	Home	Single case	Disruptive behaviors	Elementary
Durand (2002)	Reference	Monitored	No	Parent	Home	Single case	Other	Preschool, elementary
Nuzzulo-Gomez, Leonard, Ortiz, Rivera, & Greer (2002)	Yes	No	High	Other	Private school	Single case	Stereotypic behaviors, other	Preschool, elementary
Koegel, Symon, & Koegel (2002)	Reference	Yes	No	Parent	Other	Single case	Other	Preschool, elementary
Clarke, Worcester, Dunlap, Murray, & Bradley-Klug (2002)	Reference	No	High	Teacher, paraprofessional, peer tutor	Public school	Single case	Disruptive, stereotypic, academic-related behaviors, psychological well-being	Elementary
Radford & Ervin (2002)	No	Yes	No	Teacher, peer tutor	Public school	Single case	Disruptive behaviors	High school
Shipley-Benamou, Lutzker, & Taubman (2002)	Yes	No	High	Researcher	Public school	Single case	Daily living skills	Elementary
D'Ateno, Mangiapanello, & Taylor (2003)	Yes	No	High	Other	Private school	Single case	Social skills	Preschool
Charlop-Christy & Daneshvar (2003)	No	No	Low	Researcher, other	Other	Single case	Social skills	Elementary
Wert & Neisworth (2003)	Yes	No	Low	Professional, parent, self	Home	Single case	Daily living skills	Preschool, elementary
Brooks, Todd, Tofflemoyer, & Horner (2003)	Yes	No	High	Teacher, researcher, self	Public school	Single case	Academic-related behavior	Elementary
Barry & Messer (2003)	Reference	Monitored	No	Teacher, self	Public school	Single case	Academic achievement, disruptive, academic-related behaviors	Elementary
Umbreit, Lane, & Dejud (2004)	Yes	Yes	No	Paraprofessional	Public school	Single case	Academic-related behaviors	Elementary

(continued)

## Appendix (continued)

Article	Operational definition of treatment	Treatment integrity assessment	Risk for treatment inaccuracy	Treatment agent (all that apply)	Location of intervention	Design (single case/group)	Dependent variable (all that apply)	Student age (all that apply)
Lohrmann & Talerico (2004)	Yes	No	High	Teacher	Other	Single case	Disruptive, academic-related behaviors	Elementary
Stichter, Sasso, & Jolivette (2004)	No	No	High	Teacher, peer tutor	Public school, other	Single case	Disruptive behaviors	Elementary
Brookman-Frazer (2004)	Reference	No	High	Professional, parent	Home, other	Single case	Social skills, psychological well-being, other	Preschool
Johnson, McDonnell, Holwarth, & Hunter (2004)	Yes	Yes	No	Teacher, paraprofessional	Public school	Single case	Academic behaviors	Elementary
Rock (2005)	Reference	No	High	Researcher	Public school	Single case	Academic achievement, disruptive, academic-related behaviors	Elementary, high school
McCartney, Anderson, & English (2005)	Yes	Yes	No	Professional, parent	Home, other	Single case	Daily living skills	Preschool, elementary
Apple, Billingsley, & Schwartz (2005)	Yes	Yes	No	Teacher, peer tutor, researcher, self	Public school	Single case	Social skills	Preschool, elementary
Cameron, Shapiro, & Ainsleigh (2005)	Yes	No	High	Parent, other	Home, other	Single case	Other	Elementary
Symon (2005)	Reference	Yes	No	Parent, researcher	Home	Single case	Social skills, daily living skills	Preschool, elementary
Moore, Anderson, Kumar (2005)	Yes	No	Low	Teacher	Public school	Single case	Disruptive behaviors	Elementary
Ingram, Lewis-Palmer, & Sugai (2005)	Reference	Yes	No	Teacher, researcher	Public school	Single case	Disruptive behaviors	Elementary, high school
Park, Singer, & Gibson (2005)	Yes	Yes	No	Teacher, professional	Public school	Single case	Academic behaviors, psychological well-being	Elementary
Kay, Harchik, & Luiselli (2006)	Yes	Yes	No	Paraprofessional, researcher	Public school, other	Single case	Daily living skills	High school
Delano & Snell (2006)	Reference	Yes	No	Peer tutor, researcher	Public school	Single case	Social skills, disruptive behaviors	Elementary
Sansosti, & Powell-Smith (2006)	Yes	Yes	No	Parent, self	Home	Single case	Social skills	Elementary
Lambert, Cartledge, Heward, & Lo (2006)	Yes	Yes	No	Teacher	Public school	Single case	Academic achievement, disruptive behaviors, academic-related behaviors	Elementary

(continued)

## Appendix (continued)

Article	Operational definition of treatment	Treatment integrity assessment	Risk for treatment inaccuracy	Treatment agent (all that apply)	Location of intervention	Design (single case/group)	Dependent variable (all that apply)	Student age (all that apply)
Maione & Miranda (2006)	Yes	Yes	No	Professional, parent, researcher	Home	Single case	Social skills	Elementary
Brown & Miranda (2006)	Yes	Yes	No	Paraprofessional	Public school	Single case	Disruptive behaviors	High school
Stahr, Cushing, Lane, & Fox (2006)	Yes	Yes	No	Teacher, professional, paraprofessional, researcher, self	Other	Single case	Academic-related behavior	Elementary
English & Anderson (2006)	No	Yes	No	Teacher, other	Other	Single case	Disruptive behaviors, stereotypic behaviors	Elementary
Hughes, Alberto, & Fredrick (2006)	Yes	No	High	Researcher, other	Other	Single case	Disruptive behaviors	High school
Westerlund, Granucci, Gamache, & Clark (2006)	No	No	High	Peer tutor	Other	Single case	Other	High school
Lee, Odom, & Loftin (2007)	Reference	Yes	No	Peer tutor	Public school	Single case	Social skills, stereotypic behaviors	Elementary
Hawken, MacLeod, & Rawlings (2007)	Yes	Yes	No	Teacher, paraprofessional	Public school	Single case	Other	Elementary
Lucyshyn, Albin, Horner, Mann, Mann, & Wadsworth (2007)	No	No	High	Parent	Home, other	Single case	Disruptive behaviors, other	Elementary, high school
Hundert (2007)	Reference	Yes	No	Teacher	Public school	Single case	Social skills, academic-related behaviors	Preschool
Butler & Luiselli (2007)	Reference	No	High	Teacher	Private school	Single case	Disruptive behaviors, stereotypic behaviors	High school
Vismara & Lyons (2007)	Reference	Yes	No	Parent, researcher	Home, other	Single case	Social skills, psychological well-being	Preschool
Todd, Campbell, Meyer, & Horner (2008)	Yes	Monitored	No	Teacher, other	Public school	Single case	Disruptive behaviors, other	Elementary
Ganz, Bourgeois, Flores, & Campos (2008)	Yes	Yes	No	Other	Private school	Single case	Social skills	Elementary/high school
Feeney & Yvisaker (2008)	Yes	No	High	Teacher, paraprofessional	Public school	Single case	Disruptive behaviors, Other	Elementary
Franzen & Kamps (2008)	No	No	High	Teacher	Charter school	Single case	Disruptive behaviors, Other	Elementary

(continued)



### Appendix (continued)

Article	Operational definition of treatment	Treatment integrity assessment	Risk for treatment inaccuracy	Treatment agent (all that apply)	Location of intervention	Design (single case/group)	Dependent variable (all that apply)	Student age (all that apply)
Sansosti & Powell-Smith (2008)	Reference	Yes	No	Teacher, paraprofessional, other	Public school	Single case	Social skills, Other	Elementary
Feng, Lo, Tsai, & Cartledge (2008)	No	Yes	No	Teacher, peer tutor	Public school	Single case	Social skills, Other	Elementary
McLaren & Nelson (2009)	Yes	Yes	No	Teacher, paraprofessional	Head start	Single case	Disruptive behavior, Other	Preschool
Carter & Horner (2009)	Reference	Yes	No	Teacher	Public school	Single case	Disruptive behavior, social skills, academic-related behavior, other	Elementary
Horner et al. (2009)	Reference	Monitored	No	Other	Public school	Group	Academic achievement, Other	Elementary
Nikopoulos, Canavan, & Nikopoulou-Smyrni (2009)	Yes	No	High	Researcher	Other	Single case	Daily living skills	Elementary
Sprague & Perkins (2009)	Reference	Yes	No	Teacher, parent, researcher	Public school	Single case	Social skills, disruptive, academic-related behaviors, other	Elementary

## Declaration of Conflicting Interests

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

## Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article:

Preparation of this article was supported by a grant from the University of Connecticut Foundation. Opinions expressed herein do not necessarily reflect the position of the University of Connecticut, and such endorsements should not be inferred.

## References

- Brown, S., & Rahn-Blakeslee, A. (2009). Training school-based practitioners to collect intervention integrity data. *School Mental Health, 1*, 143–153.
- Burns, M. K., & Ysseldyke, J. E. (2009). Reported prevalence of evidence-based instructional practices in special education. *Journal of Special Education, 43*, 3–11.
- Carr, E. G., Dunlap, G., Horner, R. H., Koegel, R. L., Turnbull, A. P., Sailor, W., . . . Fox, L. (2002). Positive behavior support: Evolution of an applied science. *Journal of Positive Behavior Interventions, 4*, 4–16.
- Cooper, J. O. (1982). Applied behavior analysis in education. *Theory Into Practice, 21*, 114–118.
- Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18*, 23–45.
- Dunlap, G., Carr, E. G., Horner, R. H., Zarcone, J. R., & Schwartz, I. (2008). Positive behavior support and applied behavior analysis: A familial alliance. *Behavior Modification, 32*, 682–698.
- Durlack, J. A., & DuPre, E. P. (2008). Implementation matters: A review on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*, 327–350.
- Fixsen, D. L., Naoom, S. F., Blase', K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231). Retrieved from [cfs.fmhi.usf.edu/resources/publications/NIRN\\_Monograph\\_Full.pdf](http://cfs.fmhi.usf.edu/resources/publications/NIRN_Monograph_Full.pdf)
- Gansle, K. A. (2005). The effectiveness of school-based anger interventions and programs: A meta-analysis. *Journal of School Psychology, 43*, 321–341.
- Gansle, K. A., & McMahon, C. M. (1997). Component integrity of teacher intervention management behavior using a student self-monitoring treatment: An experimental analysis. *Journal of Behavioral Education, 7*, 405–419.
- Gersten, R., Baker, S., & Lloyd, J. W. (2000). Designing high-quality research in special education: Group experimental design. *Journal of Special Education, 34*, 2–18.
- Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children, 71*, 149–164.
- Gresham, F. M. (1989). Assessment of treatment integrity in school consultation and prereferral intervention. *School Psychology Review, 18*, 37–50.
- Gresham, F. M., Gansle, K. A., & Noell, G. H. (1993). Treatment integrity in applied behavior analysis with children. *Journal of Applied Behavior Analysis, 26*, 257–263.
- Gresham, F. M., Gansle, K. A., Noell, G. H., & Cohen, S. (1993). Treatment integrity in applied behavior analysis with children. *Journal of Applied Behavior Analysis, 26*, 257–263.
- Gresham, F. M., MacMillan, D. L., 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*, 198–205.
- Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single subject research to identify evidence-based practice in special education. *Exceptional Children, 71*, 165–179.
- Individuals with Disabilities Education Improvement Act, 20 U.S.C. § 1400 Et. Seq. (2004).
- Johnston, J. M., Foxx, R. M., Jacobson, J. W., Green, G., & Mulick, J. A. (2006). Positive behavior support and applied behavior analysis. *The Behavior Analyst, 29*, 51–74.
- Jones, H. A., Clarke, A. T., & Power, T. J. (2008). Expanding the concept of intervention integrity: A multidimensional model of participant engagement. *In Balance, 23*, 4–5.
- Larsen, T., & Samdal, O. (2007). Implementing Second Step: Balancing fidelity and program adaptation. *Journal of Educational and Psychological Consultation, 17*, 1–29.
- McIntyre, L. L., Gresham, F. M., DiGennaro, F. D., & Reed, D. D. (2007). Treatment integrity of school-based interventions with children in *Journal of Applied Behavior Analysis* studies from 1991 to 2005. *Journal of Applied Behavior Analysis, 40*, 659–672.
- Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review, 11*, 247–266.
- Morgan, D. L., & Morgan, R. K. (2001). Single-participant research design: Bringing science to managed care. *American Psychologist, 56*, 119–127.
- Mortenson, B. P., & Witt, J. C. (1998). The use of weekly performance feedback to increase teacher implementation of a prereferral academic intervention. *School Psychology Review, 27*, 613–627.
- No Child Left Behind, 20 U.S.C. § 16301 Et Seq. (2001).
- Noell, G. H. (2008). Research examining the relationships among consultation process, treatment integrity, and outcomes. In W. P. Erchul & S. M. Sheridan (Eds.), *Handbook of research*

- in school consultation: Empirical foundations for the field (pp. 315–334). Mahwah, NJ: Lawrence Erlbaum.
- Noell, G. H., Witt, J. C., Gilbertson, D. N., Ranier, D. D., & Freeland, J. T. (1997). Increasing teacher intervention implementation in general education settings through consultation and performance feedback. *School Psychology Quarterly, 12*, 77–88.
- Perepletchikova, F., Treat, T., & Kazdin, A. E. (2007). Treatment integrity in psychotherapy research: Analysis of the studies and examination of the associated factors. *Journal of Consulting & Clinical Psychology, 75*, 829–841.
- Peterson, L., Homer, A., & Wonderlich, S. (1982). The integrity of independent variables in behavior analysis. *Journal of Applied Behavior Analysis, 15*, 477–492.
- Power, T. J., Blom-Hoffman, J., Clarke, A. T., Riley-Tillman, T. C., Kellerher, C., & Manz, P. (2005). Reconceptualizing intervention integrity: A partnership-based framework for linking research with practice. *Psychology in the Schools, 42*, 495–507.
- Ringwalt, C. L., Ennet, S., Johnson, R., Rohrbach, L. A., Simons-Rudolph, A., Vincus, A., & Thorne, J. (2003). Factors associated with fidelity to substance use prevention curriculum guides in the nation's middle schools. *Health Education & Behavior, 30*, 375–391.
- Sanetti, L. M. H., Gritter, K. L., & Dobey, L. M. (2011). Treatment integrity of interventions with children in the school psychology literature from 1995 to 2008. *School Psychology Review, 40*, 72–84.
- Sanetti, L. M. H., & Kratochwill, T. R. (2008). Treatment integrity in behavioral consultation: Measurement, promotion, and outcomes. *International Journal of Behavioral Consultation and Therapy, 4*, 95–114.
- Sanetti, L. M. H., & Kratochwill, T. R. (2009a). Toward developing a science of treatment integrity: Introduction to a special series. *School Psychology Review, 38*, 445–459.
- Sanetti, L. M. H., & Kratochwill, T. R. (2009b). Treatment integrity assessment in the schools: An evaluation of the Treatment Integrity Planning Protocol (TIPP). *School Psychology Quarterly, 24*, 24–35.
- Sanetti, L. M. H., & Kratochwill, T. R. (in press). Teachers' self-report of treatment integrity while implementing the good behavior game: Daily versus weekly assessment. *Journal of Educational and Psychological Consultation*.
- Snell, M. E., Chen, L. Y., & Hoover, K. (2006). Teaching augmentative and alternative communication to students with severe disabilities: A review of intervention research 1997–2003. *Research and Practice for Persons with Severe Disabilities, 31*, 203–214.
- Sugai, G., Horner, R. H., Dunlap, G., Hieneman, M., Lewis, T. J., Nelson, C. M., . . . Rued, M. (2000). Applying positive behavior support and functional behavior assessment in schools. *Journal of Positive Behavior Interventions, 2*, 131–143.
- Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. S. (1993). Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology, 61*, 620–630.
- Wheeler, J. L., Baggett, B. A., Fox, J., & Blevins, L. (2006). Treatment integrity: A review of intervention studies conducted with children with autism. *Focus on Autism and Other Developmental Disabilities, 21*, 45–54.
- Witt, J. C., Noell, G. H., LaFleur, L. H., & Mortenson, B. P. (1997). Teacher use of interventions in general education settings: Measurement and analysis of the independent variable. *Journal of Applied Behavior Analysis, 30*, 693–696.

### About the Authors

**Lisa M. Hagermoser Sanetti**, PhD, is an assistant professor in the Neag School of Education and a research scientist with the Center for Behavioral Education and Research at the University of Connecticut. Her primary areas of research interest include treatment integrity assessment and promotion and evidence-based practice in schools.

**Lisa M. Dobey**, MA, is a graduate of the school psychology program at the University of Connecticut. She is currently a school psychologist in the Belmont, Massachusetts, public schools.

**Katie L. Gritter**, MS, is a doctoral candidate in the School Psychology program at the University of Connecticut. Topics of interest include treatment integrity, behavioral consultation, and evidence-based interventions.