

*TREATMENT INTEGRITY OF SCHOOL-BASED INTERVENTIONS
WITH CHILDREN IN THE JOURNAL OF APPLIED
BEHAVIOR ANALYSIS 1991–2005*

LAURA LEE MCINTYRE

SYRACUSE UNIVERSITY

FRANK M. GRESHAM

LOUISIANA STATE UNIVERSITY

AND

FLORENCE D. DIGENARO AND DEREK D. REED

SYRACUSE UNIVERSITY

We reviewed all school-based experimental studies with individuals 0 to 18 years published in the *Journal of Applied Behavior Analysis (JABA)* between 1991 and 2005. A total of 142 articles (152 studies) that met review criteria were included. Nearly all (95%) of these experiments provided an operational definition of the independent variable, but only 30% of the studies provided treatment integrity data. Nearly half of studies (45%) were judged to be at high risk for treatment inaccuracies. Treatment integrity data were more likely to be included in studies that used teachers, multiple treatment agents, or both. Although there was a substantial increase in reporting operational definitions of independent variables, results suggest that there was only a modest improvement in reported integrity over the past 30 years of *JABA* studies. Recommendations for research and practice are discussed.

DESCRIPTORS: treatment integrity, child studies, school interventions, applied behavior analysis

The field of applied behavior analysis has always rested on the fundamental principle that the empirical demonstration of measurable changes in behavior must be related to systematic and controlled manipulations in the environment. That is, the observed changes in the dependent variable (behavior) must be attributed to changes in the independent variable (some environmental event). Without this empirical demonstration, a true science of human behavior is an impossibility (Skinner, 1953). Without objective and documented

specification of an independent variable as well as accurate independent variable application, definitive conclusions regarding the relation between an independent variable and a dependent variable are compromised. The best way to ensure accurate application of the independent variable is to measure the extent to which treatment is implemented as intended.

Documentation of independent variable implementation has been discussed in the literature under the rubric of *treatment fidelity* (Moncher & Prinz, 1991) or *treatment integrity* (Gresham, Gansle, & Noell, 1993; Gresham, Gansle, Noell, & Cohen, 1993; Peterson, Homer, & Wonderlich, 1982; Yeaton & Sechrest, 1981). Treatment integrity refers to the degree to which treatments are implemented as planned, designed, or intended and is concerned with the accuracy and consistency with which interventions are implemented

We thank Heidi Olson-Tinker, Lisa Dolstra, Veronica McLaughlin, and Mai Van for assistance with the initial preparation of this article. We also are grateful to Michael J. Vance for assistance with data collection.

Address correspondence to Laura Lee McIntyre, Department of Psychology, 430 Huntington Hall, Syracuse University, Syracuse, New York 13244 (e-mail: llmcinty@syr.edu).

doi: 10.1901/jaba.2007.659–672

(Peterson et al.). Therefore, treatment integrity is necessary but insufficient for demonstrating a functional relation between intervention procedures and behavior change (Gresham, 1989).

A number of studies have been published in recent years that have examined variables associated with adequate treatment integrity (DiGennaro, Martens, & Kleinmann, 2007; DiGennaro, Martens, & McIntyre, 2005; Mortenson & Witt, 1998; Noell, Witt, Gilbertson, Ranier, & Freeland, 1997; Noell et al., 2000; Sterling-Turner, Watson, Wildmon, Watkins, & Little, 2001; Witt, Noell, LaFleur, & Mortenson, 1997). Most of these studies have focused on schools as the primary setting for investigation. Investigating the degree to which interventions are carried out with integrity in schools is valuable for several reasons. First, research suggests that teachers fail to implement interventions with accuracy despite receiving high levels of initial training (e.g., DiGennaro et al., 2005; Noell et al., 2000). This is a waste of time and resources for both teachers and consultants if, after training, the interventions are not implemented as intended. Second, findings also suggest that student problem behaviors are negatively correlated with treatment accuracy, such that low levels of problem behavior are associated with high levels of treatment integrity (DiGennaro et al., 2005, 2007; Wilder, Atwell, & Wine, 2006). Thus, a teacher's failure to implement recommended interventions may result in poor outcomes for students, in that behaviors will not improve in the desired direction. Third, the extent to which teachers implement plans with accuracy influences a behavior analyst's ability to effectively conduct formative evaluations. Specifically, a behavior analyst will be unable to determine if a student's resistance to treatment is a result of an ineffective intervention or a lack of intervention implementation (Moncher & Prinz, 1991) because the treatment's effect size is positively correlated with internal validity

(Smith, Glass, & Miller, 1980). Having this knowledge would focus a behavior analyst's efforts on problem solving with teachers and students (i.e., change the intervention or directly work to improve teachers' implementation of the current plan). Finally, recent legislation, such as the No Child Left Behind Act (U.S. Department of Education, 2002) and Individuals with Disabilities Improvement Act (2004), necessitates that school-based practitioners and teachers be accountable for their practices. As a result, there has been a recent push for evidence-based practices in academic settings as well as demonstrations of accurate plan implementation over time.

How common is the measurement of treatment integrity in the behavior analysis literature? Several reviews of the literature suggest that the measurement of treatment integrity is uncommon (Gresham, Gansle, & Noell, 1993; Peterson et al., 1982; Wheeler, Baggett, Fox, & Blevins, 2006). Peterson et al. reviewed 539 studies published in the *Journal of Applied Behavior Analysis (JABA)* between 1968 and 1980; they found that only 20% of the 539 studies reported data on treatment integrity, and over 16% of these studies did not provide an operational definition of the independent variable. There were no trends suggesting an improvement in treatment integrity over time. Gresham, Gansle, and Noell provided an update of Peterson et al.'s review by examining 158 studies published in *JABA* between 1980 and 1990 that were child studies (<19 years of age). Of these 158 studies, only 32% provided an operational definition of the independent variable and only 16% (25 studies) systematically measured and reported levels of treatment integrity.

Wheeler et al. (2006) focused on intervention studies of children with autism published between 1993 and 2003. Of the 60 studies included in the review, more than half (60%) were published in *JABA*, with the remaining studies ($n = 26$) drawn from eight other

journals (e.g., *Research in Developmental Disabilities*, *Journal of Autism and Developmental Disorders*). The results of Wheeler et al.'s review were consistent with previous studies. Of these 60 studies, only 18% ($n = 11$) reported data on treatment integrity. On the other hand, nearly all (92%) included operational definitions of independent variables. Closer analysis of the results of Wheeler et al.'s review provides some insight into the treatment-integrity reporting trends for child-based autism treatment studies. Although most of the included studies were published in *JABA*, only 14% ($n = 5$) included treatment integrity data. This figure is lower than what others have reported (e.g., Gresham, Gansle, & Noell, 1993) for *JABA* studies. In contrast, studies published in *Research in Developmental Disabilities*, *Focus on Autism and Other Developmental Disabilities*, *Journal of Autism and Developmental Disorders*, and the *Journal of Positive Behavioral Interventions* included treatment integrity data in 25% to 33% of studies. Studies that met inclusionary criteria published in *Education and Treatment of Children* and the *Journal of Early Intervention* reported treatment integrity 50% and 100% of the time, respectively. In contrast, the three studies published in *Education and Training in Mental Retardation and Developmental Disabilities* and the *Journal of Developmental and Physical Disabilities* did not report treatment integrity data. Although these findings are limited due to the scope of Wheeler et al.'s review criteria, they are helpful in placing treatment integrity reporting in *JABA* in context.

Based on the foregoing reviews, it is clear that the majority of treatment outcome studies published in *JABA* and other behavioral journals either did not measure or did not report levels of treatment integrity. As can be derived from the above discussion on the importance of treatment integrity, the failure to gather data on the integrity of independent variables may compromise the precision and

rigor of our experimental procedures (Baer, Wolf, & Risley, 1968; Johnston & Pennyacker, 1993; Kazdin, 1973). The basic concern is that when data are not collected regarding the status of the independent variable, researchers and practitioners alike cannot objectively conclude that the independent variable was implemented as planned or intended (Kennedy, 2005; Moncher & Prinz, 1991). This problem may be especially problematic in practice settings (Wilder et al., 2006), such as interventions that are implemented in schools.

The current article updates and extends the findings of the Peterson et al. (1982) and Gresham, Gansle, and Noell (1993) reviews by another 15 years. All school-based interventions with children (<19 years old) published in *JABA* between 1991 and 2005 were reviewed for possible inclusion. The clinical relevance of investigating treatment integrity combined with the importance of demonstrating that the independent variable was accurately applied in school-based intervention research serves as the basis of this study.

METHOD

Criteria for Review

A total of 995 articles (excluding book reviews and remembrances) were reviewed to determine possible inclusion. Five features of each study were considered. First, the study had to be experimental, in that the effects of intervention on behavior were examined (i.e., the study had to manipulate some aspect of the environment to create changes in a dependent variable). Because we were evaluating school-based intervention studies, articles that were assessment only (e.g., functional analysis, preference assessment) were excluded. If a study contained an initial functional analysis followed by an intervention, the intervention experiment was included. Second, participants had to be younger than 19 years old, an inclusionary criterion previously employed by Gresham, Gansle, and Noell (1993). Third, studies without a clear baseline

or control condition were excluded from further review. Studies that were not true experimental designs (e.g., AB designs) were excluded. Fourth, all studies had to be conducted in school settings; however, *school* was liberally defined to include a continuum of school placements, including residential programs. Inpatient hospital units (e.g., Neurobehavioral Unit at the Kennedy Krieger Institute) and outpatient clinics were excluded. Fifth, brief reports of three or fewer pages in length were excluded, as outlined by Peterson *et al.* (1982). Because articles of three or fewer pages typically do not provide sufficient methodological detail (e.g., lengthy descriptions of independent variables or integrity monitoring), we chose to exclude these studies so we would not artificially underestimate independent variable operational definition and integrity reporting. Thus, a total of 142 articles met these inclusionary criteria over the 15-year period. Because some of the articles contained multiple experiments, a total of 152 studies met inclusionary criteria for this review. (A full list of articles meeting inclusionary criteria is available from the first author.)

Coding

This review focused on the operational definition of the independent variables and the extent to which these variables were described, monitored, and measured. Following the procedural guidelines set forth by Peterson *et al.* (1982), the risk for treatment inaccuracies was also investigated. In addition, we were interested in assessing whether treatment integrity reporting trends varied by publication year and by whom the intervention was implemented (treatment agent; e.g., teacher, researcher, etc.). Coding schemes for each of these variables are described below.

Operational definition of the independent variable. Each study was coded “yes,” “no,” or “footnote” in answer to the question: Is the independent variable (treatment) operationally defined? To answer this question, each rater was

given the following criterion: “If you could replicate this treatment with the information provided, the intervention is considered operationally defined.” This criterion was proposed by Baer *et al.* (1968) and later used by Gresham, Gansle, and Noell (1993) in their review. Those studies that referred to more extensive sources (e.g., book chapters, manuals, or technical reports) were coded as “footnote” (i.e., contained directions to contact the author or see published details elsewhere).

Monitoring treatment integrity. Studies were coded according to their inclusion of treatment integrity data. Studies that systematically monitored and reported treatment integrity on at least one independent variable were coded “yes.” Specifically, this included studies that (a) specified a method of measurement (observer present, videotaping of sessions, component checklist) and (b) reported data as percentage of implementation (i.e., percentage of implemented steps in the intervention). Studies that monitored treatment integrity but failed to report data were coded as “monitored.” For example, “treatment integrity was assessed to ensure the fidelity of this intervention” was coded as “monitored.” Likewise, studies that mentioned statements such as “deviations from intervention protocol were not observed” were also coded as “monitored” (no method of measurement was described). The key difference between “yes” and “monitored” categories was the provision of percentage data regarding implementation and a specified data-collection method. Studies that made no mention of treatment integrity were coded “no.” We chose to replicate Gresham, Gansle, and Noell’s (1993) treatment integrity coding because, unlike Peterson *et al.* (1982), this method allowed differentiating the categories of “yes” and “monitored.”

Risk for treatment inaccuracies. Treatments were coded as either no, low, or high risk for treatment inaccuracies based on the guidelines set forth by Peterson *et al.* (1982). Treatments

were coded as “no risk” if the implementation of the treatment was reported as monitored or measured (i.e., monitoring of treatment integrity was coded as either “yes” or “monitored”). Treatments were coded as “low risk” if the treatment was not reported to be monitored or measured but was judged to be at low risk for inaccuracies. Low-risk treatments included treatments that were (a) mechanically defined (e.g., computer mediated), (b) permanent products (e.g., posting of classroom rules), (c) continuously applied (e.g., noncontingent access to preferred items or activities), or (d) single components (e.g., escape contingent on work completion). Treatments were coded as “high risk” if the treatment was not reported to be monitored or measured but was necessary. According to Peterson et al. (1982) treatments in the high-risk category were those in which “the administration of the independent variable was not exempted by any of the cases cited in category B [low risk], and the potential for error was judged to be high” (p. 485). Operationally defined, these included person-implemented interventions that included multiple behavioral components (e.g., contingent reinforcement with response cost).

Publication year. The publication year of the article was recorded (i.e., 1991 to 2005).

Treatment agent. The individuals who implemented the intervention were classified into one of the following mutually exclusive categories: (a) teacher, (b) professional (nonteacher), (c) paraprofessional, (d) parent or sibling, (e) researcher or research assistant, (f) peer tutors, (g) self, (h) multiple, (i) other, or (j) not specified. Examples of the teacher category included early childhood educators, general education classroom teachers, or discrete-trial instructors. The professional category included other nonteacher professionals (e.g., school psychologists, speech–language pathologists). Paraprofessionals included support staff such as classroom aides, teaching assistants (nonteachers), or playground or lunchroom moni-

tors. Researchers and research assistants were individuals who collected data for the purpose of the published study and were not involved in other service delivery roles (e.g., classroom teacher). Peer tutors were other children, typically in the target child’s classroom, who were not the focus of the intervention. “Self” was recorded if the intervention was self-administered or self-mediated (e.g., self-monitoring interventions). “Multiple” was coded if more than one category of treatment agent was used. If the treatment agent described in the study did not fit in any of the aforementioned categories, “other” was coded. There were a small handful of studies that did not specify the treatment agent. In these cases “not specified” was coded.

Rater Training and Interobserver Agreement

A PhD-level behavior analyst (faculty member) and four doctoral students with advanced training in behavior analysis served as raters, with each rater coding 20% of the studies. Prior to coding, all raters received four 2-hr training sessions to discuss assigned practice articles (i.e., *JABA* articles published prior to 1991) and to revise ambiguous codes. During these training sessions, all raters reached 100% agreement (via consensus) on whether an assigned article met inclusionary criteria. Five articles were assigned per training session, yielding a total of 20 training articles used prior to conducting independent coding. In addition, a random sample of 20% of studies meeting inclusionary criteria was selected for interobserver agreement coding. Studies were coded on five categories: (a) operational definition of the independent variable (three categories), (b) integrity assessment (three categories), (c) risk for treatment inaccuracies (three categories), (d) publication year (15 categories), and (e) treatment agent (10 categories). Percentage agreement was calculated by dividing the number of agreements by the number of agreements plus disagreements and multiplying by 100%. Percentage agreement

averaged 93% across the five codes (98% operational definition of the independent variable; 87% integrity assessment; 88% risk for treatment inaccuracies; 100% publication year; 92% treatment agent).

RESULTS

The majority of studies ($n = 144$; 95%) provided operational definitions of treatments, with an additional five studies (3%) reporting references or contact information to allow readers to gather more information about the interventions (e.g., treatment manuals, previously published studies, etc.). The remaining three studies (2%) did not provide operational definitions adequate for replication purposes or cite other sources for more information.

Approximately one third ($n = 46$; 30%) of the studies provided treatment integrity data in the form of percentage of implementation. Studies that reported these data showed a high percentage of integrity ($M = 93\%$; $SD = 9.93$). The majority of studies that reported integrity data ($n = 36$; 78%) reported procedural fidelity of 90% or greater. Thirteen studies (8%) mentioned that treatment integrity was monitored but did not provide data for percentage of steps accurately implemented. Over 60% of the studies ($n = 93$) did not report treatment integrity data nor did they report monitoring the implementation of their interventions.

Approximately 39% of studies ($n = 59$) were considered to be at no risk for treatment inaccuracies, in that the authors reported treatment integrity data or that treatment integrity was monitored. Just under half of the included studies ($n = 69$; 45%) were considered to be at high risk for treatment inaccuracies in that information on the implementation of treatments or the assessment of independent variables was not included but should have been (Peterson *et al.*, 1982). The remaining 16% of studies ($n = 24$) did not include information

on treatment integrity but were judged to be at low risk for treatment inaccuracies.

Reporting treatment integrity data did not appear to differ consistently by publication year; however, there was ample variability across the 15-year period. Figure 1 depicts the percentage of studies that included treatment integrity data by publication year. On average, treatment integrity data were included in one third of the included studies ($M = 34\%$; $SD = 19.23$). The publication years 1996, 1998, 1999, and 2005 included relatively more studies that reported treatment integrity data (range, 50% to 67%) than the remaining 12 years. Figure 2 shows treatment integrity data from 1968 to 2005 based on Peterson *et al.*'s (1982) review; Gresham, Gansle, and Noell's (1993) review; and the present review. These data are based on 834 studies published in *JABA* from 1968 to 2005. Of these 834 studies, 179 (21%) reported treatment integrity data (range, 0% to 67%).

We were interested in exploring whether studies that used particular treatment agents (e.g., teachers, researchers) reported treatment integrity data more frequently. As shown in Table 1, there were a variety of reported treatment agents in the included studies. The most common were researchers ($n = 52$), teachers ($n = 38$), multiple ($n = 19$), and professionals ($n = 15$). Although only seven studies used peer tutors as treatment agents, 57% ($n = 4$) reported treatment integrity data. Of the 19 studies that used multiple treatment agents, nearly a third ($n = 6$; 32%) included treatment integrity data. Likewise, for the 38 studies that used teachers as treatment agents, 37% ($n = 14$) reported treatment integrity data. Studies that used professionals, parents or siblings, researchers, or self-administered treatments had lower reporting of treatment integrity data (range, 0% to 25%).

DISCUSSION

The present review of school-based interventions with children published in *JABA* demon-

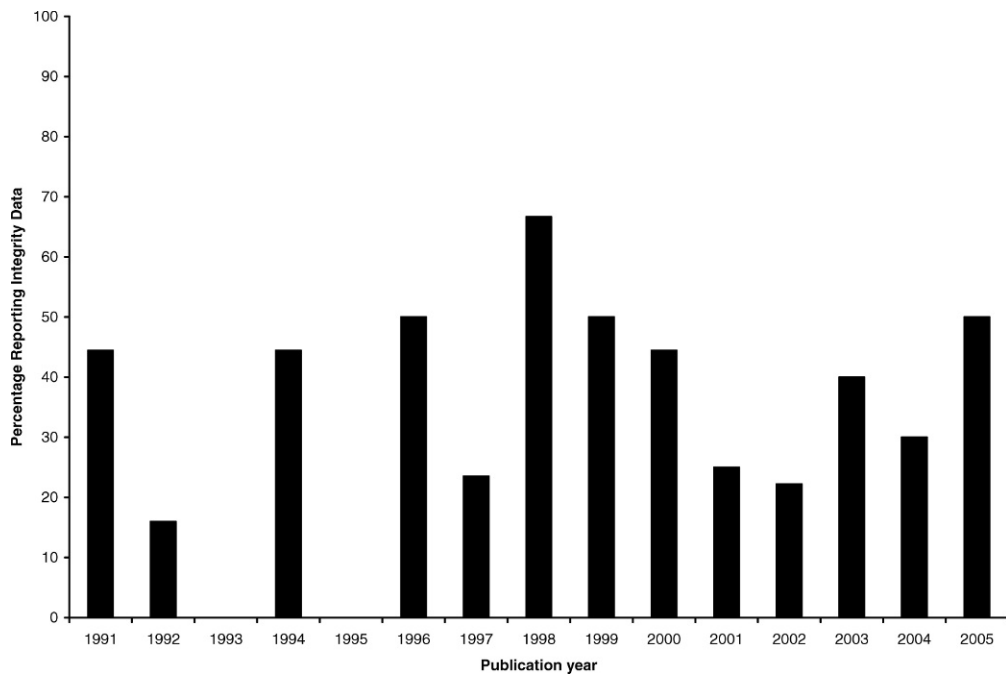


Figure 1. Percentage of *JABA* school-based studies reporting treatment integrity data by year (1991 to 2005).

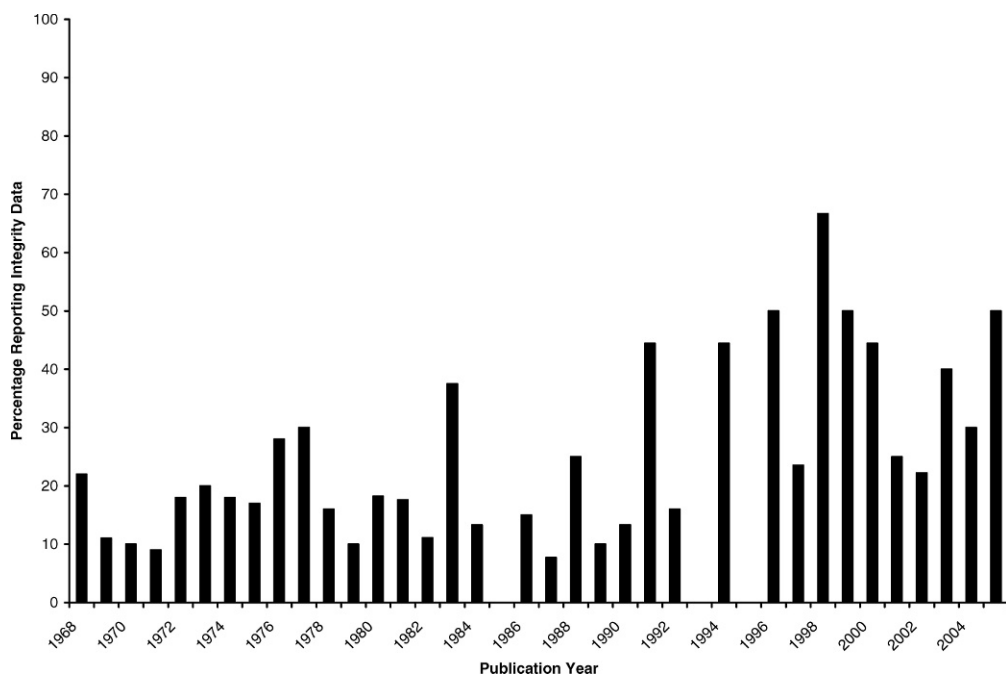


Figure 2. Percentage of *JABA* studies reviewed by Peterson et al. (1982); Gresham, Gansle, and Noell (1993); and the current review reporting treatment integrity data by year (1968 to 2005).

Table 1
Treatment Integrity Monitoring by Treatment Agent

Treatment agent	Yes + data <i>n</i> (%)	Monitored <i>n</i> (%)	No <i>n</i> (%)	Total
Teacher	14 (37)	4 (10)	20 (53)	38
Professional (nonteacher)	3 (20)	1 (7)	11 (73)	15
Paraprofessional	2 (33)	0 (0)	4 (67)	6
Parent or sibling	0 (0)	1 (50)	1 (50)	2
Researcher	13 (25)	4 (8)	35 (67)	52
Peer tutors	4 (57)	1 (14)	2 (29)	7
Multiple	6 (32)	2 (10)	11 (58)	19
Does not specify	3 (33)	0 (0)	6 (67)	9
Self	0 (0)	0 (0)	2 (100)	2
Other	1 (50)	0 (0)	1 (50)	2
Total	46 (30)	13 (8)	93 (61)	152

strates that reporting rates of treatment integrity data have been remarkably stable (and low) over the past 15 years. Approximately one third (30%) of studies that met our inclusionary criteria reported treatment integrity data. This figure is slightly higher than the Peterson *et al.* (1982) and Gresham, Gansle, and Noell (1993) reviews of this literature that showed 20% and 16% integrity, respectively. Although somewhat different inclusionary criteria were used in the two earlier reviews, treatment integrity reporting has been remarkably stable over the past 37 years (1968 to 2005) (Figure 2). Of interest is the large increase in treatment integrity reporting that was seen from 1993 to 1994. Although attributions about the cause of this increase cannot be made, this spike occurred the year following Gresham, Gansle, and Noell's review. Gresham, Gansle, and Noell reported a similar increase from 1982 to 1983 (the year following Peterson *et al.*'s review). It is plausible that papers of this nature may increase *JABA* authors' and editors' awareness of the need to include treatment integrity data. Alternatively, there may be other variables that contributed to the spike in treatment integrity reporting, such as the sharp increase seen from 1997 to 1998. To the best of our knowledge, however, editorial guidelines for preparing manuscripts to be submitted to *JABA* did not change during this time.

Reporting of treatment integrity data has been relatively stable and low over the years.

Reasons for low rates of treatment integrity reporting are not entirely clear; however, low reporting may be a function of the editorial process (*i.e.*, space limitations in journals warrant cutting out treatment integrity data) or may be due to logistics (*e.g.*, lack of skills in treatment integrity assessment, lack of resources). There may also be a publication bias favoring the reporting of treatment integrity data when integrity is high. In addition, it is plausible that researchers do not view treatment integrity data collection as important, especially if interventions produce the desired effects. We argue that without collecting integrity data, it becomes difficult to make conclusions regarding intervention results.

Having access to treatment integrity data can help behavior analysts to make decisions about treatments in school-based settings. If, for example, an intervention is being implemented accurately yet does not produce the desired effects, the behavior analyst will likely modify the treatment. If the intervention is being implemented inaccurately and does not produce the desired effects, the behavior analyst will likely institute additional training or programmed consequences to increase implementation accuracy. On the other hand, if the intervention is not being implemented with integrity yet still produces the desired effects, the behavior analyst will likely change the treatment protocol to reflect the modified intervention. Finally, if the intervention is being

implemented with integrity and the desired treatment outcomes are produced, a causal relation between independent variable manipulations and changes in the dependent variable can be inferred. Thus, we argue that including regular treatment-integrity assessments is necessary but insufficient for making treatment-related decisions (Gresham, 1989).

In contrast to the rates of treatment integrity reporting, reporting of operationally defined independent variables has dramatically increased, with nearly all (95%) studies including detailed descriptions of the interventions. This figure is consistent with a recent review of interventions for children with autism (Wheeler et al., 2006) but is a much improved rate over the 34% reported by Gresham, Gansle, and Noell (1993). Including operational definitions of independent variables contributes to the replicability of our science of behavioral interventions (Bellg et al., 2004).

Although treatment integrity measures are important for virtually all experimental studies, including assessment studies and interventions conducted in other settings, we chose to sample interventions with children in school settings. This population and setting were selected because it is the focus of our own research; however, this may be of interest to other researchers in its own right. Furthermore, interventions carried out in school settings, in which treatment agents are less likely to be researchers with significant training in experimental methods, may be at greatest risk for inaccurate implementation of interventions. When treatment integrity is not systematically assessed and reported, there is little basis for judgments about how closely an implemented intervention approximates an intended intervention. Because the current review focused on school settings, the extent to which these findings generalize to published studies conducted with other populations is unknown.

Our findings suggest that when school-based interventions are carried out by teachers,

paraprofessionals, peers, or multiple treatment agents, authors are more likely to report treatment integrity data. It may be the case that these treatment agents were judged to be at high risk for procedural inaccuracies and the authors therefore went to great lengths to ensure that these agents implemented the treatments as planned. Although definitive conclusions cannot be made based on these descriptive data, it appears that the treatment agent used in school-based studies may influence the likelihood of reporting treatment integrity data. What is unknown, however, is how many other authors collected treatment integrity data but failed to report it in their published articles. Failure to include a brief statement on the extent to which treatments were implemented as planned may be especially problematic for interventions judged to be at high risk for treatment inaccuracies (Peterson et al., 1982). If treatment integrity data are not regularly included, inferences based on the study results may be significantly limited (Kennedy, 2005). Thus, we recommend that if treatment integrity data are collected or if intervention implementation is monitored, this information should be included in published studies.

Although we have seen marked improvement in descriptions of independent variables, publications in *JABA* continue to focus on clear specifications of the dependent variables and do not include measurements of the independent variables. Indeed, a "curious double standard" so aptly recognized by Peterson et al. (1982) still remains. This observation continues to be recognized by various task forces and organizations within the fields of education, psychology, and mental health. For example, the Task Force on Evidence-Based Practice in Special Education of the Council for Exceptional Children stated that the integrity of intervention implementation is critical in single-case designs because the independent variable is implemented continuously over time (Horner et al., 2005).

Similarly, other task forces within the American Psychological Association on evidence-based treatments such as Divisions 16 (school psychology), 53 (clinical child/adolescent), and 54 (pediatric) have called for the assessment and monitoring of treatment integrity. Furthermore, researchers who submit single-case experimental design grant applications to the U.S. Department of Education's Institute of Education Sciences (IES) now must describe "how treatment fidelity will be measured, frequency of assessments, and what degree of variation in treatment fidelity will be accepted over the course of the study" (IES, 2006, p. 50). These recommendations have also been made by the National Institutes of Health (NIH). Specifically the NIH Behavior Change Consortium recommends that treatments be monitored and reported and that treatment agents be trained and supervised in the delivery of treatments (Bellg *et al.*, 2004). Monitoring and reporting treatment fidelity are especially important in clinical treatments that are considered to be at high risk for treatment inaccuracies or complex in other ways (e.g., multisite). Furthermore, the special NIH report on treatment fidelity in research specifies that "it is particularly important that funding agencies, reviewers, and journal editors who publish behavioral change research consider treatment fidelity issues" (Bellg *et al.*, p. 451).

With the increased attention paid to issues of accurate treatment implementation and reporting of treatment integrity, both within the field of behavior analysis and in other fields (e.g., psychology, behavioral medicine), it may be particularly important for *JABA* authors and readers to consider some additional ways to strengthen the influence of behavior analysis in the larger scientific community. Outlined are several recommendations for treatment integrity research and recommendations for practice.

Recommendations for Research and Practice

Although accurate implementation of the independent variable is assumed to be func-

tionally related to desired changes in the dependent variable, there has been relatively little research that demonstrates this relation (Wilder *et al.*, 2006). Furthermore, it may be the case that high levels of treatment integrity are necessary for some interventions but may not be necessary for others. Only a handful of behavior-analytic studies have addressed this issue, unfortunately coming to somewhat different conclusions. For example, Wilder *et al.* systematically manipulated different levels of treatment integrity of a three-step prompting procedure on children's compliance. Wilder *et al.* concluded that the level of treatment accuracy had a large impact on children's compliance. Northup, Fisher, Kahng, Harrel, and Kurtz (1997), on the other hand, found very little difference between time-out treatments implemented at 100% accuracy and those implemented at 50% accuracy. Vollmer, Roane, Ringdahl, and Marcus (1999) evaluated the effects of differential reinforcement of alternative behavior and found that degree of treatment accuracy did affect treatment outcomes. Because of the small number of studies that have addressed the varying effects of treatment integrity on behavior change, we recommend that additional studies include treatment integrity variation as an independent variable and consider that various treatments may actually require different levels of treatment integrity to produce desired changes in the dependent variable. Regular documentation of treatment integrity may help to improve our knowledge base in this regard.

An additional area of research for behavior-analytic studies may be to separate the components of treatment packages to identify the variables that are functionally responsible for producing behavior change. It is plausible that some components of a treatment package may be excluded, whereas others may be necessary to produce treatment effects. Thus, a treatment that is implemented with 80% accuracy but is missing a key ingredient may produce poorer

outcomes than a treatment that is implemented at 70% accuracy but includes the components that are functionally responsible for changes in the dependent variable.

Behavioral interventions, especially those implemented in applied settings (e.g., schools), may be at high risk for treatment inaccuracies due to the setting, treatment agent, complexity of the protocol, and demands placed on teachers' time and resources. Interventions that include programmed consequences for teachers (or other treatment agents) contingent on accuracy of treatment implementation may produce higher levels of treatment integrity. For example, Noell et al. (1997) found that a performance feedback package increased teachers' accurate implementation of treatments. Furthermore, DiGennaro et al. (2007) found that programmed consequences including performance feedback and negative reinforcement (escape from a meeting with the behavior analyst) produced higher levels of treatment integrity than a single programmed consequence or no programmed consequence. Additional research using programmed consequences for treatment agents in applied settings may help to elucidate conditions in which treatments are more or less likely to be implemented with accuracy in applied settings.

Data to support Peterson et al.'s (1982) no risk, low risk, and high risk for treatment inaccuracies may help the field to flesh out the construct of risk for treatment inaccuracies. Although it is assumed that some treatments may be at higher risk for inaccuracies, treatment integrity data have not been reported for studies with more or less complex interventions. It is recommended that treatment integrity be collected on a number of interventions to determine whether complexity of treatments or other features of the treatment (e.g., acceptability; Sterling-Turner & Watson, 2002) are related to treatment integrity. Furthermore, although Peterson et al.'s criteria have served as an important heuristic for the field of

behavior analysis, it may be appropriate to update our thinking with respect to what constitutes risk for treatment inaccuracies. Peterson et al.'s criteria were based on Kelly's (1977) definition of risk that he developed based on reviewing reliability reporting trends in *JABA*. This conceptualization of risk for independent variable inaccuracies does not include treatment agent (e.g., certified behavior analyst vs. novice therapist), years of experience, setting, or other variables that may be germane to our consideration of risk. In addition, Peterson et al. considered monitoring integrity and reporting treatment integrity data to be equivalent with respect to risk for treatment inaccuracies. We posit that monitoring interventions may be less informative for both research and practice than the provision of integrity data.

In terms of practical recommendations, we suggest that treatment integrity plans be specified at the outset of studies (Hellg et al., 2004). That is, researchers should specify when treatment integrity will be assessed and how the assessment will occur. Clearly specifying intervention steps in a treatment protocol may help the implementation and assessment of the intervention. Given that a number of school-based intervention studies published in *JABA* are considered to be high risk for treatment inaccuracies, it is probable that treatments implemented in practice (and not published) may be at greater risk for treatment inaccuracies.

Other practical recommendations include providing initial training for treatment agents at the study onset and training to a criterion rather than training for a prespecified period of time (Bellg et al., 2004). Training should be viewed as an ongoing activity due to factors such as therapist drift or failure to implement the treatment as outlined (e.g., DiGennaro et al., 2005; Noell et al., 2000). Spot checks of treatment integrity could be performed with the assistance of well-developed procedural checklists and protocols. We have found that providing intervention protocols (see the exam-

ple in Appendix A) and using simple procedural checklists (see the example in Appendix B) can be a helpful way to train teachers to implement interventions and collect integrity data that reflects the percentage of treatment steps implemented accurately. Depending on the intervention, protocols could provide a step-by-step guide to treatment implementation or a list of components that must occur (or may not occur) during treatment. For example, it may be important to specify when reinforcement should occur (e.g., contingent on task completion) as well as when reinforcement should not occur (e.g., in the presence of target problem behavior).

Lastly, we recommend that a small sample of treatment integrity assessments be collected on all interventions considered to be at high risk for treatment inaccuracies. Although the demands placed on the time of behavior analysts, teachers, and support staff are great, we have never skimped on conducting assessments of the reliability of dependent variables (e.g., interobserver agreement checks). If, for example, interobserver agreement data are collected on 35% of all observations, researchers and practitioners alike could decide that the number of agreement data-collection observations could be reduced (e.g., to 20%) and 15% of observations could be used for treatment integrity assessments. Because research conducted in applied settings may be at particularly high risk for treatment inaccuracies, including treatment integrity spot checks may be especially important. We believe that it is important to have some methods in place to ensure that treatments are implemented as planned. Furthermore, regularly including such data in studies published in *JABA* may help the field of applied behavior analysis to have a better understanding of the concepts and strategies applied researchers can use to strengthen our science.

REFERENCES

- Baer, D., Wolf, M., & Risley, T. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 1, 91–97.
- Bellg, A. J., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D. S., Ory, M., et al. (2004). Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychology*, 23, 443–451.
- DiGennaro, F. D., Martens, B. K., & Kleinmann, A. E. (2007). A comparison of performance feedback procedures on teachers' treatment implementation integrity and students' inappropriate behavior in special education classrooms. *Journal of Applied Behavior Analysis*, 40, 447–461.
- DiGennaro, F. D., Martens, B. K., & McIntyre, L. L. (2005). Increasing treatment integrity through negative reinforcement: Effects on teacher and student behavior. *School Psychology Review*, 34, 220–231.
- 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., & 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 of school-based behavioral intervention studies: 1980–1990. *School Psychology Review*, 22, 254–272.
- 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. (2004). Public Law 108–446. Retrieved December 30, 2006, from <http://www.ed.gov/policy/speced/guid/idea/idea2004.html>
- Institute for Education Sciences. (2006). Special education research grants 2007 request for applications. Retrieved August 29, 2006, from <http://ies.ed.gov/ncser/pdf/2007324.pdf>
- Johnston, J., & Pennypacker, H. (1993). *Strategies and tactics of behavioral research* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Kazdin, A. E. (1973). Methodological and assessment considerations in evaluating reinforcement programs in applied settings. *Journal of Applied Behavior Analysis*, 6, 517–531.
- Kelly, M. B. (1977). A review of the observational data-collection and reliability procedures reported in the *Journal of Applied Behavior Analysis*. *Journal of Applied Behavior Analysis*, 10, 97–101.
- Kennedy, C. H. (2005). *Single-case designs for educational research*. Boston: Allyn & Bacon.
- Moncher, F. J., & Prinz, F. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review*, 11, 247–266.
- 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.

- 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.
- Noell, G. H., Witt, J. C., LaFleur, L. H., Mortenson, B. P., Ranier, D. D., & LeVelle, J. (2000). Increasing intervention implementation in general education following consultation: A comparison of two follow-up strategies. *Journal of Applied Behavior Analysis, 33*, 271-284.
- Northup, J., Fisher, W., Kahng, S., Harrel, B., & Kurtz, P. (1997). An assessment of the necessary strength of behavioral treatments for severe behavior problems. *Journal of Developmental and Physical Disabilities, 9*, 1-16.
- Peterson, L., Homer, A., & Wonderlich, S. (1982). The integrity of independent variables in behavior analysis. *Journal of Applied Behavior Analysis, 15*, 477-492.
- Skinner, B. F. (1953). *Science and human behavior*. New York: The Free Press.
- Smith, M. L., Glass, G. V., & Miller, T. I. (1980). *The benefits of psychotherapy*. Baltimore: Johns Hopkins University Press.
- Sterling-Turner, H. E., & Watson, T. S. (2002). An analog investigation of the relationship between treatment acceptability and treatment integrity. *Journal of Behavioral Education, 11*, 39-50.
- Sterling-Turner, H. E., Watson, T. S., Wildmon, M., Watkins, C., & Little, E. (2001). Investigating the relationship between training type and treatment integrity. *School Psychology Quarterly, 16*, 56-67.
- U.S. Department of Education. (2002). No Child Left Behind Act of 2001. Public Law 1-7-110. Retrieved December 30, 2006, from <http://www.ed.gov/legislation/ESEA02/>
- Vollmer, T. R., Roane, H. S., Ringdahl, J. E., & Marcus, B. A. (1999). Evaluating treatment challenges with differential reinforcement of alternative behavior. *Journal of Applied Behavior Analysis, 32*, 9-23.
- Wheeler, J. J., Baggett, B. A., Fox, J., & Blevins, L. (2006). Treatment integrity: A review of intervention studies conducted with children with autism. *Focus on Autism and Other Developmental Disabilities, 21*, 45-54.
- Wilder, D. A., Atwell, J., & Wine, B. (2006). The effects of varying levels of treatment integrity on child compliance during treatment with a three-step prompting procedure. *Journal of Applied Behavior Analysis, 39*, 369-373.
- Witt, J. C., Noell, G. H., La Fleur, L. H., & Mortenson, B. P. (1997). Teacher use of interventions in general education: Measurement and analysis of the independent variable. *Journal of Applied Behavior Analysis, 30*, 693-696.
- Yeaton, W. H., & Sechrest, L. (1981). Critical dimensions in the choice and maintenance of successful treatments: Strength, integrity, and effectiveness. *Journal of Consulting and Clinical Psychology, 49*, 156-167.

Received September 6, 2006
Final acceptance April 10, 2007
Action Editor, Louis Hagopian

APPENDIX A

School-Based Intervention Protocol for Student Jamie

1. Jamie will use the reinforcement system at all times throughout the school day.
2. Jamie's behavior plan is specific and targets the following:
 - a. Follows directions: complies with teacher's instructions within 10 s without redirection.
 - b. Completes work: eyes and head oriented to academic task.
 - c. Body still: appropriate motor movement in the context of classroom instruction
3. Jamie will select a reinforcer from a prepared list of items or activities. The teacher will write Jamie's selection on the bottom of the reward slip.
4. Jamie will receive three checks contingent on successfully following directions, completing work, and keeping his hands and feet to himself (one check for each behavior) within a 20-min period.
5. Immediately after receiving the final check, Jamie is allowed to earn the selected reinforcer.
6. The teacher should then cycle back through the previous steps repeatedly through the day.

APPENDIX B

Treatment Integrity Protocol Checklist for Student Jamie

Date of observation: ___/___/___ Time of observation: _____ to _____

Teachers present: _____ Observer: _____

Directions: Please indicate that a treatment step was completed by marking a \checkmark in the corresponding box.

- Reward slip present targeting the following behaviors:
- Following directions
 - Completing work
 - Keeping body still
- The selected reward is written at the bottom of the slip.
- Teacher (or aide) provides a \checkmark contingent on appropriate target behavior.
- Jamie earns a reward of his choosing approximately every 20 min.
- Verbal praise is paired with receipt of reward.
- Jamie is asked to select another reward at the start of the next 20-min interval.

of steps completed: _____ % steps completed: _____