

The Journal of Special Education

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

Intervention Fidelity in Special and General Education Research Journals

Elizabeth Swanson, Jeanne Wanzek, Christa Haring, Stephen Ciullo and Lisa McCulley
J Spec Educ published online 26 August 2011
DOI: 10.1177/0022466911419516

The online version of this article can be found at:
<http://sed.sagepub.com/content/early/2011/08/24/0022466911419516>

Published by:
Hammill Institute on Disabilities



and
SAGE

<http://www.sagepublications.com>

Additional services and information for *The Journal of Special Education* can be found at:

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

Subscriptions: <http://sed.sagepub.com/subscriptions>

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

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

Intervention Fidelity in Special and General Education Research Journals

The Journal of Special Education
XX(X) 1-11
© Hammill Institute on Disabilities 2011
Reprints and permission: <http://www.sagepub.com/journalsPermissions.nav>
DOI: 10.1177/0022466911419516
<http://journalofspecialeducation.sagepub.com>


Elizabeth Swanson¹, Jeanne Wanzek², Christa Haring¹,
Stephen Ciullo¹, and Lisa McCulley¹

Abstract

Treatment fidelity reporting practices are described for journals that published general and special education intervention research with high impact factors from 2005 through 2009. The authors reviewed research articles, reported the proportion of intervention studies that described fidelity measurement, detailed the components of fidelity measurement reported, and determined whether the components of fidelity reported differed based on the research design, the type of intervention, or the number of intervention sessions. Results indicate that even intervention research articles in high-quality general and special education journals inconsistently report fidelity (less than 70% of the articles). Authors of single-case studies most frequently reported the collection of intervention fidelity data (81.3% of articles, compared with 67.4% of treatment-comparison study articles). Of the 67% of articles that provided information about intervention fidelity procedure, only 9.8% provided data about the quality of the treatment intervention.

Keywords

treatment fidelity, intervention fidelity, intervention research

Broadly defined, intervention fidelity refers to the delivery of an intervention or program as designed (Gresham, MacMillan, Beebe-Frankenberger, & Bocian, 2000; O'Donnell, 2008). Collection and reporting of fidelity data in research reports are critical for determining why interventions succeed or fail (Dusenbury, Brannigan, Falco, & Hansen, 2003) and for determining whether an intervention can be scaled up to larger settings. Fidelity data are especially important when trying to account for negative or ambiguous findings and allow researchers to determine whether unsuccessful outcomes are due to ineffective interventions or due to a failure to implement the intervention as intended. This information not only helps researchers explain findings but can also serve to highlight areas of further study. For example, if one can attribute unsuccessful outcomes to poor implementation, it may signify the need for strengthening training materials or increasing classroom supports prior to generalizing with a wide variety of schools or teachers. However, unsuccessful outcomes may be attributed to poorly designed intervention, in which case intervention redesign is warranted.

Fidelity results also have implications for scaling up interventions in school settings. When an intervention produces positive outcomes and fidelity is high, the intervention can be adequately defined for scaling up in a broader variety of settings. Indeed, practitioners are becoming increasingly aware of the importance of choosing evidence-based interventions (see the No Child Left Behind Act of

2001 and the Individuals With Disabilities Education Act of 2004) that are feasible and effective and can be implemented with fidelity. Therefore, reports of research should describe the utility of the intervention, that is, the degree to which it is feasible and practical for implementation in schools (U.S. Department of Education, 2011).

Because fidelity data has the potential to inform researchers' work and practitioners' intervention choices, it is receiving increased attention from funders and evaluators of research. For example, the National Institute of Child Health and Human Development expects researchers to attend to issues of fidelity measurement stating, "broader examination and measurement of the instructional context is strongly encouraged to document and inform our understanding of fidelity of implementation" (National Institutes of Health, 2011). The Institute of Education Sciences encourages applicants to not only examine the efficacy of interventions but also to gather data to help explain the level of fidelity of implementation that is attained to help researchers identify the

¹The University of Texas at Austin, Austin, TX, USA

²Florida State University, Tallahassee, FL, USA

Corresponding Author:

Elizabeth Swanson, The Meadows Center for Preventing Educational Risk, The University of Texas at Austin, 1 University Station D4900, Austin, TX 78712, USA
E-mail: easwanson@mail.utexas.edu

“conditions, tools and procedures” that are needed to support implementation of the intervention (U.S. Department of Education, 2011). Likewise, organizations that evaluate the quality of intervention research, such as the What Works Clearinghouse and the National Center on Response to Intervention’s Technical Review Committee list several features of research design that improve confidence in findings. These include (a) the use of random assignment, (b) fidelity of implementation that is conducted with adequacy and provides evidence that the intervention was implemented as intended, and (c) the use of measures that are psychometrically reliable (National Center on Response to Intervention, 2011; U.S. Department of Education, 2008).

Although intervention fidelity data are critical to interpreting the outcomes of intervention research, practitioners’ intervention delivery, and successful funding for efficacy research, O’Donnell (2008) reported that fidelity of implementation is rarely reported in large-scale education studies or used for interpreting results. However, the issue of intervention fidelity is thoroughly addressed in the health literature (O’Donnell, 2008) and can serve as a resource to inform the field of education as we work to make forward progress in effective intervention fidelity data collection and reporting. In fact, among the top 20 most frequently cited articles with the term *fidelity* in the title (i.e., articles that most likely address issues directly related to fidelity), 14 are published in health-related, peer-reviewed journals. Only 2 are published in special education journals and 1 in general education journals. Therefore, it seems that the health field may contribute to our understanding of not only how to collect and report fidelity data but also how to use fidelity data to explain outcomes.

Fidelity Reporting in Health Research

Researchers conducting studies in the health sciences, where precise intervention delivery and accurate description of treatment components are critical to replication efforts, have influenced much of the understanding about intervention fidelity. Education research parallels health research, in that replication to broader populations is expected. In the past few years, authors in the health research community have written extensively about improving intervention fidelity reporting (e.g., Dusenbury et al., 2003).

In particular, the Treatment Fidelity Workgroup of the National Institutes of Health Behavior Change Consortium identified fidelity measurement strategies that may be applied to health studies (Bellg et al., 2004). First, the Fidelity Workgroup developed guidelines for monitoring the delivery of intervention. These guidelines encourage researchers to control for provider differences through close monitoring of participants’ perceptions of the provider (e.g., asking participants to provide feedback on

provider effectiveness, monitoring participant complaints, and having providers work with multiple-treatment groups). Second, the Fidelity Workgroup suggested strategies to reduce differences within treatment to ensure that providers in the same condition deliver the same intervention. These strategies include using scripted intervention protocols, providing a treatment manual, and reviewing audio- and videotaped sessions. Third, the Fidelity Workgroup stressed adherence to treatment protocol, including content and dosage. The work group encouraged researchers to review the audio- or videotapes for protocol adherence and to ask providers to complete self-report checklists, indicating components delivered during the treatment. Finally, the Fidelity Workgroup emphasized minimizing contamination between treatment and control conditions, especially when implemented by the same provider. Here, the work group encouraged researchers to provide explicit training to providers regarding the rationale for keeping the conditions separate, frequent provider supervision, and audiotaped or in-person observations with review and feedback. The Fidelity Workgroup also suggested that participant exit interviews measure the extent to which participants in the control condition received treatment components. Within health research, authors not only suggest procedures for collecting fidelity data during intervention delivery but also provide guidelines for standardizing treatments to ensure that all participants receive the same treatment and systematically improving protocol adherence. Perhaps the idea of actively effecting protocol adherence is one way health research can inform general and special education intervention researchers’ work.

Fidelity Reporting in Education Intervention Research

General education reviews of fidelity reporting. In 2004, the National Research Council (NRC) reviewed the quality of K-12 mathematics instructional evaluations funded by the National Science Foundation and reported that 44% of the 63 minimally adequate studies measured fidelity. Only one of these adjusted outcomes and conclusions was based on the collected fidelity data (NRC, 2004). O’Donnell (2008) recently examined a corpus of 23 intervention studies that reported fidelity of implementation. Among these, 10 studies provided evidence of fidelity data collection of K-12 intervention for a core subject (e.g., science, social studies, math, reading) that can be implemented by a single teacher. Of these, 9 provided a quantitative measure of fidelity, whereas only 5 investigated how fidelity relates to outcomes.

General education suggestions for fidelity. O’Donnell (2008) made several recommendations to researchers based on his review of fidelity reporting in the literature: (a) establish a theory that drives the intervention and determine what it means to implement with fidelity, (b) clearly define critical

components of the intervention, (c) collect fidelity data on a randomly selected subset of intervention deliverers or on the entire sample, (d) measure fidelity to critical components and processes present in the intervention, and (e) test for and report the reliability and validity of the fidelity data collected.

Special education reviews of fidelity reporting. Similar to general education, previous studies examining the extent to which fidelity has been reported in special education research have not yielded promising findings. In a review of all school-based experimental studies published in the *Journal of Applied Behavior Analysis* between 1991 and 2005, authors reported that only 30% of studies provided treatment fidelity data (McIntyre, Gresham, DiGennaro, & Reed, 2007). In a similar review, Dane and Schneider (1998) evaluated intervention fidelity reports among primary and early secondary prevention studies of behavioral, social, or academic “maladjustment” from 1980 to 1994. Of the 162 studies reviewed, only 39 specified procedures for reporting treatment fidelity. The authors concluded that not conducting treatment fidelity checks threatens the internal validity of published studies.

In 2000, Gresham et al. examined three special education journals (*Journal of Learning Disabilities*, *Learning Disability Quarterly*, and *Learning Disabilities Research and Practice*) spanning a 5-year period to determine the extent to which fidelity measures were reported for intervention studies. Although 49% of the studies described treatment fidelity to some extent, 32.3% did not mention treatment fidelity at all. The authors reported that only 18.5% of the articles included evidence that treatment fidelity was measured at a rigorous level. The authors concluded that an absence or lack of consistency of fidelity reporting makes it difficult to determine whether intervention studies are effective.

Special education suggestions for fidelity. In 2005, a panel of researchers (Gersten, Fuchs, et al.) identified and described “quality indicators” for randomized control trials and quasiexperimental studies in special education. The resulting guidelines encouraged careful research design and transparent reporting of findings. Important recommendations included providing detailed participant descriptions, using multiple measures to capture an intervention’s effect, providing effect size calculations, and matching data analysis techniques to key research questions. Fidelity reporting was listed as essential for identifying interventions that are “evidence based,” suggesting that treatment and control conditions in experimental research should undergo complete and thorough investigations to establish intervention fidelity. Intervention fidelity reports, the panel suggested, should include data regarding the “surface features of fidelity implementation” through (a) reporting fidelity data collection and resulting scores, (b) conducting regular observations of the intervention, (c) using a checklist of treatment components to record whether the most critical

aspects of the intervention occurred, (d) providing a record of the number of days or sessions the intervention was conducted, and (e) reporting interrater reliability among observers. In addition, the panel suggested that quality of implementation should be investigated and reported to determine whether an intervention was implemented with varying degrees of integrity (Gersten, Fuchs, et al., 2005).

Quality indicators for single-case research have been developed in recent years as well (Horner, Carr, Halle, McGee, Odom, et al., 2005; Kratochwill, Hitchcock, Horner, Levin, Odom, Rindskopf, & Shadish, 2010). Fidelity measurement was listed among others, as an essential component of high-quality single-case studies. “Fidelity of implementation is a significant concern within single-subject research because the independent variable is applied over time” (Horner et al., 2005, p. 168). Researchers are encouraged to document and report fidelity of implementation via ongoing direct measurement during each phase of the study (Horner et al., 2005).

Purpose

Work conducted in the health, general education, and special education fields provides fidelity data collection guidelines for intervention researchers. Given these guidelines, we were interested in the extent to which fidelity data have been collected and reported over the past few years. Therefore, the purpose of this article is twofold. First, to highlight the issue of fidelity reporting, it is important to examine the status of intervention fidelity reporting in the most widely cited journals in the fields of special education and general education. Second, it is possible that reports of fidelity procedures and data are lacking in identifiable ways. Identifying weaknesses in reporting provides authors a focus point for improving the collection and reporting of fidelity data.

We expand on previous examinations of treatment fidelity reporting in intervention research in several ways. First, a broader and larger group of journals in both general and special education were purposely selected based on two criteria: the journal published primarily intervention research and demonstrated high impact scores during a 5-year period. As previously mentioned, impact scores are indicative of how often articles are read and cited. Review of the most frequently cited articles is a good way to gauge the status of a field of study. Second, several components of fidelity (as suggested by Gersten, Fuchs, et al., 2005; O’Donnell, 2008; and What Works Clearinghouse) were investigated, including the type of fidelity measure, the number of fidelity data points collected, and the extent to which fidelity was reported in different research designs (e.g., experimental study, single-case study).

The following research questions were addressed separately for general and special education journals:

Research Question 1: What proportion of intervention studies published in peer-reviewed journals reported fidelity?

Research Question 2: What components of fidelity measurement were reported in published intervention studies?

Research Question 3: Did reported components differ according to research design, intervention type, or number of intervention sessions?

Method

Journal-Selection Criteria

For inclusion in this study, journals had to meet three criteria. First, journals had to be identified as either “general education” or “special education” by the Institute for Scientific Information (Institute for Scientific Information, 2010). General education and special education journals were confirmed by publishers’ descriptions. Second, the journal must have published intervention research in the fields of reading, writing, and/or mathematics. Because the purpose of this article was to examine use of fidelity in academic intervention research, we excluded several general education and special education journals. For example, some journals focused primarily on reviews of research (e.g., *Journal of Research in Special Education Needs*) or other subject areas such as language or behavioral interventions (e.g., *Journal of Speech, Language, and Hearing Research*). Journals that disseminated primarily theory-driven information (e.g., *Educational Philosophy and Theory*) were also excluded.

Third, we selected the five journals with the highest impact factors in the areas of general and special education. The ISI defines an impact factor as “a measure of the frequency with which the average article in a journal has been cited in a given period of time” (http://thomsonreuters.com/products_services/science/free/essays/impact_factor/). A journal’s 5-year impact factor is determined by counting the number of times articles from a 5-year time period were cited over the course of the following year and dividing that number by the total number of articles published over the same 5-year period. The resulting impact scores of educational journals most often fall within a range of 0.05 to 5.0 and allow researchers to compare the relative impact of publications of varying sizes, topics, or readerships without bias. The five general education journals that met the criteria for inclusion and their associated impact factors are as follows: *American Education Research Journal* (2.874), *Scientific Studies of Reading* (2.834), *Early Childhood Research Quarterly* (1.940), *Elementary School Journal* (1.876), and *Reading Research Quarterly* (1.717). The five special education journals with the highest impact factors were also included in the study: *Annals of Dyslexia*

(3.259), *Exceptional Children* (2.894), *Journal of Learning Disabilities* (2.157), *Dyslexia* (1.938), and *Journal of Special Education* (1.938).

Article-Selection Criteria

Authors hand-searched all issues of the identified peer-reviewed journals published between 2005 and 2009 (matching the years used to calculate the 5-year impact factor) and chose articles that met the following criteria:

1. Participants were in a school setting (prekindergarten through 12th grade).
2. The independent variable was described as a student-level reading, mathematics, or writing intervention. Observation studies of classroom practices with no manipulated independent variable (e.g., Swanson & Vaughn, 2010) were excluded.
3. A teacher, paraprofessional, or researcher delivered the intervention.
4. Student-level data were collected and reported.
5. Authors reported experimental, quasiexperimental, or single-subject designs.

Article-selection reliability checks were conducted, whereby a second author hand-searched 10% of the issues in each journal. Reliability for article selection was 97% across journals. The search yielded 76 intervention articles that met selection criteria.

Code Sheet

Using a code sheet designed to guide systematic examination of reported fidelity procedures, each article was coded for variables, including journal title, type of intervention (i.e., math, reading, or writing), research design, number of intervention sessions, length of sessions, duration of intervention in weeks, frequency of intervention, fidelity data collection procedure, fidelity measures (e.g., self-administered checklist, observation, audio recording, video recording), and use of fidelity data in analysis and reported results. For each fidelity measure listed in an article, an additional code sheet was completed. For example, if observations were conducted, the coder completed an additional set of items that included number of observations, frequency of observations, and focus of observation (e.g., teacher or student).

Procedure

Training and interrater reliability. The coding processes suggested by Cooper, Hedges, and Valentine (2009) were adapted for the needs of our study. The first author trained two senior doctoral students on the use of the coding

instrument during a series of meetings. All coders have experience conducting federally funded randomized control trials and participated in fidelity data collection during the intervention studies. First, an overview of the synthesis was provided. Each item on the code sheet was read and discussed. Then, for practice in using the code sheet, the first author guided several coding sessions where four separate articles were coded. After training, reliability of coding was established. The coding team (made up of the first author and two senior doctoral students) independently coded an article, then met to discuss each item. Percentage of agreement was established (agreements divided by agreements plus disagreements). Coders were asked to justify why codes were assigned, which provided an additional training opportunity. This process was conducted twice to meet the 90% interrater reliability requirement. Reliability after the first round of coding ranged from 85% to 97%. A second round of coding resulted in 90% to 95% agreement across items.

Coding. The set of 76 articles were equally divided between the two senior doctoral students. Coding was completed through a four-step procedure that relied on independent coding and repeated meetings to discuss coding difficulties and ambiguities (Cooper et al., 2009). These procedures not only increase the accuracy of codes but also serve to guard against coding bias. First, each doctoral student independently coded his or her assigned articles. Second, the coders traded articles. This time, the second coder checked all codes. Any discrepancies were noted. Third, the coders met to discuss discrepancies and came to consensus on the correct codes. Finally, the first author reviewed every code sheet for accuracy. Any questionable codes were discussed with the team to reach consensus on correct codes.

Results

Of the 76 intervention research articles that met selection criteria, 50 included a report of intervention fidelity. Table 1 summarizes treatment fidelity by research design, sample size, intervention type (mathematics, reading, writing, or a combination), duration of intervention (number of sessions and length of sessions), and implementer. A more in-depth examination of the characteristics of intervention fidelity data collection and reporting follows.

Report of Fidelity Data Collection and Resulting Scores

By research design. Of the 76 identified intervention studies, most used treatment-comparison designs ($n = 43$); the others reported multiple-treatment (with no “business-as-usual” comparison group; $n = 17$), single-group ($n = 3$), or single-case designs ($n = 13$). Authors of the reviewed single-case studies most frequently reported the collection of intervention fidelity data, 81.3% of the articles, compared with

67.4% of treatment-comparison study articles. Authors of multiple-treatment studies least frequently reported the collection of intervention fidelity data (56.3% of the articles). Of the 43 treatment-comparison articles, 29 reported fidelity procedures alone (67.4%) and 23 reported fidelity procedures and scores (53.4%). Within the set of treatment-comparison articles, we were also interested in identifying any fidelity reporting differences based on experimental ($n = 27$) and quasiexperimental ($n = 16$) study design. Among experimental studies, 19 of them (70%) reported fidelity procedures and 15 reported fidelity procedures and scores (56%). This trend was similar among quasiexperimental designed studies, with 10 studies reporting fidelity procedures (63%) and 8 studies reporting procedures and scores (50%).

By intervention type. The greatest number of studies investigated the effectiveness of reading interventions ($n = 55$), followed by mathematics ($n = 10$), writing ($n = 7$), mathematics plus reading ($n = 2$), and reading plus writing ($n = 2$). Neither of the mathematics-plus-reading studies reported fidelity data collection procedures or scores. Authors of reading-plus-writing studies most frequently reported fidelity procedures (100% of the articles), followed by authors of writing studies (85.7% of the articles), mathematics-only studies (80% of the articles), and reading-only studies (63.6% of the articles). It is worth noting that the type of intervention with the smallest number of articles, reading plus writing, most frequently reported fidelity procedures, and the intervention type with the most articles, reading, had the lowest percentage of fidelity reporting. Authors of mathematics-only studies most frequently reported fidelity scores (80% of the articles), followed by authors of writing studies (71.4% of the articles), reading-only studies (41.8% of the articles), and the combination studies (no articles reported fidelity scores). Five articles reported quality of implementation: two reading-only studies and one each for mathematics-only, writing-only, and reading-plus-writing studies.

Fidelity Data Collection Procedures

By research design. Most authors of treatment-comparison studies who reported fidelity data collection relied on observations ($n = 24$). The number of observations ranged from 1 to 176. None of the articles included a copy of the observation form. However, descriptions of the observation form were provided in 12 articles. Although only 1 article explicitly reported random selection of sessions for observation and 2 articles explicitly reported nonrandom selection of sessions for observation, the other 22 articles did not report whether sessions were randomly chosen. In all, 12 articles reported an interrater reliability score among observers for fidelity. See Table 2 for other types of fidelity data collection that researchers used.

Seven authors of single-case studies who reported fidelity collection relied on self-report checklists completed by

Table 1. Number of Articles by Category That Report Fidelity Data Collection

Category	General education				Special education			
	Total articles	Articles reporting fidelity procedures	Articles reporting fidelity scores	Articles reporting quality of implementation	Total articles	Articles reporting fidelity procedures	Articles reporting fidelity scores	Articles reporting quality of implementation
Research design								
Treatment comparison	19	11	9	3	24	18	14	0
Single case	1	1	1	0	12	8	8	0
Multiple treatment	4	3	1	1	13	7	2	1
Single treatment	2	2	1	0	1	0	0	0
Sample size								
1–10	0	0	0	0	12	9	8	0
11–20	0	0	0	0	0	0	0	0
21–50	4	1	2	0	15	8	7	0
51–100	7	4	3	1	11	8	5	0
101–200	7	3	1	1	8	5	2	0
>200	8	8	6	2	4	4	2	1
Intervention type								
Reading only	20	12	8	1	35	23	15	1
Mathematics only	3	3	3	1	7	5	5	0
Writing only	2	1	1	1	5	5	4	0
Mathematics and reading	0	0	0	0	2	0	0	0
Reading and writing	1	1	0	1	1	1	0	0
Number of sessions								
<10	3	1	1	0	7	2	2	0
10–20	2	0	0	0	7	6	4	0
21–40	1	1	1	0	3	2	1	0
41–60	1	0	0	0	3	1	1	0
>60	3	3	2	2	3	2	2	1
NR	12	12	8	2	31	21	14	0
Length of sessions								
<30 min	9	6	3	2	13	7	5	0
31–45 min	8	6	5	0	21	16	11	0
46–60 min	4	4	3	2	7	5	3	0
>60 min	0	0	0	0	2	2	2	0
NR	4	1	1	0	8	4	3	1
Implementer								
Teacher	17	14	10	2	20	16	11	0
Researcher	2	0	0	0	11	7	6	0
Research assistant	3	3	2	1	4	3	2	0
Computer	1	0	0	0	6	0	0	0
Paraprofessional	0	0	0	0	3	3	2	0
Other	2	0	0	0	5	5	3	0
NR	1	0	0	0	1	0	0	0

Abbreviation: NR, not reported.

the teacher or researcher who implemented the intervention for the purpose of “protocol integrity” and to “ensure that instruction was delivered as planned.” All but two articles indicated that these checklists were completed for every lesson. One article reported completing the checklist once per week, and the final article did not report frequency of

administration. One article provided a copy of the self-administered checklist, and one article described the checklist. In six of the studies, authors also conducted observations of instruction in addition to self-reports. The number of observations ranged from one lesson to 40% of all lessons delivered. No observation protocols were included or

Table 2. Types of Fidelity Data Collection by Research Design

Research design	Self-administered checklist (n)		Observation (n)		Audio recording (n)		Intervention dosage (n)		Video recordings (n)		Type not reported (n)	
	GE	SE	GE	SE	GE	SE	GE	SE	GE	SE	GE	SE
Treatment comparison	2	1	8	16	2	1	0	3	0	1	0	0
Single case	0	7	1	5	0	2	1	0	1	0	0	0
Multiple treatment	0	0	3	5	0	0	0	0	1	0	0	1
Single treatment	0	0	1	0	0	0	0	0	0	0	0	0

Abbreviations: GE, = general education; SE, special education.
Some articles collected several types of fidelity data.

Table 3. Types of Fidelity Data Collection by Intervention Type

Intervention type	Self-administered checklist (n)		Observation (n)		Audio recording (n)		Intervention dosage (n)		Video recordings (n)		Type not reported (n)	
	GE	SE	GE	SE	GE	SE	GE	SE	GE	SE	GE	SE
Reading only	0	3	11	20	0	1	1	3	2	1	0	0
Math only	1	0	1	4	1	1	0	0	0	0	0	0
Writing only	1	5	0	2	1	1	0	0	0	0	0	0
Reading plus writing	0	0	1	0	0	0	0	0	0	0	0	1

Abbreviations: GE, general education; SE, special education.
Some articles collected several types of fidelity data.

described in any of the articles. One article reported interrater reliability scores.

By intervention type. Most authors of reading-only interventions relied on observations to collect intervention fidelity data ($n = 31$). The number of observations ranged from 1 to 176. No articles included a copy of the observation form; 14 articles contained a description of the form. One of the 31 articles reported random selection of intervention periods observed. A total of 19 articles reported interrater reliability scores. Of the 3 articles that reported use of a self-administered checklist, none reported the number completed. A copy of the checklist was provided in 1 article; a description was provided in the other 2. In all, 4 reading-only intervention articles reported using intervention dosage as a measure of intervention fidelity. They defined dosage as either number of pages read plus time spent reading or the number of lessons completed. In all cases, dosage was reported daily. Finally, video recordings were used in 3 articles. Only 1 article reported the number of videos collected (two per teacher). No article included a copy of the code sheet used to interpret the contents of the video, but 1 article described it. In one case, authors reported randomly selecting sessions to be video recorded.

Most mathematics-only intervention studies relied on observations to collect fidelity data. No article included a copy of the observation form, and only one included a description of the form. No authors reported whether observed

intervention sessions were randomly selected. Two articles included reports of interrater reliability procedure and the resulting scores. One additional article provided an interrater reliability score but no procedure for establishing it. See Table 3 for other types of fidelity data collection that researchers used.

By number of intervention sessions. Gersten, Fuchs, et al. (2005) indicated that reporting the number of days and/or sessions during which intervention was conducted allowed reviewers to determine the intensity of the intervention and provide guidelines for replication. Although the authors did not offer specific guidelines for the number and frequency of data collection, they suggested that researchers collect fidelity data over the course of the study and on a regular basis. To determine the extent to which fidelity data were collected on a comprehensive and regular basis in the studies reviewed for this article, an evaluation of fidelity reporting practices according to the number of sessions follows.

A total of 32 of the 76 articles reported the number of sessions provided to students during the intervention period. Refer to Table 4 for information about the type of fidelity measures used. Of these 32 articles, 12 conducted fidelity observations. Authors of 6 articles provided enough information to calculate the proportion of sessions observed for fidelity purposes. Of these, 2 articles reported conducting observations for between 2% and 3% of intervention sessions (range of intervention sessions for these two articles was 41–155). In the article with 115 sessions, authors

Table 4. Types of Fidelity Data Collection by Number of Sessions

Number of sessions	Self-administered checklist (n)		Observation (n)		Audio recording (n)		Intervention dosage (n)		Video recordings (n)		Type not reported (n)	
	GE	SE	GE	SE	GE	SE	GE	SE	GE	SE	GE	SE
< 10	0	1	1	2	0	0	0	0	0	0	0	0
10–20	0	4	0	3	0	1	0	1	0	0	0	0
21–40	1	1	0	1	0	0	0	0	0	0	0	0
41–60	9	0	0	1	0	0	0	0	0	0	0	0
> 60	0	0	2	2	1	0	1	0	0	0	0	0
NR	0	2	10	17	1	2	0	2	0	0	0	1

Abbreviations: GE, general education; SE, special education.
Some articles collected several types of fidelity data.

conducted fidelity observations at the beginning, middle, and end of the year, indicating that fidelity data were collected over the course of the study but perhaps not on a regular basis. One article reported that 15% of the sessions were observed but provided no information about the frequency of collection during the 53rd to 126th intervention sessions. Three articles reported observing between 30% and 40% of sessions. These 3 small-scale, multiple-treatment, or single-case studies reported fidelity observations over the course of the study and on a regular basis.

Report of Intervention Quality

In the articles that reported fidelity data collection, authors most often cited reasons such as “to ensure treatment is delivered as intended” for data collection. No author explicitly stated that fidelity data were collected to also ascertain the degree to which interventions were delivered with quality. However, authors of five studies did report quality-related fidelity data.

Fidelity Data Used to Explain Conclusions

One of the primary reasons to collect fidelity data during a study is to provide evidence of internal validity, ensuring that the intervention is related to the effect, and external validity, the degree to which study results may generalize to other settings. Of the 50 studies that reported fidelity data collection, only 2 provided a discussion of issues related to internal validity. One study (Clements & Sarama, 2008) provided a discussion of the way in which fidelity was related to achievement gains in the study sample. Although authors reported that fidelity scores were not related to gains at a statistically significant level, they were positively correlated with students' gains in achievement. In another study (O'Connor, White, & Swanson, 2007), authors suspected that a lack of differences between groups in vocabulary outcomes was due to differential exposure to text. Through the use of fidelity data, they established that

the amount of text read was different between groups; however, authors could not rule that the type of text read by the poor readers in the study contained too few unknown vocabulary words to provide the advantage of practicing new words in the context of connected text. Similarly, 4 studies included discussion of external validity, claiming that high fidelity ratings provided evidence that paraprofessionals and/or teachers were able to deliver interventions as designed. Fidelity scores were not included as moderators of student outcomes in any study.

Fidelity Data Reporting in General and Special Education Journals

The last indicator we examined was the extent to which fidelity data were reported in general and special education journals (as defined by ISI Web of Knowledge Journal Citation Reports described in the “Method” section). Considerably more intervention studies were reported in special education journals ($n = 50$) than general education journals ($n = 26$), with a higher percentage of treatment fidelity reported in special education journals (68%) than general education journals (65%). The types of fidelity data collected were similarly reported in both general and special education journals.

Discussion

The purpose of this study was to examine the reporting of intervention fidelity in articles published in high-impact general and special education journals. We used high-impact journals as a “best evidence” case of how intervention fidelity is being reported in research studies. As has been noted in previous research (e.g., Mastropieri et al., 2009), only a small portion of the published articles in these journals reported intervention research. A total of 76 articles were reviewed, and 67% of those articles provided fidelity procedures for the interventions implemented. In all, 47% provided quantitative data (or scores) on fidelity

implementation. These percentages are substantially higher than have been reported in previous research (e.g., Gresham et al., 2000; NRC, 2004; O'Donnell, 2008), suggesting that there may be an increase in the collection and reporting of fidelity data in recent years. Although fidelity of implementation procedures were provided in 67% of the intervention articles, only 6.6% of the articles provided information on the quality of implementation as Gersten, Fuchs, et al. (2005) have recommended. There are several reasons why the quality of intervention implementation may not be a focus of data collection within current academic intervention research. First, it is possible that researchers consider quality but embed it in their fidelity procedures, rather than code it as a separate component. For example, on a checklist of intervention components, an interventionist may receive credit for implementing a comprehension strategy component only if the strategy is modeled with clear language. If this level of detail were not provided in the corresponding research article, the quality indicator would be missing. Second, quality of implementation can be difficult to measure because it requires some subjectivity in scoring. However, there is sufficient previous research to suggest that interrater reliability in scoring quality of instruction is possible and that quality of instruction and student outcomes are related (Gersten, Baker, Haager, & Graves, 2005). As Gersten, Fuchs, et al. suggested, collecting data on the quality of implementation may provide information that data collection only on the components of the intervention does not sufficiently capture. Thus, data on the quality of implementation may be particularly helpful in efficacy studies that intend to provide causal links between intervention and student outcomes.

Observation (live, audio, or video) was the most common form of data collection in the studies reviewed. Among the 50 studies that reported fidelity procedures, 88% ($n = 44$) reported using some type of classroom observation. Observations can provide precise estimates of intervention implementation and may be more reliable data than self-reports or questionnaires that depend on the implementer to objectively remember and report about the implementation. Observations can also be one of the most expensive forms of fidelity data collection, often requiring additional personnel to attend intervention sessions or to engage in time-consuming coding of intervention sessions. Our findings suggest that when researchers are reporting fidelity, they most often choose observation data. The time and cost of observation data may have been prohibitive for some of the studies that did not report fidelity data.

Both Gersten, Fuchs, et al. (2005) and O'Donnell (2008) highlighted the need for reporting reliability in fidelity data. However, only 45 of the articles collecting fidelity information with observation techniques provided information on the reliability of the raters collecting the observation data. Interrater reliability was generally high (more

than 90%) when it was reported in the articles. When fidelity data are reliable, important information regarding the internal and external validity of the study can be ascertained; however, without evidence of reliability of the data, it is impossible to establish the extent to which the intervention was implemented as intended.

Limitations

The data presented in this study represent articles in high-impact general and special education journals publishing intervention research from 2005 to 2009. Of course, many journals were not included in this sample, making the sample not representative of all peer-reviewed articles. For example, only studies that manipulated academic interventions were included in this study. It is possible that fidelity reporting trends among studies that manipulate behavior-based interventions are quite different. While this synthesis cannot provide information related to fidelity data collection among behavior-based interventions, it is viable population of studies to investigate.

Our findings of fidelity data collection and reporting do not necessarily reflect the total quality of intervention studies being published in high-impact journals. There are many aspects related to the quality of intervention studies that we do not address here, with fidelity representing only one key area.

Implications

Although we found that two thirds of the intervention research articles published in these high-impact journals provided information on fidelity of implementation, the fact that not all of the articles incorporated even basic fidelity information suggests that there is not yet a widespread standard for reporting fidelity data in intervention research. Admittedly, page limitations may have constrained the authors of the articles in terms of the amount of fidelity information that could be provided. For example, it is not surprising, considering page limitations, that none of the articles included the checklist used for fidelity data collection. Nevertheless, our findings suggest that reporting of fidelity data is becoming more common in published articles in both general and special education, at least in high-impact journals. In order for educators and other researchers to adequately interpret the results of intervention research, precise collection and reporting of fidelity data are required. In addition, the precision of fidelity information can be increased with reporting of fidelity scores and the examination of quality of instruction in addition to the occurrence of intervention steps or components (Gersten, Fuchs, et al., 2005). Currently, only about half of the studies include fidelity scores, and very few studies have included quality of implementation as a component of fidelity data collection

and reporting. Thus, there continues to be room for improvement for intervention researchers to collect and report more precise fidelity information in the form of quantitative scores and quality of intervention implementation. In addition, fidelity data are difficult to interpret when the reliability of the data are unknown. Less than half of the studies reporting on fidelity provided reliability information. To improve the quality of fidelity reporting in published articles, it is important for researchers to plan for fidelity data collection prior to study implementation so that critical components of the intervention can be identified, and research personnel can be trained adequately for collecting data allowing for adequate reporting of the reliability of the data collected. Continued increases in the reporting of fidelity data in published articles of intervention research along with increases in the quality of reporting will allow readers to better interpret article findings. Finally, journal editors serve as a “gate-keeper” in determining the quality level of research that enters into publication. If researchers submit manuscripts with key components of fidelity data collection and reporting missing, it is the journal editor’s charge to encourage authors to reveal more information. Although there is still much to learn about ways in which fidelity data can be used to explain results, it is clear that fidelity data used at a basic level (i.e., descriptively to explain whether the intervention was implemented as intended) are essential to high-quality research.

Authors’ Note

The opinions expressed are those of the authors and do not represent views of the Institute of Education Sciences or the U.S. Department of Education.

Declaration of Conflicting Interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by an award granted to The Meadows Center for Preventing Educational Risk at The University of Texas at Austin by the Institute of Education Sciences, U.S. Department of Education, the Reading for Understanding Research Initiative, R305F100013.

References

- Bellg, A. J., Resnick, B., Minicucci, D. S., Ogedegbe, G., Ernst, D., Borrelli, B., . . . Czajkowski, S. (2004). Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the NIH behavior change consortium. *Health Psychology, 23*, 443–451.
- Clements, D. H., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Educational Research Journal, 45*, 443–494.
- Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis*. New York, NY: Russell Sage Foundation.
- 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*, 23–35.
- Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research, 18*, 237–256.
- Gersten, R., Baker, S. K., Haager, D., & Graves, A. W. (2005). Exploring the role of teacher quality in predicting reading outcomes for first-grade English learners. *Remedial and Special Education, 26*, 197–206. doi:10.1177/07419325050260040201
- 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., 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 & 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.
- Institute for Scientific Information. (n.d.). *Web of Knowledge Journal Citation Reports*. Available from <http://apps.isiknowledge.com>
- Institute for Scientific Information. (2010). *Journal citation reports*. Philadelphia, PA: Institute for Scientific Information.
- Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). *Single-case designs technical documentation*. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf
- Mastropieri, M. A., Berkeley, S., McDuffie, K. A., Graff, H., Marshak, L., Connors, N. A., . . . Cuenca-Sanchez, Y. (2009). What is published in the field of special education? An analysis of 11 prominent journals. *Exceptional Children, 76*, 95–109.
- McIntyre, L. L., Gresham, F. M., DiGennaro, F. D., & Reed, D. D. (2007). Treatment integrity of school-based interventions with children in the Journal of Applied Behavior Analysis 1991–2005. *Journal of Applied Behavior Analysis, 40*, 659–672.
- National Center on Response to Intervention. (2011). *Instruction tools chart*. Washington, DC: Author. Retrieved from http://www.rti4success.org/tools_charts/instruction.php
- National Institutes of Health. (2011). *Learning disabilities research center request for application* (FOA Number RFA-HD-12-202).

- Washington, DC: Author. Retrieved from <http://grants.nih.gov/grants/guide/rfa-files/RFA-HD-12-202.html>
- National Research Council. (2004). *On evaluating curricular effectiveness: Judging the quality of K-12 mathematics evaluations*. Washington, DC: The National Academies Press.
- O'Connor, R. E., White, A., & Swanson, H. L. (2007). Repeated reading versus continuous reading: Influences on reading fluency and comprehension. *Exceptional Children, 74*, 31–46.
- O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research, 78*, 33–84.
- Swanson, E. A., & Vaughn, S. (2010). An observation study of reading instruction provided to elementary students with learning disabilities in the resource room. *Psychology in the Schools, 47*, 481–492.
- U.S. Department of Education. (2008). *WWC Procedures and Standards Handbook: Version 2.0*. Washington, DC: Institute of Education Sciences. Retrieved from <http://ies.ed.gov/ncee/wwc/references/idocviewer/doc.aspx?docid=19&tocid=1>
- U.S. Department of Education. (2011). *Special education research grant request for application* (CFDA Number 84.324A). Washington, DC: Institute for Education Sciences.

About the Authors

Elizabeth Swanson is a research assistant professor at The Meadows Center for Preventing Educational Risk, The University of Texas at Austin. She is currently a coprincipal investigator of an Institute for Education Sciences–funded Reading for Understanding project investigating instructional interventions

designed to promote reading comprehension and remediate reading difficulties among adolescent readers.

Jeanne Wanzek, PhD, is an assistant professor in special education at Florida State University and on the research faculty at the Florida Center for Reading Research. Her research interests include learning disabilities, reading, effective instruction, and response to intervention.

Christa Haring is a doctoral student at the University of Texas of Austin. She currently works as an intervention coach at the Meadows Center for Preventing Educational Risk. Her research interests include assessment of students with disabilities, educational practices that increase student engagement and instructional coaching.

Stephen Ciullo is a doctoral candidate at the University of Texas at Austin and a researcher for the Meadows Center for Preventing Educational Risk at The University of Texas at Austin. His primary research interests are reading comprehension strategies and interventions for students with learning disabilities or struggling readers in middle grades (4th–8th).

Lisa McCulley is a research associate at The Meadows Center for Preventing Educational Risk and a doctoral student in the Department of Special Education, The University of Texas at Austin, with a concentration in learning disabilities and behavior disorders. Her current interests include investigating effective instructional methods and interventions that improve reading comprehension for struggling adolescent learners.