

TRAINING TEACHERS TO USE EVIDENCE-BASED PRACTICES FOR AUTISM: EXAMINING PROCEDURAL IMPLEMENTATION FIDELITY

AUBYN C. STAHMER AND SARAH RIETH

Rady Children's Hospital, San Diego and University of California, San Diego

EMBER LEE

Rady Children's Hospital, San Diego

ERICA M. REISINGER AND DAVID S. MANDELL

The Children's Hospital of Philadelphia Center for Autism Research

JAMES E. CONNELL

AJ Drexel Autism Institute

The purpose of this study was to examine the extent to which public school teachers implemented evidence-based interventions for students with autism in the way these practices were designed. Evidence-based practices for students with autism are rarely incorporated into community settings, and little is known about the quality of implementation. An indicator of intervention quality is procedural implementation fidelity (the degree to which a treatment is implemented as prescribed). Procedural fidelity likely affects student outcomes. This project examined procedural implementation fidelity of three evidence-based practices used in a randomized trial of a comprehensive program for students with autism in partnership with a large, urban school district. Results indicate that teachers in public school special education classrooms can learn to implement evidence-based strategies; however, they require extensive training, coaching, and time to reach and maintain moderate procedural implementation fidelity. Procedural fidelity over time and across intervention strategies is examined. © 2014 Wiley Periodicals, Inc.

Special education enrollment for children with autism in the United States has quadrupled since 2000 (Scull & Winkler, 2011), and schools struggle to provide adequate programming to these students. A growing number of interventions for children with autism have been proven efficacious in university-based research settings, but much less attention has been given to practical issues of implementing these programs in the classroom, where most children with autism receive the majority of their care (Sindelar, Brownell, & Billingsley, 2010). In general, evidence-based practices for children with autism are rarely incorporated into community settings (Stahmer & Ingersoll, 2004). Teachers in public schools report receiving inadequate training and rate their personal efficacy in working with children with autism as low (Jennett, Harris, & Mesibov, 2003). Training public educators to provide evidence-based practices to children with autism is a central issue facing the field (Simpson, de Boer-Ott, & Smith-Myles, 2003).

One major challenge to implementing evidence-based practices for children with autism in community settings is the complexity of these practices. Strategies based on the principles of applied behavior analysis have the strongest evidence to support their use (National Standards

This research was funded by grants from the National Institute of Mental Health (5R01MH083717) and the Institute of Education Sciences (R324A080195). We thank the School District of Philadelphia and its teachers and families for their collaboration and support. Additionally, Dr. Stahmer is an investigator with the Implementation Research Institute at the George Warren Brown School of Social Work, Washington University, St. Louis, through an award from the National Institute of Mental Health (R25MH080916).

Correspondence to: Aubyn C. Stahmer, Child and Adolescent Services Research Center & Autism Discovery Institute, Rady Children's Hospital, San Diego, 3020 Children's Way, MC5033, San Diego, CA 92123. E-mail: astahmer@ucsd.edu

Project, 2009). These practices vary greatly in structure and difficulty. Some strategies, such as discrete trial teaching (DTT; Leaf & McEachin, 1999; Lovaas, 1987), are highly structured and occur in one-on-one settings, whereas others are naturalistic, can be conducted individually or during daily activities, and tend to be more complex to implement (e.g., incidental teaching; Fenske, Krantz, & McClannahan, 2001; or pivotal response training [PRT]; Koegel et al., 1989). There are also classroom-wide strategies and structures based on applied behavior analysis, such as teaching within functional routines (FR; Brown, Evans, Weed, & Owen, 1987; Cooper, Heron, & Heward, 1987; Marcus, Schopler, & Lord, 2000; McClannahan & Krantz, 1999). Although all of these evidence-based practices share the common foundational principles of applied behavior analysis, each is made up of different techniques. These and other intervention techniques are often packaged together as “comprehensive interventions” (Odom, Boyd, Hall, & Hume, 2010) or used in combination in the field to facilitate learning and expand the conditions under which new student behaviors occur (Hess, Morrier, Heflin, & Ivey, 2008; Stahmer, 2007).

Teachers can learn these evidence-based strategies within the context of a research study (e.g., Suhrheinrich, 2011); however, studies report a highly variable number of hours of training needed to master the intervention strategy. For example, the amount of time required to train classroom educators in DTT in published studies ranges from 3 hours (Sarokoff & Sturmey, 2004) at its most brief, to recommendations of 26 to 60 hours of supervised experience (Koegel, Russo, & Rincover, 1977; Smith, Buch, & Gamby, 2000; Smith, Parker, Taubman, & Lovaas, 1992). Teachers have been trained to fidelity in PRT in 8 to 20 hours (Suhrheinrich, 2011). To achieve concurrent mastery of several different intervention techniques and to incorporate the development of appropriate student goals, some researchers have suggested that teachers may need a year or more of full-time, supervised practicum training (Smith, Donahoe, & Davis, 2000).

There are several reasons why teachers may not implement evidence-based practices the way they were designed. First, teachers typically receive limited instruction in specific interventions. For example, instruction often comprises attendance at a didactic workshop and receipt of a manual. Teachers are then expected to implement evidence-based practices without the ongoing coaching and feedback that is critical for intervention mastery (Bush, 1984; Cornett & Knight, 2009). Second, most evidence-based practices were not designed for school settings and therefore may be difficult to implement appropriately in the classroom (Stahmer, Suhrheinrich, Reed, Bolduc, & Schreibman, 2011). Perhaps as a result, teachers often report that they combine or modify evidence-based practices to meet the specific needs of their classroom and students (Stahmer, Collings, & Palinkas, 2005). Finally, school administrators sometimes mandate the use of programs that may not align with teachers’ classroom environment, beliefs, or pedagogy (Dingfelder & Mandell, 2011).

A major indication of the quality of the implementation of any evidence-based practices is treatment fidelity, also known as implementation fidelity (Gersten et al., 2005; Horner et al., 2005; Noell, Duhon, Gatti, & Connell, 2002; Noell et al., 2005; Proctor et al., 2011; Schoenwald et al., 2011). Implementation fidelity is the degree to which a treatment is implemented as prescribed, or the level of adherence to the specific procedures of the intervention (e.g., Gresham, 1989; Rabin, Brownson, Haire-Joshu, Kreuter, & Weaver, 2008; Schoenwald et al., 2011). There are several types of implementation fidelity. Procedural fidelity (Odom et al., 2010; also called program adherence; Schoenwald et al., 2011) is the degree to which the provider uses procedures required to execute the treatment as intended. Other types of fidelity include treatment differentiation (the extent to which treatments differ from one another), therapist competence (the level of skill and judgment used in executing the treatment; Schoenwald et al., 2011), and dosage (Odom et al., 2010). Although, ideally, all types of fidelity would be examined to determine the fit of an intervention in a school program (Harn, Parisi, & Stoolmiller, 2013), procedural fidelity provides one important avenue for

measuring the extent to which an intervention resembles an evidence-based practice or elements of evidence-based practice (Garland, Bickman, & Chorpita, 2010).

Procedural implementation fidelity is likely a potential mediating variable affecting student outcomes, with higher fidelity resulting in better outcomes (Durlak & DuPre, 2008; Gresham, MacMilan, Beebe-Grankenberger, & Bocian, 2000; Stahmer & Gist, 2001); however, it is not often measured. In behavioral services research, three separate reviews of reported implementation fidelity data have been published. In the *Journal of Applied Behavior Analysis*, fidelity data were reported in only 16% to 30% of published articles (Gresham, Gansle, & Noell, 1993; McIntyre, Gresham, DiGennaro, & Reed, 2007; Peterson, Homer, & Wonderlich, 1982). Three separate reviews indicated that only 13% to 32% of autism intervention studies included fidelity measures (Odom & Wolery, 2003; Wheeler, Baggett, Fox, & Blevins, 2006; Wolery & Garfinkle, 2002). A recent review of special education journals found that fewer than half (47%) of intervention articles reported any type of fidelity scores (Swanson, Wanzek, Haring, Ciullo, & McCulley, 2011). Indeed, limited reporting of implementation adherence is evident across a diverse body of fields (Gresham, 2009). The lack of reporting (and therefore, the presumable lack of actual measurement of implementation) limits the conclusions that can be drawn regarding the association between student outcomes and the specific treatment provided. Therefore, examination of implementation fidelity, although complicated, is important to advance the understanding of how evidence-based interventions are being implemented in school settings.

Our research team recently completed a large-scale randomized trial of a comprehensive program for students with autism in partnership with a large, urban public school district. Procedural implementation fidelity of the overall program (which includes three evidence-based practices) was highly variable, ranging from 12% to 92% (Mandell et al., 2013). The three strategies included in this program, DTT, PRT, and FR (see description in the Method section), share an underlying theoretical base, but rely on different specific techniques. The purpose of this study was to examine the extent to which public school teachers implemented evidence-based interventions for students with autism in the way these practices were designed. Examining implementation fidelity of each strategy individually may provide insight into whether specific interventions are more easily implemented in the classroom environment. In particular, we examined whether special education classroom teachers and staff: 1) mastered specific strategies that form the backbone of applied behavioral analysis programs for autism; 2) used the strategies in their classroom; and 3) maintained their procedural fidelity to these strategies over time.

METHOD

Participants

Participants were classroom teachers and staff in an urban school district's kindergarten-through-second-grade autism support classrooms (each in a different school) participating in a larger trial of autism services. Of the 67 total autism support classrooms in the district at the time of the study, teachers and staff from 57 (85%) of the schools participated. Each classroom included one participating teacher and 0 to 2 classroom assistants ($M = 1$). Throughout the district, staff were required to participate in intervention training as part of professional development, but were not required to consent to participate in the study. Data from the current study are reported only for the 57 teachers and staff who consented to participate.

Teachers received intensive training in Strategies in Teaching Based on Autism Research (STAR) during their first year of participation in the project. During the second year, continuing teachers received in-classroom coaching every other week. From the original 57, 38 teachers (67%)

Table 1
Teacher Demographic Characteristics

<i>N</i>	% Female	Total Years Teaching, <i>M</i> (range)	Years Teaching Children with ASD, <i>M</i> (range)	Education Level % Bachelor's Degree/% Master's Degree
57	97.3	10.8 (1–38)	6.8 (1–33)	30/70

participated in the second year of the study. See Table 1 for teacher demographics. A complete description of adult and student participants can be found elsewhere (Mandell et al., 2013).

Intervention

Strategies for Teaching Based on Autism Research. The goal of the Strategies for Teaching Based on Autism Research (STAR) program is to develop children’s skills in a highly structured environment and then generalize those skills to more naturalistic settings. The program includes a curriculum in which each skill is matched to a specific instructional strategy. The STAR program includes three evidence-based strategies: DTT, PRT, and FR.

DTT relies on highly structured, teacher-directed, one-on-one interactions between the teacher and student. In these interactions, the teacher initiates a specific stimulus to evoke the child’s response, generally a discrete skill, which is an element of a larger behavioral repertoire (Krug et al., 1979; Krug, Rosenblum, Almond, & Arick, 1981; Lovaas, 1981, 1987; Smith, 2001). DTT is used in STAR for teaching pre-academic and receptive language skills, where the desired behavior takes a very specific form, such as learning to identify colors, sequencing events from a story into a first-next-then-last structure or counting with one-to-one correspondence. The consequence of the desired behavior is an external reinforcer, such as a token or a preferred edible (Lovaas, 2003; Lovaas & Buch, 1997).

PRT can occur in both one-on-one interactions and small-group interactions with the teacher. It is considered student directed because it occurs in the regular classroom environment, where the teaching area is pre-arranged to include highly preferred activities or toys that the student will be motivated to acquire. In PRT, students initiate the teaching episode by indicating interest in an item or activity or selecting among available teaching materials. Materials are varied frequently to enhance student motivation and generalization of skills and make PRT appropriate for targeting expressive and spontaneous language (Koegel, O’Dell, & Koegel, 1987; Koegel et al., 1989; Laski, Charlop, & Schreibman, 1988; Pierce & Schreibman, 1997; Schreibman & Koegel, 1996). After the student expresses interest in an activity or item, he or she is required to perform a specific behavior related to the item. The consequence of the desired behavior is getting access to the activity or item. For example, students’ attempts to label and request items are reinforced by the delivery of the item, which may then provide the opportunity to focus on other skills, such as joint attention, imitation, play skills, and generalization of other skills learned in the DTT format.

FR are the least structured of the STAR instructional strategies. FR strategies are routines that occur throughout the day and include school arrival and dismissal, mealtime, toileting, transitions between classroom activities, and recreational activities. Each routine is broken into discrete steps called a task analysis and then chained together using behavior analytic procedures such as stimulus prompts (visual and verbal) and reinforcement of each step in the routine (Brown et al., 1987; Cooper et al., 1987; Marcus et al., 2000; McClannahan & Krantz, 1999). For example, a routine to change activities may include cuing the transition (verbal prompt), checking a schedule (visual prompt), pulling a picture card from the schedule to indicate the next activity, taking the card to the location of

the new activity, putting the card into a pocket utilizing a match-to-sample technique, and beginning the new activity, followed by a token for routine completion. The advantage of this strategy is that each transition component is taught within the context of performing the routine, so that the child learns to respond to natural cues and reinforcers. FR strategies are conducted in both individual and group formats, depending on the skills being taught (e.g., toileting versus appropriate participation in snack time).

Training

STAR training occurred in accordance with the STAR developers' training protocols. The research team contracted with the program developers to provide training directly to the teachers. Training included workshops, help with classroom setup, and observation and coaching throughout the first academic year of STAR implementation (described in detail in the following sections). Six local coaches also were trained by the STAR developers to provide ongoing consultation to classroom staff during the second year of STAR implementation. The training protocol for STAR is manualized and publicly available. Additional information about the STAR program can be found at www.starautismsupport.com. Training provided to classroom teachers and staff included the following components:

Workshops. The STAR program developers provided a series of trainings on the use of the STAR program. The training began in September and consisted of 28 hours of intensive workshops that covered the STAR program, including the use of the curriculum assessment, classroom setup, and training in DTT, PRT, and FR. Workshops included didactic teaching, video examples, role-playing, and a visit to each classroom to help with classroom setup. STAR workshops took place outside the school day (i.e., during professional development days, at night, and on the weekends).

Observation and coaching. During the first year, program developers observed classroom staff during regular school hours and provided feedback on use of STAR strategies with students. Trainers provided 5 days of observation and coaching immediately following training, 3 days of follow-up coaching throughout the academic year, and ongoing advising and coaching by e-mail and phone. On average, classrooms received 26.5 (range, 1.5–36) hours of coaching over 5.7 (range, 3–7) visits in the first year. During the second year, local coaches trained by the STAR developers provided coaching in the STAR strategies. Coaching was provided September through May on a monthly basis. On average, classrooms received 36.1 (range, 0–59) hours of coaching over 10 (range, 0–10) visits in the second year.

Data Collection Procedures

Data on adherence to the instructional strategies used in STAR were collected throughout the academic year via video recording of teaching interactions with students for coding of implementation fidelity in each of the three STAR intervention methods.

Classroom staff members were filmed for 30 minutes every month in Years 1 and 2. Research assistants trained in filming methods recorded the intervention during a specified date each month. Visits were timed to coincide with regularly scheduled use of each of the intervention methods. The 30-minute film was composed of 10 minutes of DTT, 10 minutes of PRT, and 10 minutes of FR to provide a sample of the use of each intervention. Recording included any consented staff member providing the intervention. The staff member filmed by the research staff varied depending on which staff member (i.e., teacher or paraprofessional) was conducting the intervention that day. The primary classroom teacher conducted the intervention in 86% of the videos collected, and paraprofessional staff conducted the intervention in the remaining 14% of videos. There were no

statistically significant differences in the proportion of videos collected by intervention provider (teacher vs. paraprofessional) for any strategy or time period ($p > .05$).

Implementation Fidelity Measures

Coding procedures. The primary method for assessing fidelity of STAR strategies was through video recordings of teachers and aides interacting with students. Coding relied on different criteria based on specific coding definitions created for each instructional component, as well as general teaching strategies (see following sections). Coding schemes for each method were developed by the first author and were reviewed by the STAR program developers.

Trained research assistants blinded to the study hypotheses coded all video recordings. For each intervention method, the core research team established correct codes for a subset of videos through consensus coding (keys). Each research assistant coder then learned one coding system (i.e., DTT, PRT, or FR) and was required to achieve 80% reliability across two keys before beginning to code any classroom sessions independently. One third of all tapes were double coded to ensure ongoing reliability of data coding throughout the duration of the project. The core research team also re-coded two tapes for each research assistant every other month, providing a measure of criterion validity. If there was less than 80% agreement between the reliability coder and the research assistant, additional training and coaching were provided until criterion was achieved and previous videos were re-coded.

Coding involved direct computer entry while viewing videos using “The Observer Video-Pro” software (Noldus Information Technology, Inc., 2008), a computerized system for collection, analysis, and management of direct observation data. For each instructional strategy, the coder observed the 10-minute segment and subsequently rated the adults’ use of each component of the strategy on a 1 to 5 Likert scale, with 1 indicating *Adult does not implement throughout segment* and 5 indicating *Adult implements consistently throughout the segment*. These Likert ratings were found to have high concordance with more detailed trial-by-trial coding of each strategy component (88% agreement) used in previous research (Stahmer, 2010). A score of 4 or 5 on a component was considered passing and correlated with 80% correct use of strategies in the more detailed coding scheme. Following are the individual components included in each strategy. Complete coding definitions are available from the first author.

Discrete trial teaching. For DTT, coders examined the use of the following components: gaining the student’s attention, choosing appropriate target skills, using clear and appropriate cues, using accurate prompting strategies, providing clear and correct consequences, using appropriate inter-trial intervals, and utilizing error correction procedures effectively (error correction evaluated against procedures described in Arick, Loos, Falco, & Krug, 2004). The criterion for passing implementation fidelity was defined as the correct use of 80% of components (score of 4 or 5) during the observation.

Pivotal response training. For PRT, coders examined the use of the following components: gaining the student’s attention, providing clear and developmentally appropriate cues related to the activity, providing the student a choice of stimuli/activities, interspersing a mixture of maintenance (previously acquired) and acquisition (not yet mastered) tasks, taking turns to model appropriate behavior, providing contingent consequences, rewarding goal-directed attempts, and using reinforcers directly related to the teaching activity. The criterion for passing implementation fidelity was defined as the correct use of 80% of components (score of 4 or 5) during the observation.

Functional routines. For FR, coders examined adherence to each step of the FR used in classrooms during group and individual routines. The use of the following components was coded: using error correction procedures appropriately, adhering to FR lesson plan, and supporting transitions

between activities. The criterion for passing implementation fidelity was defined as correct use of 80% of components (score of 4 or 5) during the observation.

Reliability of Data Recording

Inter-rater reliability, as measured by percent agreement within 1 Likert point, was calculated for coding of each instructional strategy and each month of videos by having a second coder, blinded to the initial codes, score one third of the videos per strategy for each month. The average overall percent agreement for each strategy was 86% for DTT (range, 60%–100%); 90% for PRT (range, 75%–100%); and 90% for FR (range, 67%–100%). A primary coder was assigned to each strategy, and those codes were used in the analyses.

Data Reduction and Analyses

Data were examined across four periods. Time 1 included the first measurement for available classrooms in Year 1, which was conducted in October, November, or December of 2008. Filming occurred after the initial training workshops. Coaching was ongoing throughout the year. If classrooms were filmed in more than one of those months, both the average and the best performance were analyzed. All classroom staff participated in their initial training prior to the Time 1 measurement. Time 2 was defined as the performance from the last three measurements of the school year (February, March, or April 2009) for Year 1. The same procedures were used for Year 2 (Times 3 and 4). Time 3 included the first observation in Year 2 (October, November, or December 2009). Time 4 included the performance during the last 3 months of observations (February, March, or April, 2010). Both average and best performance from each period were utilized to provide an estimate of the staff's capacity to implement the strategy in the classroom environment (best) and variability in competency of use (average).

Data from Year 1 and Year 2 were analyzed. One-way within-subject (or repeated measures) analyses of variance (ANOVAs) were conducted for each intervention strategy to examine change in implementation fidelity scores for over time. Post-hoc comparisons were made using paired sample *t* tests between time periods when ANOVA results indicated statistically significant differences. In addition, we examined differences in fidelity of implementation across intervention strategies using a one-way ANOVA with paired sample *t* tests to follow up on significant results. Type I error probability was maintained at .05 (two-tailed) for all analyses using a Bonferroni correction.

Pearson correlations were conducted to examine the relationship between fidelity of implementation of each intervention strategy and teaching experience, experience working with children with autism spectrum disorder (ASD), level of education, and number of hours of coaching received.

RESULTS

Use of the Strategies

Because teachers who did not allow filming in their classrooms cited staffing difficulties or lack of preparation as the reason, they were considered not to be implementing DTT, PRT, or FR in their classrooms on a regular basis. At Time 1, two teachers (4%) explicitly indicated that they did not use DTT at any time, and 13 teachers (23%) indicated that did not use PRT at any time. The percentage of classrooms filmed using the strategy is displayed in Figure 1. In Year 1, classrooms were filmed most often conducting DTT at both Time 1 (70% of classrooms) and Time 2 (96%). Only 23% of classrooms were filmed conducting PRT at Time 1, and 68% were filmed at Time 2. FR was filmed in 67% of classrooms at Time 1 and 81% at Time 2. In Year 2, filming was much more consistent across strategies. DTT and PRT were both filmed in 92% of classrooms at Time 3

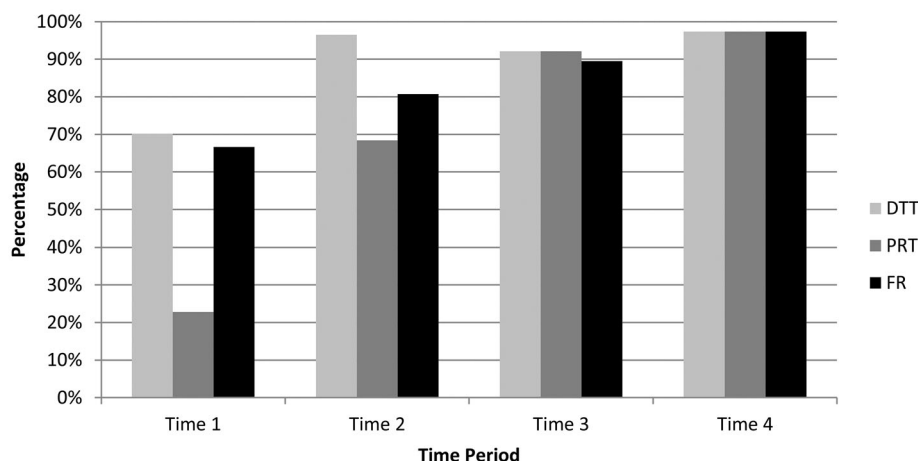


FIGURE 1. The percentage of classrooms using the strategy during each time period.

and 97% of classrooms at Time 4. For FR, 89% of classrooms were filmed at Time 3 and 97% at Time 4.

Overall Competence in the Instructional Strategies

Discrete trial training. The percentage of DTT components on which teachers met fidelity (i.e., a score of 4 or 5 during the observation) was used as the dependent variable for these analyses. Mean results are displayed in Table 2. No statistically significant changes over time were found in average or best DTT fidelity over time. In general, classrooms had a relatively high average and best DTT fidelity during all time periods. The range of scores for individual performance was variable at both time periods, as evidenced by the large standard deviations.

The percentage of classrooms in which teachers met DTT fidelity (i.e., correct implementation of 80% of the DTT strategies during the observation) was examined. Fifty-six percent of classrooms met fidelity at Time 1 based on the average of all observations at Time 1, 47% at Time 2, 46% at Time 3, and 59% at Time 4. When considering only the best example, 65% of classrooms met fidelity at Time 1, and this increased to 81% by Time 4 (see Figure 2).

Pivotal response training. The dependent variable for these analyses was the percentage of PRT components on which teachers met fidelity (i.e., a score of 4 or 5 during the observation). Mean results are displayed in Table 2. No statistically significant changes were found in average PRT fidelity over time. There was a statistically significant increase in best scores over time, $F(3, 108) = 2.85, p = .04$. In pairwise comparisons, only the difference in best scores between Time 1 and Time 4 was statistically significant, $t(9) = -2.45, p = .04$. The range of scores for individual performance was variable at both time periods, as evidenced by the large standard deviations.

The percentage of classrooms in which teachers met PRT fidelity was examined (i.e., correct implementation of at least 80% of PRT components during the observation). For average performance, only 15% of classrooms met fidelity at Time 1, 31% at Time 2, 11% at Time 3, and 19% at Time 4. When examining best performance at each time period, 23% of classrooms met fidelity at Time 1, 41% at Time 2, 17% at Time 3, and 30% at Time 4 (see Figure 2).

Table 2
Mean Fidelity of Implementation by Time and Intervention Strategy for Average and Best Fidelity^a

Time	Intervention			
	Discrete Trial Teaching <i>M (SD)</i>	Pivotal Response Training <i>M (SD)</i>	Functional Routines <i>M (SD)</i>	Overall Fidelity <i>M (SD)</i>
Average Fidelity across All Assessments During Time Period (%)				
Time 1	78.54 (24.33)	53.41 (24.09)	56.43 (16.42)	65.14 (16.47)
Time 2	73.94 (21.16)	58.43 (26.66)	69.77 (19.05)	68.45 (15.39)
Time 3	71.04 (27.79)	68.39 (20.25)	75.56 (24.17)	71.66 (20.01)
Time 4	80.46 (17.55)	60.19 (21.39)	78.51 (19.80)	73.58 (12.98)
Best Fidelity for Each Time Period (%)				
Time 1	81.64 (24.93)	54.64 (25.60)	63.53 (20.38)	69.86 (18.00)
Time 2	84.53 (19.77)	65.22 (23.38)	79.96 (21.33)	77.72 (16.28)
Time 3	79.21 (26.94)	73.78 (21.21)	81.59 (23.78)	81.33 (11.19)
Time 4	90.74 (13.00)	74.16 (21.96)	91.45 (16.50)	85.70 (11.19)

^aFidelity of implementation is defined as the percentage of strategy components implemented correctly.

Teaching in FR. The percentage of FR components on which teachers met fidelity was used as the dependent variable for these analyses. Mean results are displayed in Table 2. Statistically significant changes over time were found in average FR fidelity, $F(3, 154) = 9.11, p = .00$ and best FR fidelity, $F(3, 155) = 12.13, p = .00$. The range of scores for individual performance was variable at both time periods, as evidenced by the large standard deviations. Statistically significant increase were seen between Time 1 and each of the other time periods, both for average fidelity (Time 2: $t = -3.71, p < .00$; Time 3: $t = -3.70, p = .00$; Time 4: $t = -6.14, p = .00$), and best fidelity (Time 2: $t = -3.83, p < .00$; Time 3: $t = -3.28, p = .00$; Time 4: $t = -6.93, p = .00$).

The percentage of classrooms in which teachers met FR fidelity was examined (i.e., correct implementation of 80% FR strategies during the observation). For average performance, 11% of classrooms met fidelity at Time 1, 34% at Time 2, 62% at Time 3, and 49% at Time 4. For best performance, 16% met fidelity at Time 1, and 78% met fidelity by Time 4 (see Figure 2).

Overall fidelity. Overall fidelity across the STAR program was examined by averaging the percentage of components implemented correctly in each strategy (DT, PRT, and FR; Table 1). No significant changes over time were seen in the average overall fidelity. However, significant increases in best overall fidelity were indicated, $F(3, 178) = 8.14, p = .00$. Post-hoc analyses indicated that best fidelity at Time 1 was significantly lower than at any of the other time periods (Time 2: $t = -2.72, p < .01$; Time 3: $t = -4.14, p = .00$; Time 4: $t = -5.03, p = .00$). The range of scores for individual performance was variable at both time periods, as evidenced by the large standard deviations.

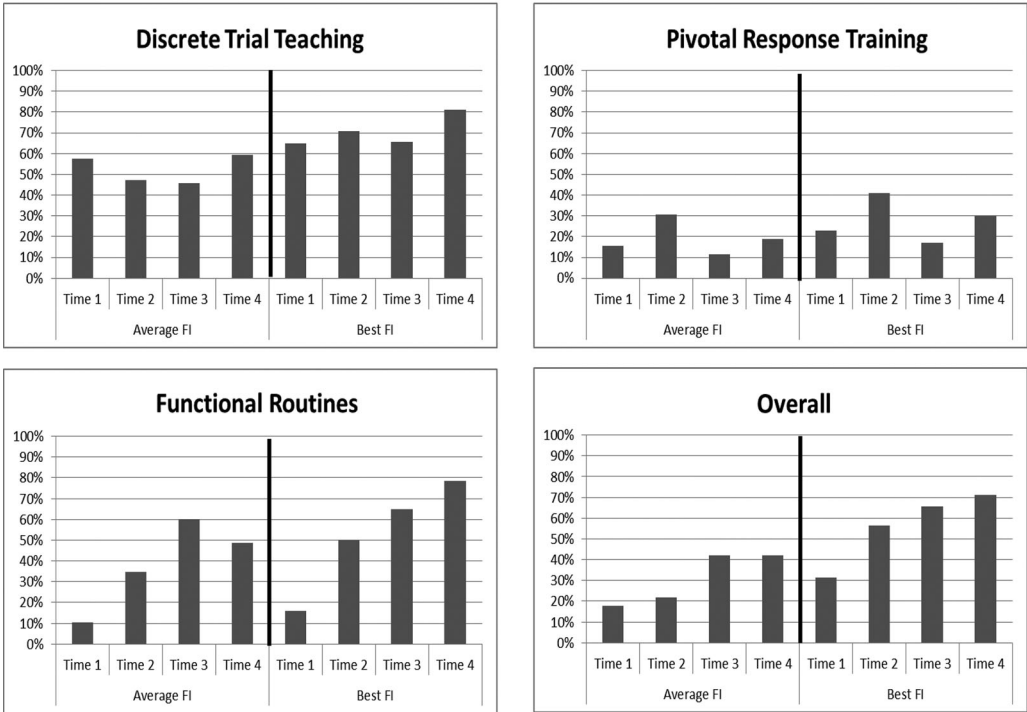


FIGURE 2. Percentage of classrooms meeting 80% implementation fidelity during each time period. FI = fidelity implementation.

The percentage of classrooms meeting overall fidelity at each time period (i.e., correctly implementing at least 80% of components in all three interventions) was examined. For average performance, 17% of classrooms met fidelity at Time 1, 22% at Time 2, and 42% at both Time 3 and Time 4. For best performance, 31% met fidelity at Time 1, and 71% met fidelity by Time 4 (Figure 2).

Comparison of Intervention Fidelity across Intervention Strategies

Mean fidelity of implementation was compared across the three intervention strategies for average and best fidelity. Significant differences in average, $F(109, 326) = 13.06, p \leq .00$, and best overall fidelity were indicated, $F(110, 327) = 3.26, p \leq .001$ (means are presented in Table 2). Analyses indicated that DTT average and best fidelity were significantly greater than were PRT average and best fidelity at Time 2 (average: $t = 4.03, p \leq .00$; best: $t = 5.14, p \leq .00$) and Time 4 (average: $t = -5.46, p \leq .00$; best: $t = -4.31, p \leq .00$). FR average and best fidelity were also significantly greater than were PRT average and best fidelity (average: $t = 5.46, p \leq .00$; best: $t = 4.31, p \leq .00$) at Time 4.

Associations between Intervention Fidelity and Experience, Education, or Coaching

Pearson correlations indicated there was no statistically significant association between the number of years of either teaching or children with autism and overall fidelity or fidelity on any specific intervention strategy at any time point. The number of hours of coaching received was not associated with overall fidelity.

DISCUSSION

These results from one of the first field trials of evidence-based practices for students with autism in public schools suggest that classrooms vary greatly in their implementation of evidence-based practices. In general, the data suggest that the complexity and structure of the intervention strategy may affect intervention use and procedural fidelity; more structured methods were more likely to be implemented with higher fidelity than were less structured strategies. Procedural fidelity continued to increase through the second year of training, suggesting the importance of continued practice for extended periods. It is important to note that the number of hours of coaching was not associated with final fidelity, suggesting that *in vivo* support may be important, but it is not sufficient to improve practice in the field.

Classrooms implemented DTT more often in Year 1 and with greater fidelity across both years than PRT or FR. The curriculum materials and steps for implementing DTT are clearly specified, highly structured, and relatively easy to follow. Components are, in general, scripted, straightforward, and with the exception of determining appropriate prompting levels, leave little room for clinical judgment.

In contrast, PRT is a more naturalistic strategy, and several of the components require clinical judgment on the part of the adult. Teachers had, in general, significantly greater difficulty implementing PRT with fidelity than either DTT or FR. During Year 1, many teachers did not implement PRT at all. By Year 2, although they were implementing the strategy, few were doing so with high fidelity. Both average and best fidelity scores across teachers are lower for PRT than either DTT or FR. Teachers may require additional time to develop and integrate these intervention strategies into the school day. It is possible that teachers have difficulty with specific components of PRT that are not well suited to the classroom environment. Recent data indicate that teachers may consistently leave out some components of PRT, which would reduce overall implementation fidelity of the comprehensive model (Suhrheinrich et al., 2013). How these adaptations affect the effectiveness of this intervention is not yet known.

FR strategies use many of the procedures of PRT in a group format, but have a specified set of goals and procedures. By the end of Year 2, procedural fidelity was greater for FR than PRT. This may indicate that the structure of the FR, including specific steps and goals, may assist with appropriate implementation of the naturalistic strategies. It may also be helpful that the STAR program uses FR strategies for activities that occur every day (e.g., snack time, toileting), providing consistent opportunities to implement the strategy independent of the classroom's schedule or structure.

Relatively high variability across classrooms and over time within classrooms was evident for both use of strategies (as measured by percentage of classrooms filmed) and implementation fidelity. It could be that classroom staff used the strategies with a different child each time they were filmed. Some students may present with behavior challenges that make the use of a particular intervention difficult. Variability in daily staffing, school activities, and student needs may affect the use of intervention strategies on any given day. It is also possible that staff characteristics, such as motivation to implement the intervention, experience, education, and training may affect how well they can use certain methods. Maintenance of all strategies may be difficult, as suggested by the decrease in fidelity at Time 3 (after summer break).

Limitations

There are several limitations to this study. First, implementation fidelity was examined during brief time periods each month. These data may provide only limited insight into whether strategies were well integrated into the daily classroom routine or used consistently over time or with a majority of students in the classroom. Second, the way fidelity was rated was relatively general and

may not have captured important aspects of the implementation that could affect student progress. Understanding the active ingredients of effective intervention and how to accurately measure those strategies is an area of growth for the field. Third, adults in the classroom knew they were being observed, and this may have altered their use of the strategies. Both the second and third limitations would lead to an overestimate of fidelity. Still, fidelity was relatively low across the three strategies. Strategies may have only been implemented on observation days or may have been implemented differently (better or worse fidelity) during the observations. Fourth, the use of filming as a proxy for use in the classroom has not been validated. In addition, for some observations, paraprofessionals rather than classroom teachers implemented the strategies. A closer examination of differences by profession may be warranted.

CONCLUSIONS

Results of this study indicate that teachers and staff in public school special education classrooms can learn to implement structured strategies that are the foundation of many autism intervention programs; however, they require a great deal of training, coaching, and time to reach and maintain implementation fidelity. A recent study indicates that ongoing classroom coaching can result in the use of important classroom practices, such as ongoing progress monitoring (Pellecchia et al., 2010). Even with ongoing support, however, not all staff will implement interventions with high fidelity. Highly structured strategies appear to be easier to learn, such that practice and coaching may be consistently required for teachers to use more naturalistic strategies with high fidelity. Naturalistic strategies may require additional training or adaptation for classroom environments. Some recent preliminary data indicate that teachers may be better able to implement a classroom-adapted version of PRT (Stahmer, Suhrheinrich, Reed, & Schreibman, 2012). Providers who achieve mastery of intervention strategies are likely to lose those skills or the motivation to use those skills over breaks from teaching; thus, ongoing consultation well past the initial didactic training is likely needed to maintain mastery. The same training and consultation strategy was used for all three practices, but with highly different results. These differential results may be related to the intervention itself or to the fit of the training and consultation model to the specific intervention, teacher, and context.

Future Research

High-quality implementation of evidence-based practices for children with autism in schools is essential for ensuring the best outcomes for this growing population of children. However, research in this area is just beginning to address the complexity of serving children with ASD using comprehensive and complex methods. The development of low-cost, accurate fidelity of implementation measurement is important for helping to ensure that teachers are accurately using evidence-based interventions. In addition, future research should address the development of training methods for naturalistic strategies that address the complexities of using these strategies in classroom settings. Integrating these strategies throughout the school day and for academic tasks can be challenging; yet, they are considered a very effective practice for children with autism. Often, paraprofessional staff spend a great deal of time working with children in the classroom. Specifically examining training needs and fidelity of implementation of paraprofessional staff compared with teachers and other professionals is needed. In addition, there are multiple interventions for ASD that are “branded” by various research groups. Often, the specific techniques or strategies overlap significantly. Research examining the key ingredients necessary for effective classroom intervention is sorely needed. This has the potential to simplify and clarify intervention for use by teachers and other community providers.

REFERENCES

- Arick, J. R., Loos, L., Falco, R., & Krug, D. A. (2004). The STAR program: Strategies for teaching based on autism research. Austin, TX: Pro-Ed.
- Brown, F., Evans, I., Weed, K., & Owen, V. (1987). Delineating functional competencies: A component model. *Journal of the Association for Persons with Severe Handicaps*, 12, 117–124.
- Bush, R. N. (1984). *Effective staff development*. San Francisco, CA: Far West Laboratory for Educational Research and Development.
- Cooper, J. O., Heron, T. E., & Heward, W. L. (1987). *Applied behavioral analysis*. New York, NY: Macmillan.
- Cornett, J., & Knight, J. (2009). Research on coaching. In J. Knight (Ed.), *Coaching: Approaches and perspectives* (pp. 192–216). Thousand Oaks, CA: Corwin Press.
- Dingfelder, H. E., Mandell, D. S., & Marcus, S. C. (2011, May). Classroom climate program fidelity & outcomes for students with autism. Paper presented at the 10th annual International Meeting for Autism Research. San Diego, CA.
- Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41, 327–350.
- Fenske, E., Krantz, P. J., & McClannahan, L. E. (2001). Incidental teaching: A non-discrete trial teaching procedure. In C. Maurice, G. Green, & R. Foxx (Eds.), *Making a difference: Behavioral intervention for autism* (pp. 75–82). Austin, TX: Pro-Ed.
- Garland, A. F., Bickman, L., & Chorpita, B. F. (2010). Change what? Identifying quality improvements targets by investigating usual mental health care. *Administration and Policy in Mental Health and Mental Health Services Research*, 37, 15–26.
- Gersten, R., Fuchs, L., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children*, 71, 149–164.
- Gresham, F. M. (1989). Assessment of treatment integrity in school consultation and prereferral intervention. *School Psychology Review*, 18, 37–50.
- Gresham, F. M. (2009). Evolution of the treatment integrity concept: Current status and future directions. *School Psychology Review*, 38, 533–540.
- Gresham, F. M., Gansle, K. A., & Noell, G. H. (1993). Treatment integrity in applied behavior analysis with children. *Journal of Applied Behavior Analysis*, 26, 257–263.
- Gresham, F. M., MacMillan, D. L., Beebe-Grankenberger, M. E., & Bocian, K. M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented? *Learning Disabilities Research and Practice*, 15, 198–125.
- Harn, B., Parisi, D., & Stoolmiller, M. (2013). Balancing fidelity with flexibility and fit: What do we really know about fidelity of implementation in schools? *Exceptional Children*, 79, 181–193.
- Hess, K. L., Morrier, M. J., Heflin, L. J., & Ivey, M. L. (2008). Autism treatment survey: Services received by children with autism spectrum disorders in public school classrooms. *Journal of Autism and Developmental Disorders*, 38, 961–971.
- Horner, R. H., Carr, E. G., Halle, J., McGee, G. G., Odom, S. L., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, 71, 165–179.
- Jennett, H. K., Harris, S. L., & Mesibov, G. B. (2003). Commitment to philosophy, teacher efficacy, and burnout among teachers of children with autism. *Journal of Autism and Developmental Disorders*, 33, 583–593.
- Koegel, R. L., O'Dell, M. C., & Koegel, L. K. (1987). A natural language teaching paradigm for nonverbal autistic children. *Journal of Autism & Developmental Disorders*, 17, 187–200.
- Koegel, R. L., Russo, D. C., & Rincover, A. (1977). Assessing and training teachers in the generalized use of behavior modification with autistic children. *Journal of Applied Behavior Analysis*, 10, 197–205.
- Koegel, R. L., Schreibman, L., Good, A., Cerniglia, L., Murphy, C., & Koegel, L. K. (Eds.). (1989). *How to teach pivotal behaviors to children with autism: A training manual*. Santa Barbara: University of California–San Diego.
- Krug, D. A., Arick, J., Almond, P., Rosenblum, J., Scanlon, C., & Border, M. (1979). Evaluation of a program of systematic instructional procedures for pre-verbal autistic children. *Improving Human Performance*, 8, 29–41.
- Krug, D. A., Rosenblum, J. F., Almond, P. J., & Arick, J. R. (1981). *Autistic and severely handicapped in the classroom: Assessment, behavior management, and communication training*. Portland, OR: ASIEP Education.
- Laski, K. E., Charlop, M. H., & Schreibman, L. (1988). Training parents to use the natural language paradigm to increase their autistic children's speech. *Journal of Applied Behavior Analysis*, 21, 391–400.
- Leaf, R. B., & McEachin, J. J. (1999). *A work in progress: Behavior management strategies and a curriculum for intensive behavioral treatment of autism*. New York, NY: DRL Books.
- Lovaas, O. I. (1981). *Teaching developmentally disabled children: Theme book*. Austin, TX: PRO-ED.
- Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual functioning of young autistic children. *Journal of Consulting and Clinical Psychology*, 55, 3–9.
- Lovaas, O. I. (2003). *Teaching individuals with developmental delays: Basic intervention techniques*. Austin, TX: Pro-Ed.

- Lovaas, O. I., & Buch, G. (1997). Intensive behavioral intervention with young children with autism. In N. Singh (Ed.), *Prevention and treatment of severe behavior problems: Models and methods in developmental disabilities* (pp. 61–86). Pacific Grove, CA: Brooks/Cole Publishing.
- Mandell, D. S., Stahmer, A. C., Shin, S., Xie, M., Reisinger, E., & Marcus, S. C. (2013). The role of treatment fidelity on outcomes during a randomized field trial of an autism intervention. *Autism*, 17, 281–295.
- Marcus, L., Schopler, E., & Lord, C. (2000). TEACCH services for preschool children. In J. S. Handleman & S. L. Harris (Eds.), *Preschool education programs for children with autism* (pp. 215–232). Austin, TX: Pro-ED.
- McClannahan, L. E., & Krantz, P. J. (1999). *Activity schedules for children with autism: Teaching independent behavior*. Bethesda, MD: Woodbine House.
- 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 Behavioral Analysis*, 40, 659–972.
- National Standards Project. (2009). *National Standards report*. Randolph, MA: National Autism Center.
- Noell, G. H., Duhon, G. J., Gatti, S. L., & Connell, J. E. (2002). Consultation, follow-up and implementation of behavior management interventions in general education. *School Psychology Review*, 31, 217–234.
- Noell, G. H., Witt, J. C., Slider, N. J., Connel, J. E., Williams, K. L., Resetar, J. L., & Koenig, J. L. (2005). Teacher implementation following consultation in child behavior therapy: A comparison of three follow-up strategies. *School Psychology Review*, 37, 87–106.
- Noldus Information Technology, Inc. (2008). *The Observer XT 8.0* [computer software]. Wageningen, The Netherlands.
- Odom, S. L., Boyd, B. A., Hall, L. J., & Hume, K. (2010). Evaluation of comprehensive treatment models for individuals with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 40, 425–436.
- Odom, S. L., & Wolery, M. (2003). A unified theory of practice in early intervention/early childhood special education: Evidence-based practices. *Journal of Special Education*, 37, 164–173.
- Pellecchia, M., Connell, J. E., Eisenhart, D., Kane, M., Schoener, C., Turkel, K., & Mandell, D. S. (2010). Group performance feedback: Consultation to increase classroom team data collection. *Journal of School Psychology*, 49, 411–431.
- Peterson, L., Homer, A. L., & Wonderlich, S. A. (1982). The integrity of independent variables in behavior analysis. *Journal of Applied Behavior Analysis*, 15, 477–492.
- Pierce, K., & Schreibman, L. (1997). Multiple peer use of pivotal response training to increase social behaviors of classmates with autism: Results from trained and untrained peers. *Journal of Applied Behavior Analysis*, 30, 157–160.
- Proctor, E., Simere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., & Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research*, 38, 65–76.
- Rabin, B. A., Brownson, R. C., Haire-Joshu, D., Kreuter, M. W., & Weaver, N. L. (2008). A glossary for dissemination and implementation research in health. *Journal of Public Health Management and Practice*, 14, 117–123.
- Sarokoff, R. A., & Sturmey, P. (2004). The effects of behavioral skills training on staff implementation of discrete-trial teaching. *Journal of Applied Behavior Analysis*, 37, 535–538.
- Schoenwald, S. K., Garland, A. F., Chapman, J. E., Frazier, S. L., Sheidow, A. J., & Southam-Gerow, M. A. (2011). Toward the effective and efficient measurement of implementation fidelity. *Administration and Policy in Mental Health and Mental Health Services Research*, 38, 32–43.
- Schreibman, L., & Koegel, R. L. (1996). Fostering self-management: Parent-delivered pivotal response training for children with autistic disorder. In E. D. Hibbs & P. S. Jensen (Eds.), *Psychosocial treatments for child and adolescent disorders: Empirically based strategies for clinical practice* (pp. 525–552). Washington, DC: American Psychological Association.
- Scull, J., & Winkler, A. M. (2011). *Shifting trends in special education*. Washington, DC: Thomas B. Fordham Institute.
- Simpson, R. L., de Boer-Ott, S. R., & Smith-Myles, B. (2003). Inclusion of learners with autism spectrum disorders in general education settings. *Topics in Language Disorders*, 23, 116–133.
- Sindelar, P. T., Brownell, M. T., & Billingsley, B. (2010). Special education teacher education research: Current status and future directions. *The Journal of the Teacher Education Division of the Council for Exceptional Children*, 33, 8–24.
- Smith, T. (2001). Discrete trial training in the treatment of autism. *Focus on Autism and Other Developmental Disabilities*, 16, 86–92.
- Smith, T., Buch, G. A., & Gamby, T. E. (2000). Parent-directed, intensive early intervention for children with pervasive developmental disorder. *Research in Developmental Disabilities*, 21, 297–309.
- Smith, T., Donahoe, P. A., & Davis, B. J. (2000). *The UCLA young autism project*. Austin, TX: Pro-Ed.
- Smith, T., Parker, T., Taubman, M., & Lovaas, O. I. (1992). Transfer of staff training from workshops to group homes: A failure to generalize across settings. *Research in Developmental Disabilities*, 13, 57–71.
- Stahmer, A. (2007). The basic structure of community early intervention programs for children with autism: Provider descriptions. *Journal of Autism and Developmental Disorders*, 37, 1344–1354.
- Stahmer, A. (2010). *Examining methods of fidelity of implementation measurement*. Unpublished raw data.

- Stahmer, A., Collings, N. M., & Palinkas, L. A. (2005). Early intervention practices for children with autism: Descriptions from community providers. *Focus on Autism & Other Developmental Disabilities*, 20, 66–79.
- Stahmer, A., & Gist, K. (2001). The effects of an accelerated parent education program on technique mastery and child outcome. *Journal of Positive Behavior Interventions*, 3, 75–82.
- Stahmer, A., & Ingersoll, B. (2004). Inclusive programming for toddlers with autism spectrum disorders: Outcomes from the Children's Toddler School. *Journal of Positive Behavior Interventions*, 6, 67–82.
- Stahmer, A., Suhrheinrich, J., Reed, S., Bolduc, C., & Schreibman, L. (2011). Classroom pivotal response teaching: A guide to effective implementation. New York, NY: Guilford Press.
- Stahmer, A. C., Suhrheinrich, J., Reed, S., & Schreibman, L. (2012). What works for you? Using teacher feedback to inform adaptations of pivotal response training for classroom use. *Autism Research and Treatment*, 2012, 1–11.
- Suhrheinrich, J. (2011). Training teachers to use pivotal response training with children with autism; Coaching as a critical component. *Teacher Education and Special Education*, 34, 339–349.
- Suhrheinrich, J., Stahmer, A., Reed, S., Schreibman, L., Reisinger, E., & Mandell, D. (2013). Implementation challenges in translating pivotal response training into community settings. *Journal of Autism and Developmental Disorders*, 43, 2970–2976.
- Swanson, E., Wanzek, J., Haring, C., Ciullo, S., & McCulley, L. (2011). Intervention fidelity in special and general education research journals. *The Journal of Special Education*, 47, 13–33.
- Wheeler, J. J., Baggett, B. A., Fox, J., & Blevins, L. (2006). Treatment integrity: A review of intervention studies conducted with children with autism. *Focus on Autism and Other Developmental Disabilities*, 21, 1–10.
- Wolery, M., & Garfinkle, A. N. (2002). Measures in intervention research with young children who have autism. *Journal of Autism and Developmental Disorders*, 32, 463–478.

Copyright of Psychology in the Schools is the property of John Wiley & Sons, Inc. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.