Article

HAMMILL INSTITUTE ON DISABILITIES

Topics in Early Childhood Special Education
2019, Vol. 39(1) 56–67

DOI: 10.1177/0271121419834343
tecse.sagepub.com
SAGE

Fidelity of a Teacher-Implemented Intervention for Preschoolers With Autism Spectrum Disorder: No, Some, and Unexpected Effects

Stephanie S. Reszka, PhD[1], Katie Belardi, PhD[2], Jessica Amsbary, MS[1], Brian A. Boyd, EdD[3], and Linda R. Watson, PhD[1]

Abstract

Advancing Social-Communication and Play (ASAP) is a classroom-based intervention targeting the social-communication and play development of preschool-age children with autism spectrum disorder. We used a multimethod approach to measuring intervention fidelity, including adherence, exposure, treatment differentiation, and quality of treatment delivery. Overall, teachers in the ASAP classrooms increased their adherence to the ASAP intervention compared with business-as-usual teachers. Teacher characteristics did not affect the ability to implement ASAP with fidelity. There were no group differences in coded video instructional quality scores. Differences in child engagement between the ASAP and business-as-usual groups were not mediated by fidelity adherence scores, contrary to our hypothesis. Finally, end-of-year summary of treatment delivery ratings did not align with adherence scores. Continued research regarding fidelity of school-based interventions should focus on distinguishing active ingredients from general instructional practices and learning from teachers about the child-level factors they consider when implementing an intervention.

Keywords

fidelity, ASAP, school interventions, autism spectrum disorder

Introduction

The Individuals with Disabilities Education Act (2004) guarantees children with disabilities, including autism spectrum disorder (ASD), between the ages of 3 and 21 the right to public education. Many children with ASD attend preschool programs, making the classroom an ideal context, and educators ideal implementers, for delivering interventions that target the core deficits of children with ASD. Early intervention in these core areas, including social-communication and play, is critical for improving later language skills and academic success (Estes et al., 2015). Although there are documented evidence-based practices for preschool children with ASD (e.g., Early Start Denver Model, Learning Experiences and Alternative Program for Preschoolers and their Parents [LEAP], Treatment and Education of Autistic and related Communication Handicapped Children [TEACCH], etc.; Wong et al., 2015), there are a limited number of efficacious teacher-implemented and school-based interventions for children with ASD (Kasari & Smith, 2013) and even fewer documenting the procedures for measuring how well the intervention was implemented (Odom, Boyd, Hall, & Hume, 2010). Our team took a first step in addressing this need by developing a classroom-based, team-implemented intervention for preschoolers with ASD, evaluating its efficacy, and including fidelity of implementation measures.

Advancing Social-Communication and Play (ASAP)

ASAP is a classroom-based intervention package for preschool-age children with ASD based on a developmental

[1]The University of North Carolina at Chapel Hill, USA
[2]Duquesne University, Pittsburgh, PA, USA
[3]The University of Kansas, Lawrence, USA

Corresponding Author:
Stephanie S. Reszka, Department of Allied Health Sciences, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7190, USA.
Email: sreszka@unc.edu

sequence of skills designed to improve the pivotal skills of social-communication and play that are related to later language outcomes. The ASAP intervention includes a systematic evidence-based curriculum (Dykstra, Boyd, Watson, Crais, & Baranek, 2012) designed to supplement the existing classroom curriculum to ensure that these key areas are addressed across the varied classroom models in which children with ASD are served (see Boyd et al., 2018 for more about ASAP and related research). ASAP is implemented in both 1:1 and group contexts within the classroom by classroom teaching teams, with support from an intervention coach.

Intervention educational teams were given ASAP manuals, a training DVD, and an ASAP goal hierarchy poster to display in the classroom at the time of the first training. In addition to this didactic training, coaches conducted bimonthly classroom observations that were discussed during in-person monthly team meetings. During these meetings, coaches and classroom teams developed action plans with ASAP goals to target over the next month based on progress and challenges in the classroom. Coaches also were available by phone or email to answer questions and to help classroom teams incorporate ASAP into their classroom practices and to implement the intervention. Incorporating evidence-based practices into teaching is important, but to maximize outcomes, it is critical that adopters implement the intervention with fidelity (Odom et al., 2010).

Fidelity of Implementation

Fidelity of implementation is the degree to which consumers adhere to delivering the critical components of a manualized intervention in the way it was designed (Century, Rudnick, & Freeman, 2010). Recently, funding agencies have required intervention studies to include measures of treatment fidelity (Institute of Education Sciences [IES], 2018; O'Donnell, 2008) in an effort to improve treatment efficacy. Fidelity of implementation is important for several reasons (Hume et al., 2011), including determining whether the treatment effects are valid (Gresham, MacMillan, Beebe-Frankenberger, & Bocian, 2000; Mandell et al., 2013) and whether the treatment is transferrable to the real world (Strain & Bovey, 2008). In conceptualizing intervention fidelity, there are a number of dimensions to consider, including (a) adherence/integrity: implementing an intervention as intended, usually prescribed in a manual; (b) exposure/dosage: frequency of delivery; (c) quality of treatment delivery; and (d) treatment differentiation: to ensure that treatment recipients are the only ones receiving the intervention.

Challenges to Measuring Fidelity

A simple explanation for the frequent lack of intervention fidelity measures is that fidelity is a difficult construct to measure (Howlin, Moss, Savage, & Rutter, 2013). The first challenge to measuring fidelity is determining what elements of the intervention to measure and *how* to measure those elements. Mowbray, Holter, Teague, and Bybee (2003) suggest that fidelity measurement begins by identifying both structural and instructional ingredients in a treatment, establishing operational definitions and determining the best method of measurement. Previous research supports a multimethod approach to fidelity, capturing both the quality of the teacher's implementation and a quantitative view of the teacher's adherence to the delivery process (Dane & Schneider, 1998; Odom et al., 2010; Power et al., 2005; Sutherland, McLeod, Conroy, & Cox, 2013) with reliable and valid measures (Smith, Daunic, & Taylor, 2007). Including a number of different methods and multiple data collection points requires extra time and resources to complete data collection related to fidelity of implementation.

Finally, there are notable challenges that may affect fidelity measurement inherent to interventions conducted in the school setting. School-based interventions are susceptible to threats to internal validity including diffusion or contamination of treatment, especially in cases in which educator interactions can influence the outcomes (Trochim & Donnelly, 2001). In addition, it can be more difficult to measure intervention dosage and quality within classroom contexts because intervention implementation is often infused into everyday routines and activities.

Purpose of Study

The purpose of this manuscript is to (a) briefly describe our multimethod approach to fidelity measurement within a 4-year cluster randomized trial (CRT) of a teacher-implemented intervention for preschoolers with ASD and (b) report findings related to fidelity, child and teacher predictors of fidelity, and fidelity as a mediator of child outcomes. This information has the potential to inform our understanding of the sensitivity of a multimethod fidelity approach and challenges and solutions to developing these types of measures. The following research questions were addressed:

Research Question 1: Does a multimethod approach measure fidelity to ASAP and differentiate ASAP from business-as-usual (BAU) classrooms?
Research Question 2: Are there student or teacher characteristics that moderate intervention fidelity?
Research Question 3: Does ASAP treatment fidelity mediate student outcomes?

Method

Design

This IES-funded Goal 3 efficacy trial, a multisite CRT, was conducted in public preschool classrooms in Florida, North

Carolina, Minnesota, Oregon, and Washington. Classrooms were randomly assigned, by a statistician who was not involved in recruitment and using a computerized randomization process, to the ASAP intervention or BAU control condition. Classroom teams in the BAU condition continued with their standard practice and received no training or coaching from ASAP staff. Randomization was at the classroom level, and there could be up to two classrooms from the same school enrolled. Educational teams were not made aware of assignment until completion of enrolled children's baseline data collection. A total of 78 classrooms, including 161 children with ASD, enrolled in this study. Of these classrooms, 40 received the ASAP intervention (85 children), and 38 classrooms served as control classrooms (76 children). Over the course of 4 years, only three classrooms including six children dropped out, either due to the participating child moving to another school or the teacher declining further participation. See Boyd et al. (2018) for the published Consolidated Standards of Reporting Trials (CONSORT) flow diagram.

Participant Eligibility

To be eligible to enroll in the ASAP study, the preschool educational team had to consent to implement ASAP together. Each educational team included a lead teacher, at least one teaching assistant, and at least one related service provider (RSP; usually a Speech-Language Pathologist or Occupational Therapist). In addition, the classroom's lead teacher must have taught for at least 1 year prior to study participation, and all teachers and RSPs must have been licensed to practice or be under the supervision of a licensed professional. Classrooms followed a variety of preschool classroom models, including inclusive/self-contained, with a variety of curricula (e.g., TEACCH, Strategies for Teaching based on Autism Research [STAR], Applied Behavior Analysis [ABA], etc.), and half/full day. As an incentive, participating ASAP teachers received US$200 in classroom educational supplies at the start of the school year, and participating RSPs received US$100 in therapy supplies (or US$50 each if there were two RSPs for the classroom). Participating BAU teachers received US$275 in educational supplies, with RSPs receiving US$25 in supplies; however, half of the supplies were provided at the start of the school year and half at the end to promote study retention. In appreciation for their participation, the BAU classroom teams were provided copies of the ASAP manual at the end of the school year. There were no significant pretreatment differences in classroom quality based on the Professional Development in Autism Program Assessment (PDA; Professional Development in Autism Center, 2008), ASAP $M = 3.67$ (0.67); BAU = 3.68 (0.59), or self-reported classroom practices as measure by the Classroom Practice Inventory (CPI; TEACCH/LEAP project team, 2007), ASAP $M = 4.67$ (2.40); BAU $M = 4.76$ (1.91). See Boyd et al. (2018) for more information on classroom variables.

Eligible children were between the ages of 3 and 5 at the time classrooms consented, were enrolled in a public preschool classroom by October 31 of that school year, had an educational classification of developmental delay or ASD and/or clinical diagnosis on the autism spectrum, and met diagnostic criteria on the Autism Diagnostic Observation Schedule–Generic (ADOS; Lord et al., 1999) administered by a research-trained project staff member (note: the ADOS-2 became available midway through this study). Children were invited to enroll in the study by their classroom teacher, and consent forms were signed by families prior to enrollment. Child exclusion criteria were having (a) an uncorrected visual or hearing impairment, (b) an uncontrolled seizure disorder, (c) a diagnosed neurogenetic syndrome comorbid with ASD (e.g., Fragile X), (d) no exposure to English at home, or (e) a level of language proficiency that made him or her eligible to receive a Module 3 ADOS. Child race was significantly different at pretest, $\chi^2(2) = 11.14$, $p = .004$, with the ASAP group having a smaller proportion of White children. No other variables approached significance for teachers, children, or parents. Because teacher fidelity of implementation is the primary focus of this article, only teacher-level variables are reported in Table 1. See Boyd et al. (2018) for additional details on parent and child-level variables.

ASAP Coaches

All ASAP coaches had at least 2 years of experience working with the public school system, and at least 3 years of direct service experience with preschool or elementary-age children with ASD (for additional coach demographics, see Boyd et al. (2018). Master (lead) coaches from the North Carolina site trained coaches from the other three study sites (Florida, Minnesota, and Oregon/Washington). In the first year, the two lead coaches were PhD candidates in Speech and Hearing Sciences with extensive experience working with preschoolers with ASD. In Years 2 through 4, the lead coach held a doctorate in Education and had previous experience teaching children with ASD. All coaches participated in a year of training in which they worked with a practice classroom at their local site and coaching quality was monitored by the lead coach throughout the year, using a checklist with randomly selected audio recordings of 20% of the monthly team meetings. ASAP coaches strictly adhered to the coaching procedures ($M = 0.91$, max possible score = 1.00; $SD = 0.14$), and there were no coaching differences between sites. Coaches at each site met with each other monthly via videoconference calls to discuss progress and needs related to their own coaching experiences. For more information on coaching in ASAP, please refer to Boyd et al. (2018).

Table 1. Classroom Teacher Demographic Data.

Demographic Information	ASAP	BAU
Teachers[a]		
Education		
AA degree	0 (0.00%)	1 (1.32%)
BA/BS degree	35 (41.18%)	20 (26.32%)
MEd/MA/MS	44 (51.76%)	52 (68.42%)
Higher	4 (4.71%)	0 (0.00%)
Years teaching	8.79 (5.62)	9.11 (5.92)
Race		
Black	5 (12.50%)	5 (12.20%)
White	33 (82.50%)	34 (85.37%)
Other	2 (5.00%)	1 (2.44%)
Hispanic	4 (10.53%)	8 (20.51%)

Note. ASAP = advancing social-communication and play; BAU = business-as-usual.
[a]Intervention missing two participants (2.35%) and BAU missing three (3.95%).

ASAP Intervention Training

Once randomized to receive the ASAP intervention, coaches provided the classroom teams with two intervention trainings over the course of the school year. The first training, approximately 4 to 6 hr in duration, provided an introduction to ASAP and occurred by the end of November of the school year in which the classroom participated in ASAP. A booster training, approximately 2 to 3 hr, occurred after the winter holiday break, to brainstorm solutions related to any ASAP implementation challenges and provide a refresher to the ASAP teams.

Fidelity Measures

Question 1: Does a multimethod approach measure fidelity to ASAP and differentiate ASAP from BAU classrooms?

Teacher interviews. To address Research Question 1, we collected teacher-report data on their adherence to the ASAP curriculum from interviews. Interviews were chosen as a measure to allow teachers to share their current level of knowledge about various aspects of the intervention. Interviews with the lead teacher in ASAP and BAU classrooms were conducted by trained research staff at the lead project site at three time points across the school year (i.e., September/October, January/February, and May/June). The interviews were conducted in person or via telephone or Skype, depending upon the teacher's preference, and lasted approximately 45 to 60 min. These interviews were conducted in both ASAP intervention and BAU control classrooms to allow us to contrast practices used in BAU classrooms versus those unique to using the ASAP intervention and how those practices change over time. An attempt was made to keep the research staff, who were graduate students and research associates, naive to treatment condition; however, there were times when they became unblinded because teachers in the ASAP condition used intervention-specific language. The interview questions focused on two randomly selected children to give a broader view of the intervention as it was used with children with a variety of skills and intervention goals. During these interviews, teachers were asked to report the following items without referring to their notes to help determine their familiarity with the ASAP curriculum: social-communication and play goals selected for each target child and details about the process of selecting goals, contexts in which they address goals (1:1 and/or group context), who addresses goals with the child, and the process for determining the next goals to target. The research staff member was asked to record teacher responses verbatim on an interview protocol. These protocols were scored by trained research staff with regard to how closely teachers adhered with ASAP goals and goal selection processes and ASAP implementation. Points were assigned based on how closely the teacher's language matched the ASAP curriculum. Child scores were averaged for each classroom. The possible score range was 0 to 20 (see Table 2 for more information on scoring). Interviews (n = 173) were independently coded by two coders. ICCs were calculated on all 173 interviews with SPSS Version 25 and were .80 for Time 1, .85 for Year 2, and .91 for Year 3.

Observational coding of teacher–child interaction quality. To further address Question 1, we collected behavioral data on teacher's instructional quality from video-recorded observations. This method was selected to allow researchers to see teachers use the intervention in their interactions with children (e.g., seeing the selected goals, materials, and strategies that the teachers typically use to address the goals). The video-recorded observations were conducted at three time points to measure the quality of instructional delivery, which was evaluated on seven dimensions (i.e., securing child's attention, teacher responsiveness and rapport, child engagement, density of opportunities, selection of goals, appropriateness of goals, and appropriateness of materials/activity). Research staff videotaped the teachers working on either a social-communication or play goal (the teacher selected which goal to target), with the two randomly selected target students (used for the teacher interviews) for 5 min each. After the teacher worked on the goal, the research staff member asked her or him to explain the selected goal as part of a brief follow-up interview. Three blinded coders, who were graduate students and research associates trained on the coding system, watched the videos and coded each dimension in the teacher's implementation and the teacher's answers to the brief follow-up interview (see Table 2 for specific items and scoring). For reliability purposes, 10% of randomly selected videos were coded.

Table 2. Fidelity Measures and Scoring Description.

Fidelity dimension	Measure	Subcomponents	Information	Scoring
Adherence differentiation	Teacher interview	Are goals related to ASAP?	If you are working on social-communication [play] with [CHILD], what is the primary goal or skill that you are targeting?	0—Not social-communication or play 1—ASAP broad category (social-interaction, requesting, joint attention) 2—ASAP specific goal language
			How did you decide to work on that particular social-communication [play] goal or skill with [CHILD]?	0—Observation, not doing it yet 1—IEP, parent, other assessments 2—ASAP hierarchy; ASAP assessments
			In this area, what goal or skill will you work on next?	0—Not social-communication or play 1—ASAP broad category (e.g., social-interaction, requesting, joint attention) 2—ASAP specific goal language
			In general, how do you decide what to work on next?	0—Observation 1—IEP, parent, other assessment 2—ASAP hierarchy; ASAP assessments
		Regularly scheduled meetings	Frequency	0 — 1x/quarter, report period, or not meeting 1—Daily, weekly, monthly
			Focus related to ASAP	0—No 1—Yes
Exposure differentiation	Teacher interview	Dosage	Did you target social-communication [play] in 1:1, group or both interactions	0—1:1 or group only 1—Both
Differentiation	Teacher–child interaction quality follow-up	Teaching and interaction quality	Securing child's attention	0—Never/rarely 1—Occasionally 2—Often 3—Almost always/always
			Teacher responsiveness/rapport - Scaffolding - Fading/wait time - Responding to child	0—Never/rarely 1—Occasionally 2—Often 3—Almost always/always
			Child engagement	0—Mostly unengaged 1—Mostly passive engagement 2—Mostly active engagement
			Frequency of attempts to engage teacher	0—No/very few attempts to engage teacher 1—Some attempts to engage teacher 2—Many frequent attempts to engage teacher
			Density of opportunities given for *initiating* target behavior	0—No/very few opportunities 1—Some opportunities 2—Many opportunities
			Goals (based on teacher statement, based on videotape)	0—Not social-communication or play 1—ASAP broad category (e.g., social-interaction, requesting, joint attention) 2—ASAP specific goal language
			Does teacher's explanation of the target goal match his or her actions during the activity	0—No 1—Yes
Summary of treatment delivery	End of year summary of ASAP delivery scale		Purposefully planned 1:1 and group activities, took advantage of teachable moments, selected appropriate materials and activities, addressed ASAP goals, all educational team members implemented the intervention; implemented intervention in 1:1 and group contexts	1–6 Likert-type scale (*strongly disagree* to *strongly agree*)

Note. ASAP = advancing social-communication and play; IEP = Individualized Education Program.

SPSS Version 25 was used to calculate the following ICC: Time 1 = .92, Time 2 = .93, and Time 3 = .82.

End of the year summary of ASAP delivery scale. In addition, we surveyed the coaches' perception of the overall quality of delivery of ASAP. At the end of the school year, individual coaches were asked to complete a brief assessment of the overall delivery of ASAP using a 6-point Likert-type scale (1 = *strongly disagree* to 6 = *strongly agree*) with possible scores ranging from 7 to 42 points. These ratings were available for the ASAP classrooms only as the coaches did not observe the BAU classrooms. Unlike the other fidelity measures, these ratings occurred at one time point and were not focused exclusively on the two randomly selected children, but rather were a more global summary of practices across the school year. Coaches rated each classroom team on seven dimensions related to ASAP implementation: (a) intentionality in planning for 1:1 and small group activities, (b) use of teachable moments to target ASAP goals, (c) selection of materials and activities, (d) implementation of ASAP in 1:1 settings, (e) implementation of ASAP in small group settings, (f) team functioning, and (g) overall implementation throughout the year. These ratings were intended to capture the coach's perception of how well the educational team was implementing ASAP. As these ratings were completed by coaches assigned to individual classrooms, and there was only one coach per classroom, the coaches' summary of ASAP delivery did not allow for reliability calculations.

Question 2: Are there student or teacher characteristics that moderate intervention fidelity? And Question 3: Does ASAP treatment fidelity mediate student outcomes?

Teacher and classroom measures. To address Research Question 2, we collected objective measures to describe the teachers and classrooms. As part of the larger study, at the beginning and end of the school year, all teachers completed the Maslach Burnout Inventory–Educator Survey (MBI-ES; Maslach, Jackson, & Schwab, 1986). The MBI-ES is a 22-item measure on which teachers rate items about their emotional exhaustion on a 7-point scale. Teachers also completed the Evidence-Based Practice Attitude Scale (EBPAS; Aarons, Cafri, Lugo, & Sawitzky, 2012) to gather information about their perspective and attitude of adopting new practices. The EBPAS has 50 items that respondents rate with a 5-point Likert-type scale about constructs, such as the fit, limitations, and teacher needs related to evidence-based practices. See Boyd et al. (2018) for more information related to project measures. The PDA (Professional Development in Autism Center, 2008) is a validated 54-item checklist measuring global classroom quality. Blinded researchers rated the classrooms structure, supports, layout, and function, including a brief teacher interview, at pretest and posttest. As previously mentioned, there were no significant pretreatment group differences.

Child measures. To address Research Questions 2 and 3, we collected behavioral data on different child characteristics. Teachers completed the Caregiver-Teacher Report Form (CTRF; Achenbach, 1997) for each child to measure baseline levels of challenging behaviors. The Mullen Scales of Early Learning (MSEL; Mullen, 1995) was administered by blinded research staff to obtain a standardized score of children's language, motor, and visual-perceptual development (using the Early Learning Composite). The ADOS was administered by a research-trained staff measure to determine severity of skills. These child measures were collected for all children enrolled in the ASAP study and there were no between-group differences at pretreatment, CTRF Total scores ASAP = 61.60 (8.90), BAU = 62.04 (8.71); Mullen scores ASAP = 52.46 (8.42), BAU = 53.90 (10.03); ADOS severity scores ASAP = 7.32 (1.75), BAU = 7.28 (1.77).

In addition, observational measures of child engagement levels in the classroom were collected at pretest and posttest by blinded research staff using tablet computers. Child engagement was coded using a continuous coding method. Child engagement states were coded as (a) unengaged, (b) onlooking, (c) object engagement, (d) person engagement, (e) supported joint engagement, and (f) coordinated joint engagement (see Boyd et al., 2018 for operational definitions of each engagement state). A "some engagement" variable (i.e., sum of object, person, supported, and coordinated states) was added to align with previous research that uses a broader definition of child engagement in classrooms (Aguiar & McWilliam, 2013; Sparapani, Morgan, Reinhardt, Schatschneider, & Wetherby, 2016). At the end of the video, coders rated the child's overall level of engagement using a 5-point Likert-type scale adapted from Kishida and Kemp (2009; 1 = *nonengaged* to 5 = *very highly engaged*). Interobserver reliability was collected for 20% of observations (ICC for pretest = .89, ICC for posttest = .65). Posttest ICCs were lower than anticipated likely due to observer drift. These states are mutually exclusive such that a child may only be in one engagement state at any given time.

Data Analysis

The data were analyzed as repeated measures of teachers, with fidelity assessments at Times 1, 2, and 3. We fit the analysis as hierarchical linear modeling (HLM) regressions with a random effects model for both the intercept and change over time. HLM manages the nonindependence that arises from repeated measures of individuals through the estimation of these random effects (Burchinal, Nelson, & Poe, 2006). Treatment differences over time were estimated as fixed effects in the model. Subsequent to the primary model, the impact of child and teacher variables on adherence was tested by the addition of teacher and child moderators as interaction terms to the above models (treatment by time by moderator). Teacher-level moderators were

Table 3. Adherence Descriptives.

	BAU				ASAP				
Timepoint	*N*	*M* (*SD*)	Minimum	Maximum	*N*	*M* (*SD*)	Minimum	Maximum	Effect size
Time 1	38	6.57 (2.76)	0.50	11.00	39	6.29 (2.45)	0.00	10.00	$d = 0.11$
Time 2	38	5.86 (2.39)	0.00	11.00	38	11.10 (3.37)	4.50	17.00	$d = -1.79$
Time 3	38	4.70 (3.12)	0.00	10.00	38	11.20 (4.99)	0.00	19.50	$d = -1.56$

Note. BAU = business-as-usual; ASAP = advancing social-communication and play.

number of years teaching, number of years teaching children with ASD, MBI-pretest, and EBPAS-pretest. Child-level moderators were CTRF-pretest, Mullen-pretest, ADOS Social Affect subscale severity score, and ADOS Total severity score. Each moderator was included in a separate model.

In addition, adherence was investigated as a mediator of child engagement outcomes (mean proportion of Some Engagement, mean proportion Unengaged, and mean Overall Engagement rating). We examined mediation models only for these child engagement variables because we found significant Group × Time interactions, in favor of the ASAP group, for these specific outcomes in the larger study. For the mediation analysis, models included direct effects for treatment, pretest, and adherence on posttest.

Results

Question 1: Does a Multimethod Approach Measure Fidelity to ASAP and Differentiate ASAP From BAU Classrooms?

Teacher interviews. Our analyses indicated that there were no statistical differences between the two groups' mean adherence scores at Time 1, but that both intercept and slope effects were significant (all $p < .001$). This reflected a change over time in ASAP classroom teachers' mean intervention adherence score, derived from the teacher interviews, contrasted to no change over time for the BAU classroom lead teachers (see Table 3), with ASAP classroom teachers demonstrating significantly higher scores at Times 2 and 3. The effect sizes at Times 2 and 3 are very large (.2 = *small effect*, .5 = *moderate effect*, .8 = *large effect*; Cohen, 1988; Lakens, 2013; 1.2 = *very large effect*, 2.0 = *huge effect*; Sawilowsky, 2009), confirming the significant difference in adherence between groups. More specifically, ASAP teachers increased their mean adherence scores to the ASAP intervention from Time 1 (6.29) to Time 2 (11.10), which was maintained at Time 3 (11.20). In contrast, the BAU classroom lead teachers' mean adherence scores remained relatively constant during the intervention period (means of 6.57, 5.86, and 4.70 at Time 1, Time 2, and Time 3, respectively). ASAP classroom teachers demonstrated an increase in adherence that control classroom teachers did not, thus discriminating between the ASAP and BAU classrooms. These patterns were confirmed via statistical analysis to differ significantly between the groups. That is, the groups are not different at Time 1 ($p = .7191$), but are at Time 2 ($p < .0001$) and Time 3 ($p < .0001$). The control group shows a significant decrease from Time 1 to Time 2 ($p = .0023$) with a marginal decrease from Time 2 to Time 3 ($p = .0566$). The treatment group increases from Time 1 to Time 2 ($p < .0001$) and remains steady from Time 2 to Time 3 ($p = .1416$). See Table 5 for post hoc analyses of differences in adherence scores over time.

These findings suggest that the ASAP model was effective in training teachers to adhere to the intervention. However, it is important to note that the mean scores at Times 2 and 3 in the ASAP group were well below the maximum possible score of 42 on this measure, and there was considerable variability within the ASAP group on adherence scores.

Teacher–child interaction quality observations. Interestingly, there were no observed differences in the coded video instructional delivery scores between ASAP and BAU classroom teachers. Total scores from these observations included ratings of the interaction itself and teacher answers to the follow-up interview. In fact, these scores remained relatively constant for both the ASAP and the control classrooms (see Table 4). There were no changes within the groups over time or differences observed between the two groups at any time point, BAU Time 1 vs. 3 = −0.75 (0.78), Time 2 versus 3 = 0.74 (0.82); ASAP Time 1 versus 3 = −1.58 (1.13), Time 2 versus 3 = 0.44 (1.19).

End of year summary of ASAP delivery scale. There was wide variability in the summary ratings completed by coaches at the end of the year. Specifically, on these forms, there was a possible total score of 42. The coaches' mean total score was 28.86 (9.32) with a range from 7 to 42. The coaches' ratings were not correlated with the adherence scores from the interviews ($r = -.13$, $p = .57$) or the teacher–child interaction quality video ($r = .12$, $p = .64$) for the ASAP group.

Table 4. Teacher–Child Interaction Quality Descriptives.

	BAU				ASAP				
Timepoint	*N*	*M* (*SD*)	Minimum	Maximum	*N*	*M* (*SD*)	Minimum	Maximum	Effect size
Time 1	51	25.05 (4.74)	9.33	33.00	65	23.69 (4.31)	9.00	32.00	$d = .30$
Time 2	48	24.14 (5.67)	4.83	31.00	50	24.39 (5.37)	7.00	32.00	$d = .05$
Time 3	43	24.26 (4.88)	11.33	32.00	50	25.23 (4.83)	10.00	32.00	$d = -.20$

Note. BAU = business-as-usual; ASAP = advancing social-communication and play.

Table 5. Adherence and Child Outcomes Using Hierarchical Linear Regression Analysis.

Effects	*M* overall	SOE	UE
Adherence direct	−.13 (.11) $p = .239$	.05 (.12) $p = .668$	−.09 (.12) $p = .434$
Pretest direct	.50 (.06) $p < .001$	.29 (.08) $p = .001$	−.43 (.06) $p < .001$
Treatment direct	.21 (.11) $p = .054$	.11 (.11) $p = .321$	.19 (.11) $p = .083$
Treat to adherence direct	.64 (.04) $p < .001$	.34 (.04) $p < .001$	.64 (.04) $p < .001$
Indirect effect	−.10 (.08) $p = .242$	.01 (.03) $p = .674$	−.07 (.09) $p = .436$

Note. SOE = some engagement; UE = unengaged.

Question 2: Are There Student or Teacher Characteristics That Moderate Intervention Fidelity?

With the significant differences observed in the ASAP classroom adherence, as indicated by scores of teacher interviews, we further examined whether there were any predictors of teachers' adherence scores in the ASAP classrooms. Specifically, we investigated whether the teachers' number of years teaching, teacher burnout, or teacher attitudes toward evidence-based practices (EBPs) predicted adherence scores. We further examined whether child characteristics at baseline (i.e., CTRF total and MSEL standard score), controlling for child age, predicted adherence scores. There was no evidence of moderation (all $p > .05$) for teacher-level variables, or for child-level moderators, with the exception of the Mullen Composite Standard Score ($p = .0408$). Mullen scores were not associated with adherence change in the BAU group either from Time 1 to Time 2 ($r = .09$, $p = .6004$) nor adherence change from Time 2 to Time 3 ($r = .19$, $p = .2359$). In the ASAP group, however, Mullen scores are negatively associated with change from Time 1 to Time 2 ($r = -.37$, $p = .0239$), but not from Time 2 to Time 3 ($r = -.10$, $p = .5634$).

Question 3: Does ASAP Treatment Fidelity Mediate Student Outcomes?

We examined whether adherence was related to the outcome of child engagement in the classroom. Child engagement ratings, specifically Unengaged, Some Engagement, and Overall Engagement, were significantly different between the ASAP and BAU groups at posttest with moderate effect sizes (.49–.56; for more discussion of these findings, see Boyd et al., 2018). The mediation models examined the indirect (mediation) effect for treatment through adherence on posttest child engagement outcomes, while covarying for the corresponding pretest child engagement variable; the indirect effect was nonsignificant in all models (see Table 5).

Discussion

Overall, our findings, based on the teacher interview data, indicate adherence to the intervention increased in the ASAP group over time. When asked about targeting social-communication and play skills as prescribed by ASAP, teachers in the ASAP group reported delivering the intervention the way it was intended. This supports the idea that teachers trained in ASAP learn how to implement ASAP in their classroom routines and activities with the support of their coach.

Although the ASAP teacher interview scores significantly increased in adherence to the ASAP model compared with BAU classrooms, teachers did not approach the highest possible scores on either the teacher interview scores or on the coaches' ratings, suggesting that there is considerable room to grow. There are multiple potential explanations for teachers not achieving higher scores on these fidelity measures, including teachers lacking paraprofessional and team resources to target goals in both 1:1 and group contexts, lack of time for holding team meetings, team buy-in and workload (e.g., some paraprofessionals rode buses home with children) affecting their availability or willingness to consistently implement the intervention, teachers not remembering target child's goals if they were using ASAP with multiple students in their classroom, and so forth.

Using the adherence scores obtained from teacher interviews, we found that teacher characteristics did not affect the degree to which they were able to implement ASAP with fidelity. Specifically, classroom teachers, regardless of experience, level of burnout, use of evidence-based practices, or global classroom quality, showed similar adherence to the ASAP intervention in their classroom. ASAP was designed to be easily adapted to any classroom serving preschoolers with ASD, and to be compatible with any existing curriculum or strategies. While only marginally significant, higher Mullen scores were related to less change in adherence to the ASAP model. This may be because children with higher Mullen scores responded well to the intervention and progressed through the hierarchy of skills ASAP targets with ease, leading to a decreased need for teachers to implement ASAP with these children. In other words, ASAP classroom teachers may have been using the intervention with higher fidelity in working with children with more severe cognitive impairments versus higher cognitive abilities because the higher functioning children needed to target higher level social-communication or play goals that were not included as part of the ASAP intervention hierarchies. It should be noted that this mediation analysis was run only for the interview data, as no fidelity changes were observed in the video-recorded interactions.

In regard to the lack of change observed in the videotaped interaction, there are some reasonable explanations. These videos represented brief (about 5 min of observation per child) one-to-one interactions between the teacher and students three times during the year. Importantly, many components of ASAP are not easily observed from such interactions, such as whether the educational team is appropriately applying knowledge of when a child mastered a goal and what goal to target next, how the classroom team planned together to address goals, and whether the targeted amount of ASAP instruction was provided at the appropriate dosage. Thus, it is possible that our observational coding was a better reflection of overall quality of instruction and teacher–child interaction, which is an important consideration for effective ASAP implementation, but not necessarily a fidelity measure that would differentiate ASAP from BAU classrooms. Although it does seem counterintuitive that an observational measure would be subjective, the variety of goals and strategies targeted as part of the ASAP intervention posed a problem for creating operational definitions of some items, such as density of opportunities. For example, some goals more easily lend themselves for repeated trials, thus naturally presenting more opportunities for practice than others.

Another interesting finding was that coaches' end of year summaries of treatment delivery did not consistently align with ASAP classroom teachers' mean adherence scores. There are a few plausible explanations for the discrepancy. First, coaches may have been more accurate in their assessment of team fidelity, based on their regular observations and interactions with classroom teams. Alternatively, coaches may have allowed their own biases related to classroom teachers, classroom teams, or schools to influence their end of year ratings. Also, although coaches worked with classroom teachers and teams for the duration of the school year, their time in the classroom was limited to observations two times per month with each visit lasting about 1 hr. Because teachers were in their classrooms every day compared with the occasional coaches' visits, and had ongoing opportunities to observe and implement ASAP, their reporting of classroom practice may more accurately reflect ASAP adherence than the coaches' ratings. However, the phone interview may have been a better measure of the teacher's knowledge about the ASAP intervention than a measure of their fidelity to the intervention in practice. This is part of the dilemma in determining how to best measure ASAP fidelity. It is promising that teachers self-report adherence to the model, but in the absence of confirmatory observational data, we cannot be confident that the teacher interviews are capturing actual implementation fidelity. In addition, the teacher interview and videotaped interaction quality were focused only on reports and behaviors of the lead teachers and may only represent teacher-implemented fidelity. In contrast, the coaches' end of year summaries may have captured overall team implementation, given that the coach was in the classroom observing implementation by the whole team. Finally, the use of teacher interviews and video observations allowed data collection related to fidelity to occur in both intervention and BAU classrooms, whereas the coaches' summaries did not allow for this comparison.

Another possible explanation for the lack of agreement between the fidelity measures is that the fidelity measures were from different measurement sources (i.e., interview, direct observation, coded videotaped interactions). These sources of fidelity ratings may be measuring different components of fidelity (Schultes, Jöstl, Finsterwald, Schober, & Spiel, 2015; Sutherland, Conroy, McLeod, Algina, & Wu, 2018). For instance, the interview may be measuring teacher knowledge of the ASAP terminology, while the videotaped interactions may be measuring the quality of teacher–child interactions.

Our findings suggested that teacher fidelity did not mediate child outcomes, and previous research reports mixed findings when examining adherence–outcomes associations (Abry, Rimm-Kaufman, Larsen, & Brewer, 2013; Berry et al., 2016; Sutherland et al., 2018). However, we only ran the mediation analysis on adherence scores as this was the score that indicated a change in fidelity. Had we seen changes in the videotaped interactions, or used an alternative measure, we may have seen mediation effects on child outcomes.

There are many notable strengths to this fidelity measurement approach. For instance, both ASAP and BAU classroom teachers were interviewed about adherence to the ASAP intervention. Oftentimes, fidelity of implementation

is measured only in the classrooms receiving intervention. Measuring fidelity by the means of conducting interviews and video-recorded observations with teachers allowed us to obtain fidelity data from both ASAP and BAU classrooms, and to compare and contrast the fidelity data. The most notable strength of this multimethod approach is that it captures the knowledge and skills specific to implementing the ASAP intervention while measuring all of the different components of fidelity. A comprehensive view of fidelity is recommended by previous research (Odom et al., 2010; Power et al., 2005).

Limitations

It also is important to note the weaknesses of this fidelity measurement approach. The interviews relied solely on teacher report and it is possible teachers were reporting implementing ASAP with fidelity only because they were familiar with the intervention and its components. Although interviewers asked teachers to put their notes away during the phone call to get a better picture of their knowledge and understanding of ASAP goals, it is possible that they referred to their notes to help them answer ASAP-specific questions (e.g., listing a specific play or SC goal). There is also some concern with coder bias in the interview or video coding. Outside factors such as the interviewer's or video coder's perceptions of teacher attitude and ASAP knowledge from earlier phone calls and videos, could certainly contribute to coder biases.

Another limitation of our fidelity measurement is that we did not measure teaching assistant and RSP implementation, relying on reports from the teacher about an intervention that is designed to be team based. There may have been other team-based factors that affected intervention implementation fidelity.

Among the lessons learned through this process is a common theme of intervention research, that of identifying and extracting active ingredients (e.g., Kasari, 2002; Pellecchia et al., 2015). In the absence of evidence that our measures of fidelity mediate student outcomes, it is likely that there are one or more important elements that we are not capturing. For example, it can be challenging to discern the difference between changes in teacher behaviors as a result of active ingredients of an intervention and those that are indicative of changes in general instructional quality, such as changes that occur naturally as the teachers discover over the course of the year which strategies are most effective for working with a given child or children. Continued examination and identification of active ingredients (i.e., components correlated with or essential to the effect of the intervention) are pertinent to any measurement of fidelity in future iterations of ASAP. It is possible that we did not see the measures of fidelity aligning with child outcomes because we did not accurately identify the active ingredients. Perhaps the presence of a coach to provide support is an active ingredient that may be more important than the intervention itself.

Another challenge was measuring the amount of treatment exposure or dosage children received. During the phone interview teachers were asked, "How often do you work on that particular goal?" Often, teachers responded, "We work on those things all the time" or "We work on it all day," which could occur in both ASAP and BAU classrooms and may have been more "overreported" in BAU classrooms. This could have happened because BAU teachers were not thinking about specific ASAP goals, whereas the ASAP classroom teachers were, after coaching, more likely to answer this question related to ASAP goal intervention dosage. Thus, this is not the most accurate way to measure exposure differentiation.

In addition, for the videotaped interactions, we have reliability information for just over 10% of the coded videos. While it appears that our coders were coding these interactions in a similar way, having 20% of the videos coded by two raters is better practice. Still, for this study, 10% of videos represents 30 of videos that were coded for reliability purposes.

Implications for Future Research and Practice

Although we are beginning to develop ways to quantify and measure fidelity, more research is needed to truly capture the construct of fidelity in intervention implementation, especially when interventions are intended to be integrated by practitioners in already existing activities and routines. The inconsistent outcomes across fidelity measures indicate the need for further refinement of ASAP fidelity procedures, and raise broader questions about measurement of fidelity within school-based interventions, especially those implemented by classroom education teams. As a result, future research should explore additional ways to capture daily implementation (e.g., measuring exposure/dosage) without being overly intrusive. Furthermore, as it relates to school-based interventions, continued research in the field of implementation science should focus on how to best distinguish active ingredients from general classroom practices.

Recommendations for practice include continuing to refine intervention coaching and training models to maximize fidelity while considering the differences between "what can work" and "what does work." One of the critical components to effective interventions is collecting feedback from stakeholders. Stakeholders can provide information about how the intervention could more easily be adopted (e.g., more coaching), factors that affected fidelity (e.g., the relationships between team members), and how fidelity could be improved (e.g., having more time for their behavior to change—more than one school year, more resources).

Conclusion

Using a multimethod approach to measure fidelity to the ASAP intervention, we have found that teachers who received ASAP training reported greater adherence to the model than the BAU classroom teachers, suggesting that our fidelity measure captured some of the active ingredients in the ASAP intervention. However, the inconsistent outcomes across fidelity measures indicate the need for further refinement of ASAP fidelity procedures, and raise broader questions about measurement of fidelity within school-based interventions, especially those implemented by classroom education teams. Continued research in the field of implementation science should focus on how to best distinguish active ingredients from general classroom practices and learn from teachers about the child-level factors they consider when implementing an intervention.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Drs. Boyd and Watson are developers of the Advancing Social-Communication and Play (ASAP) manuals and accompanying intervention and training materials, but receive no royalties. Dr. Boyd receives royalties for a book published through the Springer International Publishing company.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant No. R324A110256-14 to the University of North Carolina at Chapel Hill.

ORCID iD

Stephanie S. Reszka https://orcid.org/0000-0001-7002-2457

References

Aarons, G. A., Cafri, G., Lugo, L., & Sawitzky, A. (2012). Expanding the domains of attitudes towards evidence-based practice: The Evidence Based Practice Attitude Scale–50. *Administration and Policy in Mental Health and Mental Health Services Research*, *39*, 331–340.

Abry, T., Rimm-Kaufman, S. E., Larsen, R. A., & Brewer, A. J. (2013). The influence of fidelity of implementation on teacher–student interaction quality in the context of a randomized controlled trial of the Responsive Classroom approach. *Journal of School Psychology*, *51*, 437–453.

Achenbach, T. M. (1997). *Guide for caregiver-teacher report form for ages 2–5*. Burlington: Department of Psychiatry, The University of Vermont.

Aguiar, C., & McWilliam, R. A. (2013). Consistency of toddler engagement across two settings. *Early Childhood Research Quarterly*, *28*, 102–110.

Berry, V., Axford, N., Blower, S., Taylor, R. S., Edwards, R. T., Tobin, K., . . . Bywater, T. (2016). The effectiveness and micro-costing analysis of a universal, school-based, social–emotional learning programme in the U.K.: A cluster-randomised controlled trial. *School Mental Health*, *8*, 238–256.

Boyd, B. A., Watson, L. R., Reszka, S. S., Sideris, J., Alessandri, M., Baranek, G. T., . . . Belardi, K. (2018). Efficacy of the ASAP intervention for preschoolers with ASD: A cluster randomized controlled trial. *Journal of Autism and Developmental Disorders*, *48*, 3144–3162. doi:10.1007/s10803-018-3584-z

Burchinal, M., Nelson, L., & Poe, M. (2006). Growth curve analysis: An introduction to various methods for analyzing longitudinal data. In K. McCartney, M. Burchinal, & K. Bub (Eds.), *Monographs of the Society for Research in Child Development: Vol. 71. Best practices in quantitative methods for developmentalists* (pp. 65–87). Hoboken, NJ: Wiley-Blackwell.

Century, J., Rudnick, M., & Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evaluation*, *31*, 199–218.

Classroom Practice Inventory. (2007). *Comparison of two comprehensive treatment models for preschool-aged children with autism* (Project staff of TEACCH/LEAP project). Chapel Hill, NC: Frank Porter Graham Child Development Institute.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, *18*, 23–45.

Dykstra, J., Boyd, B. A., Watson, L. R., Crais, E. R., & Baranek, G. T. (2012). The impact of the advancing social-communication and play intervention on preschoolers with autism spectrum disorders. *Autism*, *16*, 33–49.

Estes, A., Munson, J., Rogers, S. J., Greenson, J., Winter, J., & Dawson, G. (2015). Long-term outcomes of early intervention in 6-year-old children with autism spectrum disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *54*, 580–587.

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.

Howlin, P., Moss, P., Savage, S., & Rutter, M. (2013). Social outcomes in mid- to later adulthood among individuals diagnosed with autism and average nonverbal IQ as children. *Journal of the American Academy of Child & Adolescent Psychiatry*, *52*, 572–581.

Hume, K., Boyd, B., McBee, M., Coman, D., Gutierrez, A., Shaw, E., . . . Odom, S. (2011). Assessing implementation of comprehensive treatment models for young children with ASD: Reliability and validity of two measures. *Research in Autism Spectrum Disorders*, *5*, 1430–1440.

Individuals with Disabilities Education Act, 20 U.S.C. § 1400 (2004).

Institute of Education Sciences. (2018). RFA 84.324A. Retrieved from https://ies.ed.gov/funding/pdf/2019_84324A.pdf

Kasari, C. (2002). Assessing change in early intervention programs for children with autism. *Journal of Autism and Developmental Disorders*, *32*, 447–461.

Kasari, C., & Smith, T. (2013). Interventions in schools for children with autism spectrum disorder: Methods and recommendations. *Autism*, *17*, 254–267.

Kishida, Y., & Kemp, C. (2009). The engagement and interaction of children with autism spectrum disorder in segregated and inclusive early childhood center-based settings. *Topics in Early Childhood Special Education*, *29*, 105–118.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, Article ID 863.

Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. (1999). *Autism Diagnostic Observation Schedule-WPS* (ADOS-WPS), Los Angeles, CA: Western Psychological Services.

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.

Maslach, C., Jackson, S. E., & Schwab, R. (1986). *Maslach Burnout Inventory Manual: Educators survey*. Menlo Park, CA: Mind Garden.

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. B. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation*, *24*, 315–340.

Mullen, E. M. (1995). *Mullen scales of early learning*. Circle Pines, MN: American Guidance Service.

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.

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.

Pellecchia, M., Connell, J. E., Beidas, R. S., Xie, M., Marcus, S. C., Kelleher, C., & Mandell, D. S. (2015). Dismantling the active ingredients of an intervention for children with autism. *Journal of Autism and Developmental Disorders*, *45*, 2917–2927.

Power, T. J., Blom-Hoffman, J., Clarke, A. T., Riley-Tillman, T. C., Kelleher, C., & Manz, P. H. (2005). Reconceptualizing intervention integrity: A partnership-based framework for linking research with practice. *Psychology in the Schools*, *42*, 495–507.

Professional Development in Autism Center. (2008). *PDA program assessment*. Seattle: PDA Center at University of Washington.

Sawilowsky, S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, *8*, 467–474.

Schultes, M. T., Jöstl, G., Finsterwald, M., Schober, B., & Spiel, C. (2015). Measuring intervention fidelity from different perspectives with multiple methods: The *Reflect* program as an example. *Studies in Educational Evaluation*, *47*, 102–112.

Smith, S. W., Daunic, A. P., & Taylor, G. G. (2007). Treatment fidelity in applied educational research: Expanding the adoption and application of measures to ensure evidence-based practice. *Education & Treatment of Children*, *30*, 121–134.

Sparapani, N., Morgan, L., Reinhardt, V. P., Schatschneider, C., & Wetherby, A. M. (2016). Evaluation of classroom active engagement in elementary students with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *46*, 782–796.

Strain, P. S., & Bovey, E. (2008). LEAP preschool. In J. Handleman & S. Harris (Eds.), *Preschool education programs for children with autism* (pp. 249–280). Austin, TX: Pro-Ed.

Sutherland, K. S., Conroy, M. A., McLeod, B. D., Algina, J., & Wu, E. (2018). Teacher competence of delivery of BEST in CLASS as a mediator of treatment effects. *School Mental Health*, *10*, 214–225.

Sutherland, K. S., McLeod, B. D., Conroy, M. A., & Cox, J. R. (2013). Measuring implementation of evidence-based programs targeting young children at risk for emotional/behavioral disorders: Conceptual issues and recommendations. *Journal of Early Intervention*, *35*, 129–149.

Trochim, W. M., & Donnelly, J. P. (2001). *Research methods knowledge base*. Cincinnati, OH: Atomic Dog.

Wong, C., Odom, S. L., Hume, K. A., Cox, A. W., Fettig, A., Kucharczyk, S., . . . Schultz, T. R. (2015). Evidence-based practices for children, youth, and young adults with autism spectrum disorder: A comprehensive review. *Journal of Autism and Developmental Disorders*, *45*, 1951–1966.