

The role of treatment fidelity on outcomes during a randomized field trial of an autism intervention

Autism

17(3) 281–295

© The Author(s) 2013

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/1362361312473666

aut.sagepub.com**David S Mandell**

University of Pennsylvania, USA

Aubyn C Stahmer

Rady Children's Hospital, USA

Sujie Shin

WestEd in San Francisco, USA

Ming Xie

University of Pennsylvania, USA

Erica Reisinger

University of Pennsylvania, USA

Steven C Marcus

University of Pennsylvania, USA

Abstract

This randomized field trial comparing Strategies for Teaching based on Autism Research and Structured Teaching enrolled educators in 33 kindergarten-through-second-grade autism support classrooms and 119 students, aged 5–8 years in the School District of Philadelphia. Students were assessed at the beginning and end of the academic year using the Differential Ability Scales. Program fidelity was measured through video coding and use of a checklist. Outcomes were assessed using linear regression with random effects for classroom and student. Average fidelity

Corresponding author:

David S Mandell, Center for Mental Health Policy and Services Research, Perelman School of Medicine, University of Pennsylvania, 3rd Floor, 3535 Market Street, Philadelphia, PA 19104, USA.

Email: mandelld@upenn.edu

was 57% in Strategies for Teaching based on Autism Research classrooms and 48% in Structured Teaching classrooms. There was a 9.2-point (standard deviation = 9.6) increase in Differential Ability Scales score over the 8-month study period, but no main effect of program. There was a significant interaction between fidelity and group. In classrooms with either low or high program fidelity, students in Strategies for Teaching based on Autism Research experienced a greater gain in Differential Ability Scales score than students in Structured Teaching (11.2 vs 5.5 points and 11.3 vs 8.9 points, respectively). In classrooms with moderate fidelity, students in Structured Teaching experienced a greater gain than students in Strategies for Teaching based on Autism Research (10.1 vs 4.4 points). The results suggest significant variability in implementation of evidence-based practices, even with supports, and also suggest the need to address challenging issues related to implementation measurement in community settings.

Keywords

autism, fidelity, implementation science, randomized trials, school-based intervention

Introduction

The test of any intervention is always the test of whether that intervention is effective in a particular context (Dingfelder and Mandell, 2011). This is especially true in human service organizations such as schools, in which interventions are delivered through the behavior of individuals, who are heavily influenced by the culture and resources of their work environment (Aarons et al., 2011). The traditional randomized controlled trial (RCT) minimizes the role of this context through exclusion criteria, use of highly trained clinicians, and strict adherence to a manualized treatment. The goal is to increase the internal validity of the results, that is, to ensure that any observed effect can be attributed to the intervention itself (Gordis, 2000). A resulting challenge, however, lies in assessing the external validity of the study: the extent to which results will generalize to the community settings in which one hopes proven-efficacious interventions ultimately will be used. In these settings, practitioners' skill, experience, and fidelity to the intervention manual all may differ greatly from the university-based research setting in which efficacy was assessed, potentially diminishing the intervention's effects (Fixsen et al., 2005).

Increasingly, studies of the treatment of child psychopathology have noted the differences in outcomes between RCTs and community practice and the need to conduct research that examines the outcomes in community practice settings (Hoagwood and Olin, 2002; Proctor et al., 2009; Weisz et al., 2005). While fraught with challenges, community-based studies have at least two important advantages over those conducted in traditional university-based research settings. First, they allow for direct measurement of the fit between the intervention and the system that may adopt it, the practitioners who will use it, and a less biased sample of the children who may benefit. Second, because the trial gives community practitioners a chance to test the intervention themselves, it can increase the probability that the intervention will be adopted after the study ends. To date, very little autism intervention effectiveness research has been published (Magiati et al., 2007; Perry et al., 2008), with the notable exception of a pragmatic trial of the picture exchange communication system (Howlin et al., 2007).

At least two decades of study suggest the efficacy of behaviorally based approaches for treating children with autism (Rogers, 1998; Rogers and Vismara, 2008). The best tested of these interventions are in the family of applied behavior analysis (ABA) and include didactic (Lord and McGee, 2001; Lovaas, 1987) and more incidental (Koegel et al., 2003; Lord and McGee, 2001; Schreibman, 2000) approaches. These methods often require extensive practitioner training, a very high

staff-to-child ratio, and have been tested at levels of intensity (i.e. 20–40 h per week) that rarely are feasible outside of grant-funded research. It is not surprising, therefore, that the few studies of community practice find considerable deviations from proven-efficacious practice and relatively poor outcomes (Bibby et al., 2001; Boyd and Corley, 2001; Chasson et al., 2007; Sheinkopf and Siegel, 1998; Stahmer et al., 2005).

Comprehensive intervention packages combine instructional strategies with curriculum content in a manner designed for classroom implementation (Odom et al., 2010). Because of the lower staff-to-child ratio required, careful manualization, and combination of instructional strategies with curricular content, these programs generally are better suited to community settings. While based on proven-efficacious methods, however, few published studies present rigorous data on their outcomes. One such program is Strategies for Teaching based on Autism Research (STAR), which combines three different ABA methods with curriculum content in six areas (Arick et al., 2004). To date, only one single-group prestudy/poststudy examines its efficacy, finding that about one-third of students made age-equivalent language gains (Arick et al., 2003).

Because of STAR's reliance on ABA techniques, which families often request, and because it is designed for classroom implementation, the School District of Philadelphia chose to implement STAR in its kindergarten-through-second-grade autism support classrooms. The district partnered with our research team to examine the implementation and outcomes. The resulting study, the Philadelphia Autism Instructional Methods Study (AIMS), represents one of the few randomized trials of a comprehensive classroom-based intervention for children with autism spectrum disorders (ASD). The trial was implemented using school district staff, increasing the external validity of the outcomes. We also implemented a rigorous comparison condition that allowed us to attribute resulting outcomes to program, rather than training effects. The current article presents the intent-to-treat outcomes of this comparison of STAR and a rigorous comparison group, as well as the moderating effect of teacher characteristics, including their fidelity to the program.

Methods

"Teaching as usual" prior to program implementation

Prior to the start of this study, no systematic, regular training was provided to autism support teachers in the district once they entered the classroom, and there was no standard program or curriculum. Teachers are required to have a masters-level certification in special education, but no autism-specific experience or training per se. In the year prior to the start of the study, the district had contracted with the same group that we contracted with to provide training in Structured Teaching (see the following) to provide 2 days of didactic training to autism support teachers.

STAR

We contracted with the STAR program developers to provide training and ongoing support to classroom staff. A complete description of STAR is available at www.starautismprogram.com. STAR relies on three teaching techniques in the family of ABA: discrete trial training (Lovaas, 2003; Lovaas and Buch, 1997), pivotal response training (Koegel and Koegel, 1995, 1999), and teaching in functional routines (Alberto and Troutman, 2002; Cooper et al., 1987; Malott et al., 2002). Discrete trial training and pivotal response training are designed for one-to-one instruction,

and functional routines are designed for use with groups. These techniques are paired with a curriculum with 169 lesson plans that address receptive, expressive, and spontaneous language; functional routines; and preacademic, play, and social concepts. Lesson choice is guided by the student learning profile, which identifies appropriate lessons and instructional strategies for teaching each skill. Teachers implement the lesson with the student until the student is able to correctly use the skill four out of five times without prompting. Classrooms include visual schedules, labels, and cues, clearly and consistently placed throughout the room and one-to-one work areas. STAR includes regular and systematic data collection for each child.

Description of the comparison condition: Structured Teaching

The district had contracted previously with a well-regarded autism school to provide training in Structured Teaching. We contracted with this school to provide training of the same intensity and duration as in the intervention condition. While the trainers are Treatment and Education of Autistic and Communication related handicapped Children (TEACCH) certified, this school is not officially connected to Division TEACCH in North Carolina.

Structured Teaching was developed through the TEACCH program. (A description of the program is available at www.teacch.com.) Structured Teaching methods are based on the observation that ASD is accompanied by a receptive language delay and difficulty with environmental inconsistencies. Instruction is systematically organized to ease transitions. Visual instructions are added to verbal directions to increase the students' understanding, responding, and level of independence. Visual structure is integrated into teaching methods using pictorial instructions, organization, and clarity to highlight important tasks, information, and concepts. Samples or pictures of finished tasks are used to help students visually conceptualize the goals of the tasks. Teachers rely on physical prompts, with prompts and cues gradually removed as the student progresses. Other teaching methods include providing only those materials necessary for students to complete a task to reduce distractions. Instruction utilizes reinforcement of correct responding. Use of specific reinforcers is maximized by identifying objects or activities that are particularly reinforcing to students (Charlop-Christy et al., 2002; Dettmer et al., 2000; Simpson, 2005). Clear transition signals are used to help children move through the day independently. Classroom staff also has clear schedules for the day and are expected to support group activities by modifying materials to meet students need and provide minimal assistance with independent work.

Curriculum content was selected independently by teachers and was based on Pennsylvania's alternate curriculum standards, which emphasize functional independence and adaptive behavior, in addition to more minimal academic achievement. Teachers developed lesson plans and activities to address content in the Structured Teaching format. Teachers were expected to develop an ongoing, systematic process for assessing student learning based on individualized education plan (IEP) goals. Teachers did not learn a specific process for data collection.

Program overlap and differences

While STAR and Structured Teaching have many differences, some similarities emerged during the training that were not apparent from reading the manuals or communication with the trainers. Specifically, the classroom organizational and scheduling/transition strategies in Structured Teaching were very similar to the use of functional routines in STAR. The primary differences between the two programs therefore were the use of one-to-one instructional strategies and a highly specified curriculum in STAR.

Staff training

Staff training in both conditions included the following: (1) 28 h of intensive workshops at the beginning of the academic year, (2) hands-on work with teachers to set up classrooms, and (3) 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.

Sample

Table 1 describes classroom staff and students participating in the study. Eligible staff comprised those working in one of the district's 42 kindergarten-through-second-grade autism support classrooms. The district sent a letter to each staff member, asking them to participate. Staff was required to participate in training in either STAR or Structured Teaching as part of professional development, but was not required to participate in the research study.

Of these 42 classrooms, staff in 38 (90%) agreed to participate. No students consented to participate in three classrooms. In two classrooms, data collection at baseline on participating students was incomplete, leaving 33 classrooms (79%). There were no statistically significant differences in demographics or years teaching between staff in these 33 classrooms and the 9 classrooms that ultimately did not participate (data not shown).

Table 1 presents the teacher and student characteristics by intervention condition. Classrooms had an average of two staff members. Twenty teachers (61%) had ≥ 3 years' experience working with children with autism. Classrooms were capped at eight students, with an average of 3.7 students per classroom enrolled in the study. There were no statistically significant differences on these variables between the two intervention conditions.

There were 121 consented students, representing 52% of all eligible students in these 33 classrooms. Baseline testing of two students was later deemed invalid; these students were excluded from the analysis, leaving a sample of 119. Children were more than 5 years and less than 9 years of age. Students were recruited through a consent form and flyer describing the program that teachers sent home with the student. Parents received US\$50 for the first wave of data collection, US\$100 for the second wave of data collection, and a summary report of their child's assessment that was designed for use in IEP planning. Inclusion criteria for students were that they have a classification of autism through the district and be enrolled at least half time in participating classrooms.

The average age of students was 6.2 (standard deviation (SD) = 0.6) years at study entry. The majority was male (85.9%) or African American (53.7%). The average IQ, as measured by the *Differential Ability Scales, Second Edition (DAS-II)*, at baseline was 58.6 (SD = 15.9). The average score on the Autism Diagnostic Observation Schedule (ADOS) severity algorithm was 6.7 (SD = 1.0). The groups differed significantly in racial/ethnic composition. The percentage of African American students in the STAR group was twice that in the Structured Teaching group and vice versa for Latino students. This difference was due to two schools with high recruitment in predominantly Latino neighborhoods being randomized to the Structured Teaching group. Because randomization and the first training occurred prior to student recruitment, we were unable to randomize again when this failure was discovered.

Measures

Child outcome was measured with the *DAS-II* (Elliott, 1990), which is designed to assess cognitive abilities in children aged 2 years 6 months through 17 years 11 months across a broad range of

Table 1. Philadelphia Autism Instructional Methods Study teacher/classroom characteristics (N = 33).

	STAR	Structured Teaching	p value
	n = 18	n = 15	
<i>Teacher characteristics</i>			
Years of autism teaching experience			
≤3 years	50.0%	60.0%	.566
>3 years	50.0%	40.0%	
Program fidelity by end of the observation period	57%	48%	.177
Low (STAR = 0.12–0.49; Structured Teaching = 0.17–0.42)	27.8%	26.7%	.917
Moderate (STAR = 0.5–0.68; Structured Teaching = 0.43–0.54)	33.3%	40%	
High (STAR = 0.69–0.92; Structured Teaching = 0.55–0.71)	38.9%	33.3%	
Hours of training/support	53.8 h	16.5 h	.010
<i>Student characteristics</i>			
	n = 60	n = 59	
Male	81.7%	89.8%	.203
Race/ethnicity			
Black	70.0%	35.6%	<.001
Hispanic	1.7%	15.3%	
White	20.0%	32.2%	
Other	8.3%	16.9%	
<i>Classroom characteristics (average within classroom)</i>			
Student age in years (mean (SD))	6.2 (0.43)	6.3 (0.77)	.676
DAS score at Time 1 (mean (SD))			
General conceptual ability	61.0 (16.87)	57.6 (13.78)	.544
Verbal reasoning	51.5 (13.99)	51.2 (14.61)	.839
Spatial reasoning	63.1 (23.65)	57.6 (19.03)	.475
Nonverbal reasoning	68.1 (17.67)	64.1 (12.05)	.455
DAS score at Time 2 (mean (SD))			
General conceptual ability	69.8 (15.21)	67.1 (12.09)	.590
Verbal reasoning	61.6 (14.78)	58.2 (15.30)	.523
Spatial reasoning	69.3 (21.08)	69.5 (13.57)	.967
Nonverbal reasoning	78.8 (15.88)	73.4 (12.63)	.296
ADOS algorithm severity score at Time 1 (mean (SD))	6.7 (1.14)	6.7 (0.95)	.914

STAR: Strategies for Teaching based on Autism Research; DAS: Differential Abilities Scale; ADOS: Autism Diagnostic Observation Schedule; SD: standard deviation.

developmental areas. Subtests measure verbal and visual working memory, immediate and delayed recall, visual recognition and matching, processing and naming speed, phonological processing, and understanding of basic number concepts. Psychologists trained to research reliability administered the *DAS-II* at baseline (September) and at the end of the school year (May).

Program fidelity measures for each program were designed based on the manuals and in consultation with the trainers. Each teacher was filmed for 30 min each month. Videoing was designed to allow for 10 min of video of discrete trial training, pivotal response training, and functional routines for the STAR group, and for a range of transitions, group, and 1:1 activities for the Structured

Teaching group. A brief overview of the classroom setup was also filmed for each classroom by panning through each area of the room.

Trained research assistants and undergraduate students blind to study condition coded fidelity based on the videos. Coding used different criteria for each teaching technique. For example, during the pivotal response training segment, teachers were coded for gaining the child's attention, providing clear and appropriate instructions, providing the child a choice of stimuli, eliciting a mixture of maintenance and acquisition tasks, and using natural reinforcers. For discrete trial training, teachers were coded on their ability to gain a child's attention, provide clear, appropriate instructions, use appropriate prompting strategies, provide clear and correct consequences, and utilize error correction procedures. Teachers were coded for each step of functional routines used in the classrooms. Coding of Structured Teaching fidelity included students' use of picture schedules and clear visual set up of independent activities.

One of the authors (A.C.S.) recorded two tapes for each coder every other month, to measure criterion validity. Less than 90% agreement with the research assistant resulted in additional training until this level of agreement was achieved. Based on the video, fidelity was coded as a percentage of components met for fidelity (possible range = 0%–100%).

Other teacher characteristics potentially associated with fidelity and student outcome included experience teaching children with autism (measured by self-report) and hours attending training and receiving consultation in the classroom (measured by attendance sheets and reports from the consultants).

Student characteristics, potentially associated with outcome, included age, *DAS-II* score at baseline, and autism severity. Autism severity was measured using the ADOS (Lord et al., 2000). The ADOS was administered at baseline in the school by a clinician trained to research reliability. ADOS scores were then converted to the ADOS severity algorithm, a 10-point, validated scale that allows for comparisons of severity across the four modules of the ADOS (Gotham et al., 2009).

Randomization and masking

Randomization occurred at the classroom level. Classrooms were randomized in block sizes of two, as teachers were consented, using a random number generator in SAS. Assessors were blinded to classroom condition. While all assessments were conducted at schools, a research assistant went into the classroom to retrieve the subject, who was then brought to another room in the school for assessment.

Analysis

Means and SDs or percentages as appropriate were calculated for each variable of interest and compared using t-test or chi-square test. Because of differences between groups, program fidelity and hours of training were percentiled separately for each group. There were no differences in the magnitude of associations of these variables with outcomes when they were jointly percentiled (data not shown). Difference in outcome by group was estimated with longitudinal linear models with random effects for classroom and student (Donner and Klar, 2000; Murray et al., 1998; Sashegyi et al., 2000). The independent variables included a categorical variable for time period (baseline, 8 months), group (STAR or comparison), and their interaction. Because the randomization was not successful with regard to race/ethnicity, the bivariate association of other variables with outcome was estimated for the entire sample. Variables significant at $p < .2$ were included in an adjusted model as covariates. The three-way interaction between time, treatment group, and the

moderator was used to assess the presence and magnitude of the moderating effect of each variable.

Statistical power

We conducted a power analysis to ascertain the minimum sample size necessary to achieve 80% power to detect a moderate effect of Cohen's $h = 0.5$ using a t-test comparing the mean difference between baseline and 8-month outcomes between the intervention and comparison groups. We assume an alpha of 0.05. We assumed a conservative intraclass correlation (ICC) within classrooms of $\rho = 0.15$ (Hsieh et al., 2003). Using the Power and Precision software package (Borenstein, 2000), we calculated that with a sample size of 128 cases (64 in each group), the study would have power of 80% to yield a statistically significant result.

Results

Table 1 describes the sample by condition. Average fidelity of implementation was 57% (range = 12%–92%) among STAR teachers and 48% (range = 17%–71%) among Structured Teaching teachers ($p = .177$). Average number of hours of training and consultation was 53.8 in the STAR group and 16.5 in the Structured Teaching group ($p = .01$). Students did not statistically significantly differ between groups in their age, sex, *DAS-II* score, or ADOS score. Sixteen percent of students floored on the *DAS-II*, with no differences between groups ($p = .456$; data not shown). Students did differ in race/ethnicity, with a larger portion of students enrolled in the Structured Teaching arm being Latino. Table 2 presents the results of two regression models predicting change in *DAS-II* score. The first comprises the intent-to-treat analysis. The second includes as main effects program fidelity, years of autism experience, and hours receiving training or support through the study.

There was a 9.2-point ($SD = 9.6$) increase in *DAS-II* score over the 8-month study period ($p < .001$; data not shown). The difference between STAR (9.7 points; $SD = 1.2$) and Structured Teaching (8.6 points; $SD = 1.2$) was not statistically significant. There was no statistically significant effect on outcome of hours of training and support, program fidelity, years of autism teaching experience, student sex, race/ethnicity, or ADOS severity score. Each additional year of student age was associated with a 3.4-point ($SD = 0.9$) average decrease in outcome ($p < .001$). Each additional point of *DAS-II* score at baseline was associated with a 0.1 ($SD = 0.04$) decrease in average *DAS-II* gain ($p < .001$). All findings persisted after adjustment.

In post hoc analysis, there was a significant interaction between fidelity and intervention group (Figure 1; $p = .026$). In classrooms with either low or high program fidelity, students in STAR experienced a greater gain in *DAS-II* score than students in Structured Teaching. In classrooms with moderate fidelity, students in Structured Teaching experienced a greater gain than students in STAR. No other interactions were statistically significant (Figure 2).

Discussion

This comparison of STAR and Structured Teaching found that training classroom staff in either resulted in a clinically meaningful IQ increase in one academic year among children in kindergarten-through-second-grade autism support classrooms, gains equivalent to those experienced by students in the single-group study of STAR (Arick et al., 2003), and much greater than would be expected from normative data.

Table 2. Intent-to-treat and adjusted regression predicting study outcomes.

	Unadjusted		Adjusted	
	Estimate	p value	Estimate	p value
Intervention status				
STAR	9.73	.531	9.11	.955
Structured Teaching	8.63		9.21	
Hours of training/support (per hour)				
Low	7.97	.703	—	—
Moderate	9.28		—	
High	9.81		—	
Program fidelity				
Low	8.49	.608	—	—
Moderate	8.40		—	
High	10.25		—	
Years of autism experience				
>3 years	9.78	.586	—	—
≤3 years	8.79		—	
Sex				
Male	9.12	.852	—	—
Female	9.59		—	
Race/ethnicity				
Black	9.29	.208	9.44	.356
Hispanic	6.18		7.03	
White	7.89		7.86	
Other	13.48		12.32	
Student age in years	-3.41	<.001	-3.23	<.001
DAS general conceptual ability score at Time 1	-0.13	<.001	-0.13	<.001
ADOS algorithm severity score at Time 1	0.30	.538	—	—

STAR: Strategies for Teaching based on Autism Research; DAS: Differential Abilities Scale; ADOS: Autism Diagnostic Observation Schedule.

These gains are especially intriguing because neither group achieved high fidelity. Typically, a benchmark of 80% fidelity is used in these types of field trials (Reichow, 2011; Wilczynski and Christina, 2008); only two classrooms achieved this benchmark in the STAR group and none achieved it in the Structured Teaching group. Teachers were provided with significantly more training and in-classroom consultation than they usually receive. Yet, in these under-resourced, urban classrooms, staffing ratios and preparation often were less than optimal. The state of classroom practice prior to program implementation may have been such that even small changes resulted in student gains.

The lack of association between hours of support and outcomes is contrary to previous findings (Beidas et al., 2011). Training may have missed critical components necessary to achieve fidelity and outcomes. For example, STAR comprises a complex program. Teachers may not have been able to assimilate and implement all components concurrently. Poststudy interviews suggest that the STAR trainers emphasized one-to-one instruction because they believed that these were the primary active mechanisms. Teachers reported that they felt competent in these techniques but did not use them frequently because it was difficult to pull one child for instruction without better

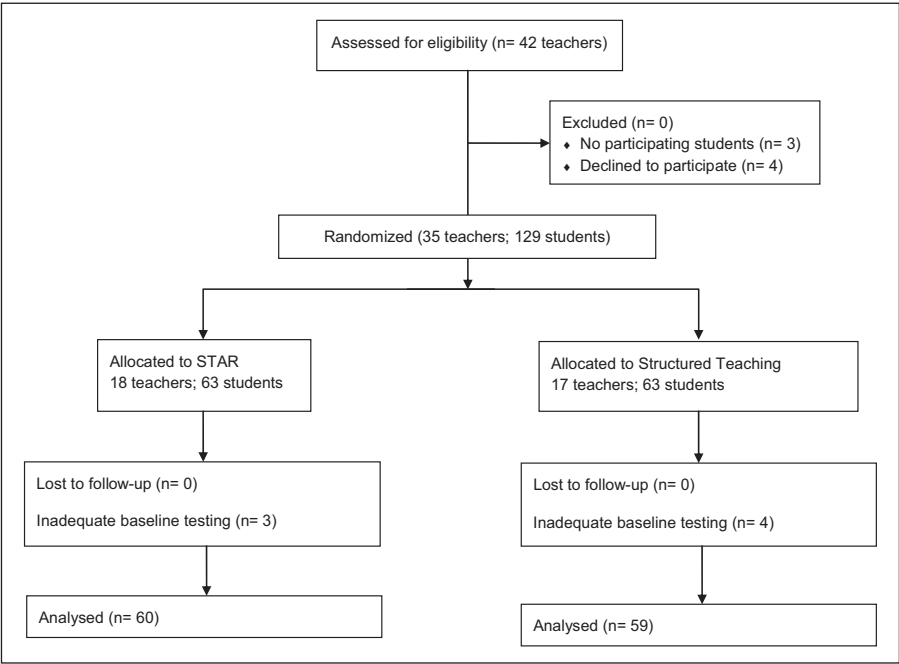


Figure 1. CONSORT flow diagram for the teacher and student assignment.
STAR: Strategies for Teaching based on Autism Research; CONSORT: Consolidated Standards of Reporting Trials.

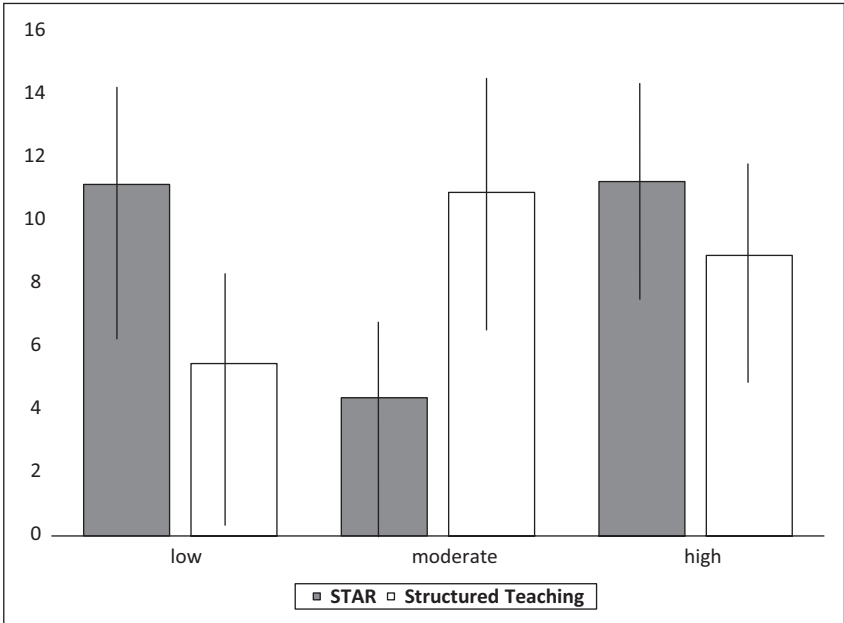


Figure 2. Interaction of fidelity and program in predicting outcome with 95% CI.
STAR: Strategies for Teaching based on Autism Research; CI: confidence interval.

classroom management strategies. If true, then training should have initially emphasized classroom management and functional routines, layering one one-to-one techniques afterward.

A second hypothesis is that when teachers were not implementing the intervention, trainers responded by providing more support. If these teachers were not responsive to coaching, then increasing hours would not have improved fidelity or outcomes. In this case, increased hours were “chasing” lack of fidelity, but did not comprise an effective strategy for improving it (Webb et al., 2010).

Despite generally low fidelity, group differences in outcome were moderated by program fidelity, such that in both the highest and lowest tertiles of fidelity, STAR outcomes were better than Structured Teaching outcomes, while the reverse was true for the middle tertile. The different associations between fidelity and outcomes in the two conditions may speak to two countervailing trends. In the STAR group, anecdotal evidence suggests that the lowest fidelity group included very experienced teachers who agreed to participate in the study but were not interested in changing their teaching strategies. These strategies did not correspond to the requirements for STAR but may have been equally effective. The middle tertile of fidelity in the STAR group may include teachers who implemented STAR components with less deep structure knowledge (Han and Weiss, 2005). These multiple components may have been too much for teachers to juggle, resulting ultimately in less effective instruction. Alternatively, Structured Teaching, with fewer components, mostly around classroom management, may have allowed for more flexibility to include effective instructional strategies while maintaining fidelity (Kendall and Beidas, 2007). It should be made clear, however, that these analyses were conducted post hoc, and beg replication prior to acceptance of this unusual association between fidelity and outcomes. Alternatively, they may speak to the importance of careful measurement of fidelity. Most field-based intervention studies use self-report or checklists. Our more thorough measures of fidelity may have resulted in the difference in its observed association with outcomes.

Structured Teaching and the classroom organization in STAR share many components, but STAR contains many more intervention components. Structured Teaching components may represent the minimum components necessary upon which to build effective instruction for children with autism. Even the low-fidelity STAR classrooms may have included these components, which would explain the poor outcomes in the low-fidelity Structured Teaching classrooms. The relatively poor outcomes in the high-fidelity Structured Teaching classrooms may relate to a phenomenon that has been observed in studies of psychotherapy, in which poor outcomes precede fidelity (Webb et al., 2010). Therapists may increase fidelity when clients appear not to be responding positively. Similarly, teachers in the Structured Teaching group may have increased their fidelity to attempt to obtain better outcomes, but Structured Teaching may have offered too limited an intervention to obtain these outcomes.

The moderating effects of two student characteristics also are worth noting. First, increased age was associated with decreased gains over the study period. Some have argued that there is a “developmental window” during which time children’s response to traditional behavioral intervention is greater. Our study may provide some evidence of this window and points to the critical importance of early intervention. Our second finding that students with lower cognitive ability at baseline achieved greater gains may reflect two phenomena. First, these interventions may be of particular benefit to this subgroup of children. Second, the results may reflect some regression to the mean in measurement of cognitive ability.

This study had several limitations. First, our randomization was not successful with regard to ethnicity, suggesting the need for future pragmatic trials to employ more stratified randomization strategies. Second, fidelity was measured through video at monthly intervals. While these

measures provided rich cross-sectional data, it is not clear the extent to which fidelity was achieved when members of the research team were not observing the classroom. Still, these measures of fidelity offer a considerable improvement over frequently used self-report from teachers (Fixsen et al., 2005). Third, the comparison group comprised a rigorous and promising intervention, rather than “teaching as usual.” While this may be considered a strength, direct observation suggests that the components comprising Structured Teaching could be considered a subset of STAR. This overlap may have attenuated differences in outcomes. Because we do not have multiple baseline measures of student progress, it is not clear the extent to which improvements observed in the current study represent a change from usual outcomes. This finding speaks to the need for different designs in comparative effectiveness studies, in which multiple baseline measures are taken. Finally, we relied on the *DAS* as an outcome measure. While measures of cognitive ability have a long tradition of use in studies such as this one (Rogers and Vismara, 2008), they may miss important components of academic achievement.

Despite these limitations, there are important implications related to these findings. First, the findings provide strong indication that school-age children with autism can experience significant cognitive gains. Perhaps more importantly, these gains can be achieved in a large, urban public school system. Large school districts such as Philadelphia represent a small proportion of all school districts in the United States; yet, they serve a huge percentage of all school children. As a case in point, the largest 500 districts comprise 2.8% of all districts, yet serve 43% of all students (Sable et al., 2010). Students in these districts are more likely to be poor and from traditionally underrepresented minority groups, both of which groups are underrepresented in autism research (Lord et al., 2005). These districts also serve a disproportionately large number of children with special education needs and often are under court order to improve their care of these children (Yell and Drasgow, 2000). Yet, researchers have traditionally not partnered with them to develop strategies to improve and test care, despite the fact that effecting change in these districts could result in an improved understanding of the best ways to implement evidence-based practice in these sometimes-challenging settings, as well as substantial gains for many children (Atkins et al., 2003; Fantuzzo et al., 2003).

A second set of implications relates to the puzzling relationship observed between program fidelity and student outcomes. General consensus is that only when programs are fully implemented, should we expect positive outcomes (Fixsen et al., 2005; Institute of Medicine, 2001). Our study suggests a more complex relationship. One possibility is that teachers may systematically or, more likely, intuitively abstract program components and apply them appropriately to students based on individual need. If this is the case, then researchers have much to learn from teachers about selectively applying intervention components based on context (Green, 2008). A second possibility is that while the fidelity measures were extracted directly from the program manuals, they did not capture the active intervention mechanisms. Comprehensive programs often represent a mix of different approaches, and the importance of core ingredients may not be empirically determined (Odom et al., 2010), suggesting the need for dismantling designs. Finally, recent studies have found that interventions addressing organizational climate as well as program content can have a powerful effect on outcomes (Aarons et al., 2011; Glisson et al., 2010). In the second phase of AIMS, we have begun to measure factors such as climate for innovation implementation (Klein and Sorra, 1996) and classroom staff cohesion to examine their effects on both fidelity and outcomes.

Autism research has not yet addressed the considerable gap between research findings and practice (Green, 2008; Weisz et al., 1995, 2004). Many other disciplines have begun to address these issues and recognize that the “gap” does not necessarily represent a failure on the part of

community practitioners, but rather a failure of researchers and policy makers to make research and policy relevant (Green et al., 2009) or to recognize the effective (but untested) work already being conducted in community settings (Van De Ven and Johnson, 2006). In its next iterations, we hope that AIMS will begin to address both the implementation of evidence-based practice in community settings and what researchers can learn from community practitioners about meeting the needs of these children. These next steps include developing and testing training and consultation strategies, determining the most efficient and effective methods to measure fidelity, determining other school and classroom characteristics that moderate the relationship between fidelity and outcomes, and observing teachers who obtain positive student outcomes, regardless of fidelity to a particular intervention model, to learn what strategies are effective in their environment.

Funding

This study was funded by grants from the National Institute of Mental Health (R01MH083717) and the Institute of Education Science (R324A080195).

References

- Aarons G, Hurlburt M and Horwitz S (2011) Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Administration and Policy in Mental Health* 38(1): 4–23.
- Alberto P and Troutman A (2002) *Applied Behavior Analysis for Teachers*. 6th ed. Upper Saddle River, NJ: Merrill.
- Arick J, Loos L, Falco R, et al. (2004) *The STAR Program: Strategies for Teaching Based on Autism Research*. Austin, TX: PRO-ED.
- Arick J, Young H, Falco R, et al. (2003) Designing an outcome study to monitor the progress of students with autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities* 18: 75–87.
- Atkins MS, Graczyk PA, Frazier SL, et al. (2003) Toward a new model for promoting urban children's mental health: accessible, effective, and sustainable school-based mental health services. *School Psychology Review* 32(4): 503–514.
- Beidas R, Koerner K, Weingardt K, et al. (2011) Training research: seven practical suggestions for maximum impact. *Administration and Policy in Mental Health* 38: 223–237.
- Bibby P, Eikeseth S, Martin NT, et al. (2001) Progress and outcomes for children with autism receiving parent-managed intensive interventions (erratum appears in *Research in Developmental Disabilities* 2002;23(1): 81–104). *Research in Developmental Disabilities* 22(6): 425–447.
- Borenstein J (2000) *Power and Precision Software*. Teaneck, NJ: Biostat Corp.
- Boyd R and Corley M (2001) Outcome survey of early intensive behavioral intervention for young children with autism in a community setting. *Autism* 5(4): 430–441.
- Charlop-Christy MH, Carpenter M, Le L, et al. (2002) Using the picture exchange communication system (PECS) with children with autism: assessment of PECS acquisition, speech, social-communicative behavior, and problem behavior. *Journal of Applied Behavior Analysis* 35(3): 213–231.
- Chasson G, Harris G and Neely W (2007) Cost comparison of early intensive behavioral intervention and special education for children with autism. *Journal of Child and Family Studies* 16: 401–413.
- Cooper J, Heron T and Heward W (1987) *Applied Behavior Analysis*. Columbus, OH: Merrill.
- Dettmer S, Simpson R, Myles B, et al. (2000) The use of visual supports to facilitate transitions of students with autism. *Focus on Autism and Other Developmental Disabilities* 15(3): 163–169.
- Dingfelder HE and Mandell DS (2011) Bridging the research-to-practice gap in autism intervention: an application of diffusion of innovation theory. *Journal of Autism and Developmental Disorders* 41(5): 597–609.
- Donner A and Klar N (2000) *Cluster Randomization Trials in Health Research*. London: Arnold.
- Elliott C (1990) *Differential Ability Scales: Administration and Scoring Manual*. 2nd ed. San Antonio, TX: Psychological Corp.

- Fantuzzo J, McWayne C and Bulotsky R (2003) Forging strategic partnerships to advance mental health science and practice for vulnerable children. *School Psychology Review* 32(1): 17–37.
- Fixsen DL, Naoom SF, Blase KA, et al. (2005) *Implementation research: a synthesis of the literature*. FMHI Publication #231. Tampa, FL: The National Implementation Research Network, Louis de la Parte Florida Mental Health Institute, University of South Florida.
- Glisson C, Schoenwald S, Hemmelgarn A, et al. (2010) Randomized trial of MST and ARC in a two-level evidence-based treatment implementation strategy. *Journal of Consulting and Clinical Psychology* 78(4): 537–550.
- Gordis L (2000) *Epidemiology*. 2nd ed. Philadelphia, PA: W.B. Saunders Company.
- Gotham K, Pickles A and Lord C (2009) Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism and Developmental Disorders* 39(5): 693–705.
- Green L (2008) Making research relevant: if it's an evidence-based practice, where's the practice-based evidence? *Family Practice* 25(Suppl. 1): 20–24.
- Green L, Glasgow R, Atkins D, et al. (2009) Making evidence from research more relevant, useful, and actionable in policy, program planning, and practice: slips “Twixt Cup and Lip.” *American Journal of Preventive Medicine* 37(6 Suppl. 1): S187–S191.
- Han SS and Weiss B (2005) Sustainability of teacher implementation of school-based mental health programs. *Journal of Abnormal Child Psychology* 33(6): 665–679.
- Hoagwood K and Olin S (2002) The NIMH blueprint for change report: research priorities in child and adolescent mental health. *Journal of the American Academy of Child and Adolescent Psychiatry* 41(7): 760–767.
- Howlin P, Gordon R, Pasco G, et al. (2007) The effectiveness of Picture Exchange Communication System (PECS) training for teachers of children with autism: a pragmatic, group randomised controlled trial. *Journal of Child Psychology and Psychiatry, and Allied Disciplines* 48(5): 473–481.
- Hsieh F, Lavori P and Cohen H (2003) An overview of variance inflation factors for sample-size calculation. *Evaluation & the Health Professions* 26: 239–257.
- Institute of Medicine (2001) *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academy Press.
- Kendall P and Beidas R (2007) Smoothing the trail for dissemination of evidence-based practices for youth: flexibility within fidelity. *Professional Psychology, Research and Practice* 38(1): 13–20.
- Klein KJ and Sorra JS (1996) The challenge of innovation implementation. *Academy of Management Review* 21(4): 1055–1080.
- Koegel L and Koegel R (1995) Pivotal response intervention II: preliminary long-term outcome data. *The Journal of the Association for Persons with Severe Handicaps* 24: 186–198.
- Koegel L and Koegel R (1999) Pivotal response intervention I: overview of approach. *The Journal of the Association for Persons with Severe Handicaps* 24: 174–185.
- Koegel L, Carter C and Koegel R (2003) Teaching children with autism self-initiations as a pivotal response. *Topics in Language Disorders* 23: 134–145.
- Lord C and McGee J (eds) (2001) *Educating Children with Autism*. Washington, DC: National Academy Press.
- Lord C, Risi S, Lambrecht L, et al. (2000) The autism diagnostic observation schedule—generic: a standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders* 30: 205–223.
- Lord C, Wagner A, Rogers S, et al. (2005) Challenges in evaluating psychosocial interventions for autistic spectrum disorders. *Journal of Autism and Developmental Disorders* 35(6): 695–708.
- Lovaas O (1987) Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology* 55: 3–9.
- Lovaas O (2003) *Teaching Individuals with Developmental Delays: Basic Intervention Techniques*. Austin, TX: PRO-ED.
- Lovaas O and Buch G (1997) Intensive behavioral intervention with young children with autism. In: Singh N (ed.) *Prevention and Treatment of Severe Behavior Problems: Models and Methods in Developmental Disabilities*. Pacific Grove, CA: Brooks/Cole Publishing Co., pp. 61–86.
- Magiati I, Charman T and Howlin P (2007) A two-year prospective follow-up study of community-based early intensive behavioural intervention and specialist nursery provision for children with

- autism spectrum disorders. *Journal of Child Psychology and Psychiatry, and Allied Disciplines* 48(8): 803–812.
- Malott R, Malott M and Trojan E (2002) *Elementary Principles of Behavior*. 4th ed. Upper Saddle River, NJ: Prentice Hall.
- Murray D, Hannan P, Wolfinger R, et al. (1998) Analysis of data from group-randomized trials with repeat observations on the same groups. *Statistics in Medicine* 17: 1581–1600.
- Odom SL, Boyd BA, Hall LJ, et al. (2010) Evaluation of comprehensive treatment models for individuals with autism spectrum disorders. *Journal of Autism and Developmental Disorders*. 40(4): 425–436.
- Perry A, Cummings A, Geier JD, et al. (2008) Effectiveness of intensive behavioral intervention in a large, community-based program. *Research in Autism Spectrum Disorders* 2: 621–642.
- Proctor E, Landsverk J, Aarons G, et al. (2009) Implementation research in mental health services: an emerging science with conceptual, methodological, and training challenges. *Administration and Policy in Mental Health* 36(1): 24–34.
- Reichow B (2011) Development, procedures and application of the evaluative method for determining evidence-based practices in autism. In: Reichow B, Doehring P, Cicchetti D, et al. (eds) *Evidence Based Practices and Treatment for Children with Autism*. New York: Springer, pp. 25–40.
- Rogers S (1998) Empirically supported comprehensive treatments for young children with autism. *Journal of Clinical Child Psychology* 27(2): 168–179.
- Rogers S and Vismara L (2008) Evidence-based comprehensive treatments for early autism. *Journal of Clinical Child and Adolescent Psychology* 37(1): 8–38.
- Sable J, Plotts C and Mitchell L (2010) *Characteristics of the 100 largest public elementary and secondary school districts in the United States: 2008–09*. NCES 2011-301. Washington, DC: US Department of Education, National Center for Education Statistics.
- Sashegyi A, Brown S and Farrell P (2000) Application of generalized random effects regression models for cluster-correlated longitudinal data to a school-based smoking prevention trial. *American Journal of Epidemiology* 152: 1192–1200.
- Schreibman L (2000) Intensive behavioral/psychoeducational treatments for autism: research needs and future directions. *Journal of Autism and Developmental Disorders* 30(5): 373–378.
- Sheinkopf S and Siegel B (1998) Home-based behavioral treatment of young children with autism. *Journal of Autism and Developmental Disorders* 28(1): 15–23.
- Simpson R (2005) Evidence-based practices and students with autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities* 20(3): 140–149.
- Stahmer A, Collings N and Palinkas L (2005) Early intervention practices for children with autism: descriptions from community providers. *Focus on Autism and Other Developmental Disabilities* 20(2): 66–79.
- Van De Ven A and Johnson P (2006) Knowledge for theory and practice. *Academy of Management Review* 31(4): 802–821.
- Webb C, DeRubeis R and Barber J (2010) Therapist adherence/competence and treatment outcome: a meta-analytic review. *Journal of Consulting and Clinical Psychology* 78(2): 200–211.
- Weisz J, Chu B and Polo A (2004) Treatment dissemination and evidence-based practice: strengthening intervention through clinician-researcher collaboration. *Clinical Psychology: Science and Practice* 11: 300–307.
- Weisz J, Donenberg G, Han S, et al. (1995) Child and adolescent psychotherapy outcomes in experiments versus clinics: why the disparity? *Journal of Abnormal Child Psychology* 23(1): 83–106.
- Weisz J, Doss A and Hawley K (2005) Youth psychotherapy outcomes research: a review and critique of the evidence base. *Annual Review of Psychology* 56: 337–363.
- Wilczynski S and Christina C (2008) The National Standards Project: promoting evidence-based practices in autism spectrum disorders. In: Luiselli J, Russo D and Christian V (eds) *Effective Practices for Children with Autism: Education and Behavioral Support Interventions that Work*. New York: Oxford University Press, pp. 37–60.
- Yell M and Drasgow E (2000) Litigating a free appropriate public education: the Lovaas hearings and cases. *The Journal of Special Education* 33(4): 205–214.

Copyright of Autism: The International Journal of Research & Practice is the property of Sage Publications, Ltd. 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.