

# Dismantling the Active Ingredients of an Intervention for Children with Autism

Melanie Pellecchia<sup>1</sup> · James E. Connell<sup>2</sup> · Rinad S. Beidas<sup>1</sup> · Ming Xie<sup>1</sup> · Steven C. Marcus<sup>1</sup> · David S. Mandell<sup>1</sup>

Published online: 25 April 2015  
© Springer Science+Business Media New York 2015

**Abstract** This study evaluated the association of fidelity to each of the components of the Strategies for Teaching based on Autism Research (STAR) program, a comprehensive treatment package for children with autism that includes discrete trial training, pivotal response training, and teaching in functional routines, on outcomes for 191 students ages 5–8 years in a large public school district. Fidelity to all components was relatively low, despite considerable training and support, suggesting the need to develop new implementation strategies. Fidelity to pivotal response training, but not discrete trial training or functional routines, was positively associated with gains in cognitive ability despite low levels of fidelity, and may be an effective intervention choice in under-resourced settings.

**Keywords** Treatment packages · Autism · Treatment fidelity · Implementation science

## Introduction

Early and intensive treatment programs based on applied behavior analysis (ABA) have been repeatedly cited as the treatment model with the most evidence to support effectiveness (National Autism Center 2009; National Research Council 2001; Weitlauf et al. 2014), although significant

variation in outcomes among different children has been noted (Schreibman and Anderson 2011). ABA-based treatment techniques are numerous, and range from highly structured didactic approaches, such as discrete trial training (DTT; Lovaas 1987), to less structured naturalistic approaches, such as pivotal response training (PRT; Koegel et al. 1999), and teaching within the natural environment and daily routines (Arick et al. 2003; Hart and Risley 1980). Considerable evidence supports the efficacy of each of these approaches (Eikeseth et al. 2002; Kane et al. 2010; Koegel et al. 1999a; b; Lovaas 1987).

Given that most children with autism receive the majority of their treatment in school, classroom-based treatment programs are particularly important. Many school districts have adopted comprehensive classroom-based treatment programs for children with autism. Comprehensive programs (e.g., Early Start Denver Model, Strategies for Teaching based on Autism Research, Competent Learner Model) include multiple instructional components, delivered using a variety of instructional techniques (Arick et al. 2003; Dawson et al. 2010; Lopata et al. 2012; Odom et al. 2010; Tucci et al. 2005). For example, programs may include both structured and naturalistic instructional techniques and may be used with groups or individual students. All of these instructional techniques require significant staff training and time to implement correctly. Perhaps as a result, there is substantial variation in the extent to which these comprehensive programs and their individual components are implemented with fidelity in community settings (Bibby et al. 2001; Boyd and Corley 2001; Stahmer et al. 2005).

Researchers have presented multiple definitions and conceptualizations of treatment fidelity. Treatment fidelity is most often described as adherence, or how much the treatment procedures are followed as intended (Gresham

---

✉ Melanie Pellecchia  
pmelanie@upenn.edu

<sup>1</sup> Perelman School of Medicine, University of Pennsylvania, 3535 Market St, 3rd Floor, Philadelphia, PA 19104, USA

<sup>2</sup> AJ Drexel Autism Institute, Drexel University, 3020 Market St, Suite 560, Philadelphia, PA 19104, USA

et al. 1993; Pereplechikova et al. 2007). Other definitions of fidelity include competence, defined as staff skills related to implementation (Pereplechikova et al. 2007), and intensity or dosage, often measured as hours per week of treatment delivery (Peters-Scheffer et al. 2010). There is some disagreement on the best way to measure fidelity. Challenges related to the cost and time associated with direct observation measures of fidelity often necessitate the use of indirect methods in implementation research, although direct observation is generally considered the gold standard (Beidas et al. 2014).

There is little research on the association between fidelity, regardless of how it is defined (staff adherence, staff skill, intervention intensity, or some combination), and outcomes for children with autism. Studies for children with autism tend to measure only treatment intensity, particularly hours per week, finding that greater intensity is associated with improved outcomes (Eikeseth et al. 2002; Granpeesheh et al. 2011; Lovaas 1987; Smith et al. 2000). A significant minority of studies, however, have found little to no differences in outcome as a function of intervention intensity (Sallows and Graupner 2005; Turner and Stone 2007; Vismara et al. 2009). Substantial differences in sample size, methods, and outcome measures likely contributed to these discrepant results. The general consensus in the field is that greater intervention intensity is likely to improve outcomes (National Research Council 2001), although the optimal levels of intensity remain unknown.

It also has been hypothesized that improved accuracy of treatment implementation will result in improved outcomes for children with autism (Detrich 1999), although few studies have examined this association. The limited studies examining this question suggest that greater intervention accuracy is associated with improved outcomes, such as improved social, cognitive, and language skills (Lopata et al. 2012; Strain and Bovey 2011). A larger-scale study that used a monthly-administered, global measure of fidelity that assessed accuracy found a nonlinear association between fidelity and improvements in cognitive ability that varied by type of intervention (Mandell et al. 2013). These studies usually do not account for treatment intensity, either because it was not measured or because it did not vary in the particular study.

Most conceptual literature on treatment fidelity suggests that it is composed of multiple components (Pereplechikova et al. 2007); yet prior studies of the effect of treatment fidelity on outcomes for children with autism did not account simultaneously for treatment intensity and accuracy of implementation. Prior research also rarely examined fidelity of the specific intervention components that comprise a comprehensive treatment package, and instead provided a global measure of overall fidelity (Lopata et al. 2012; Mandell et al. 2013). Some of these

studies may have had limited ability to address these questions because of small sample size (Lopata et al. 2012; Strain and Bovey 2011). The field of autism treatment research lags behind other fields in this regard, as this systematic dismantling of treatment packages has effectively identified the core components of interventions designed for other populations, resulting in cost-effective and individualized mental health care (Chaffin et al. 2011; Damschroder et al. 2009; Micco et al. 2007; Van Brunt 2000).

The purpose of the present study was to evaluate the association of treatment fidelity to each of three components of a comprehensive treatment package with outcome for elementary school-aged children with autism. This study builds upon a recently completed randomized field trial (Mandell et al. 2013) of the Strategies for Teaching based on Autism Research (STAR) Program (Arick et al. 2004). The STAR program is a comprehensive, manualized program for children with autism that includes three types of ABA-based interventions: discrete trial training, pivotal response training, and teaching within functional routines. For the current study, we evaluated differences in outcome as a function of the amount of intensity and accuracy with which teachers implemented each of the three types of behavioral interventions comprising STAR. The current study expands upon previous research evaluating fidelity to the STAR program (Mandell et al. 2013) by systematically examining the association between fidelity to each of the components of STAR (rather than the whole program) and student outcomes, measuring intensity as well as accuracy (only accuracy was measured in the 2013 study), and by relying on a larger sample of students receiving STAR.

Each treatment component of any comprehensive treatment package like STAR has an associated cost to implement and sustain. Data on specific associations between intervention components and outcome can increase the cost effectiveness of these programs, direct implementation strategies, and potentially increase sustainability by reducing teacher burden.

## Method

### Participants

Participants from the third and final year of a randomized trial conducted in partnership with the School District of Philadelphia (Mandell et al. 2013) were included in this study. In this year of the study, all participating classroom staff received training in STAR, which is the intervention the district chose to implement after reviewing the year 1 study results. This is the only year in which intervention intensity data was collected, in response to a new set of study questions; therefore only data from the third year was

available for the current study. The two sets of participants, teachers ( $N = 54$ ) and students ( $N = 191$ ), were in kindergarten-through-second grade autism support classrooms. Teachers ranged in age from 22 to 64, with a mean age of 37. The majority of teachers were female (85 %) and Caucasian (67 %); 24 % were African-American, 2 % were Asian, 2 % identified as bi-racial, and 5 % of teachers did not provide race or ethnicity information. Mean student age was 6 years 3 months ( $SD .87$  years, range 5–8 years). The students were ethnically diverse; 53 % were African-American, 30 % Caucasian, 10 % Hispanic, 5 % Asian, and 2 % were of other ethnicities. Consistent with the epidemiology of autism, 87 % were male (see Table 1). Three quarters of the students were eligible for free or reduced-price lunches. All participants had an educational classification of autism provided by a licensed practitioner; diagnosis was confirmed using the Autism Diagnostic Observation Schedule (ADOS; Lord et al. 1999), administered by a research-reliable clinician.

### Program Overview

The STAR program (Arick et al. 2004) is a comprehensive, manualized program for children with autism that includes three types of ABA-based interventions: discrete trial training, pivotal response training, and teaching within functional routines. The curriculum is divided into six major areas: expressive language, receptive language, spontaneous language, functional routines, pre-academic concepts, and play and social interaction skills. The program has three levels of instruction to meet the needs of children up to 8 years of age at different developmental stages, and provides lesson plans for highly specific activities designed to improve skills in various curriculum

areas. The lesson plans prescribe the intervention type (DT/PRT/FR) to be used for each specific lesson.

Discrete trial training (DT) is implemented using an intensive one-to-one teaching session in a highly structured setting free from distractions. Discrete trial training generally involves mass trials, or the repeated practice of the same response for several successive teaching episodes and the use of reinforcers that are functionally unrelated to the response (e.g., providing access to a small amount of food for correctly identifying a car). Instruction within DT involves breaking down complex skills into small component parts, and teaching each component part individually. For example, to teach a student with autism to play appropriately with toys, an instructor may first teach them to imitate actions with objects, such as pushing a toy car when provided with an imitative cue.

Pivotal response training (PRT) typically consists of loosely structured sessions that are initiated and paced by the child, take place in a variety of locations, and employ a variety of teaching materials. During PRT the child chooses the instructional object or activity, and the reinforcer is related to the response (e.g., providing access to a car for correctly identifying a car; Arick et al. 2004). Instruction within PRT involves following the child's lead and capturing and contriving teachable moments related to the context.

Functional routines are predictable activities with an expected sequence of steps that occur naturally throughout the day (Arick et al. 2004). Examples of functional routines common to most school-aged children include arriving into the classroom, transitioning between activities, using the bathroom, and having a snack. Functional routines instruction involves providing systematic prompts and cues to teach the child to participate independently in common school and self-care routines (Arick et al. 2004).

The STAR program recommends that each student receive at least two DT sessions and at least one PRT session per day. Functional routines instruction occurs naturally throughout the daily activities, and it is recommended that each student receive targeted instruction on one-to-two functional routines per day.

### Teacher Training

Teachers and classroom staff received intensive training and support on the implementation of the STAR program as part of the parent study by the developers of the STAR program (Mandell et al. 2013). The STAR program developers trained consultants who provided frequent and regular teacher training and classroom support. Training included 3 days of intensive workshops at the start of the school year, hands-on work in the classrooms with teachers to set up classrooms and plan student lessons at the start of

**Table 1** Descriptive characteristics of the sample

Characteristic	Students ( $N = 191$ )	Teacher ( $N = 54$ )
Chronological age		
Mean age (SD)	6.3 (.87)	37 (4.84)
Age range	5–8	22–64
Ethnicity		
African American (%)	53	24
Caucasian (%)	30	67
Hispanic (%)	10	0
Asian (%)	5	2
Other ethnicity (%)	2	2
Unknown ethnicity (%)	0	5
Gender (%)		
Male	87	15
Female	13	85

the school year, ongoing full-day quarterly workshops during the school year, and ongoing coaching provided in the classroom with classroom staff for 2–3 h per week, twice per month.

## Measures

### *Primary Independent Variables*

**Implementation Accuracy** Implementation accuracy for each intervention component was measured via hour-long monthly observations. Trained research assistants recorded the occurrence or non-occurrence of each step of each program component (discrete trial training, pivotal response training, functional routines) on a data collection form created for this study. The research assistants were trained to reliability with the first author on all steps of each of the three program components. Training involved didactic instruction on data coding procedures, role-playing, and practice coding. Training continued until research assistants were 80 % reliable with the first author. Accuracy was coded for each of the three components on a 4-point scale using criteria specific to each teaching technique: 0 (does not implement), 1 (poor use), 2 (somewhat accurate), 3 (mostly accurate), and 4 (highly accurate). Accuracy ratings of 3 or 4 indicated acceptable levels of implementation accuracy for each intervention component. Discrete trial training accuracy included gaining the child's attention, providing clear instructions, using specific prompting strategies, error correction procedures, and reinforcement. Pivotal response training accuracy included items such as gaining the child's attention, providing the child with a choice of tasks, incorporating a mixture of target and maintenance tasks, and using natural reinforcers. Functional routines accuracy was specific to the daily routines used in the classroom, such as lining up for transitions or sitting and attending to group instruction, and included the use of clear instructions, verbal and visual cues, and reinforcement.

**Intervention Intensity** Levels of intervention intensity for each student were monitored through teacher report for each of the instructional strategies in the STAR program (discrete trial training, pivotal response training, and functional routines). Classroom teachers reported how often they implemented the intervention with each student throughout the week. The discrete trial training and pivotal response training components of the intervention were delivered in intensive teaching sessions lasting approximately 30 minutes each session. Intervention intensity was defined as number of intervention sessions for these intervention components per week (e.g., 2 DT sessions per week). The functional routines component was

designed to be delivered throughout the day. For example the classroom routine of lining up was taught as the opportunity for lining up occurred naturally during the day. Intervention intensity was coded using a Likert scale ranging from 0 to 4 with the following criteria for each score: 0 (less than one time per week), 1 (one time per week), 2 (two to four times per week), 3 (one time per day), and 4 (two times per day).

**Composite Fidelity Score** A composite fidelity score variable was created because of the high correlation between intensity and accuracy for each intervention component ( $r = .53-.73$ ), and because fidelity is a function both of intensity and accuracy. The composite score was calculated as the product of intensity and accuracy for each technique for each student. For example, a teacher who implemented the intervention with high accuracy during scheduled observations (as indicated by an accuracy score of 4), but did not implement the intervention with a student during daily instruction (as indicated by an intensity score of 0), would receive a composite fidelity score of 0 ( $4 \times 0$ ). The scale for the composite fidelity score, therefore ranged from 0 to 16 for each treatment component ( $0-4$  on intensity  $\times 0-4$  on accuracy). This is a more accurate rating of the teacher's overall fidelity than that which would be captured by the sum. For example, when conducting a planned observation a teacher could get a four for accuracy. However, that teacher may never independently implement the intervention, and thus have a zero for intensity (e.g., 4 on accuracy + 0 on intensity) resulting in a summed composite score of four, even though the teacher never implemented the intervention when the researchers were not present. A composite fidelity rating for each component (discrete trial training, pivotal response training, and functional routines) was calculated, rather than an overall STAR fidelity rating, in order to examine the individual effects of fidelity to each treatment component on student outcome. The teachers' fidelity scores at the end of the academic year were used in the analyses.

### *Primary Dependent Variable*

**Child Outcome** The Differential Ability Scales, 2nd Edition (DAS-II; Elliott 1990) was used to measure child outcomes. The DAS-II was designed to assess a wide range of cognitive abilities in children aged 2 years 6 months through 17 years 11 months and has been used with children with autism in many previous studies (e.g., Anderson et al. 2007; Thurm et al. 2007). Licensed psychologists and doctoral level graduate students in psychology, trained to 80 % reliability, administered the DAS-II at the beginning and end of the school year. Child outcomes are commonly reported as change in overall cognitive ability or IQ in

autism intervention studies, especially outcome studies of behavioral interventions similar to the one used in the present study (e.g., Eikeseth et al. 2007; Lovaas 1987; and Smith 1999). The DAS-II is viewed as an appropriate assessment tool for evaluating the cognitive abilities of children with autism and other disabilities because it relies less on expressive language ability than other cognitive assessments, and therefore was selected as the outcome measure for the current study.

#### *Other Covariates of Interest*

**Autism Severity** Autism symptoms were measured using the Autism Diagnostic Observation Schedule (ADOS). The ADOS is a semi-structured standardized observational measure of social interaction, communication skills, restrictive and repetitive behavior, and play or imaginative use of materials for the assessment of autism spectrum disorders (Lord et al. 1999). The ADOS has strong psychometric properties, with high sensitivity and specificity reported for each of the four modules (80–94 %; Lord et al. 1999). ADOS scores were converted to the ADOS severity algorithm, a validated measure that allows for comparison of autism severity across the modules of the ADOS (Gotham et al. 2009). Clinicians trained to 80 % reliability with a research certified administrator administered the ADOS at the beginning of the school year.

Other teacher and student characteristics potentially associated with outcome were included in the analyses. These included teachers' years of experience teaching children with autism (measured by self-report) and student age (parent report). These characteristics were included in the analyses in order to identify potential child and teacher level characteristics that may be associated with student outcome.

#### **Data Analytic Plan**

Means, standard deviations, and ranges were calculated for all variables of interest. Correlations for intervention intensity and intervention accuracy were calculated within and between each intervention component (discrete trial training, pivotal response training, and functional routines). We then estimated differences in cognitive ability between baseline and follow-up using linear regression models with random effects for classroom. Separate models were used where the independent variables of interest in turn included intervention intensity, accuracy, and the composite fidelity score for each intervention component. This was done first controlling only for baseline cognitive ability score (our "unadjusted model") and then repeated in a fully adjusted model controlling for baseline cognitive ability scores, student age, autism severity score, and years of teacher experience.

#### **Results**

Descriptive characteristics for the sample are presented in Table 2. The mean composite fidelity score for discrete trial training was 3.6 out of 16 (range 0–12), mean composite fidelity for pivotal response training was 2.7 out of 16 (range 0–12), and mean composite fidelity for functional routines was 7.3 out of 16 (range 0–16). Mean discrete trial training intensity was 1.4 out of 4 (range 0–3.5), mean pivotal response training intensity was 1.5 out of 4 (range 0–3), and mean functional routines intensity was 2.8 out of 4 (range 0–4). Mean accuracy ratings for each intervention were: discrete trial training, 2.0 out of 4 (range 0–4), pivotal response training, 1.6 out of 4 (range 0–4), and functional routines, 2.4 out of 4 (range 0–4). Average number

**Table 2** Descriptive characteristics of independent variables by student

Variable	N	Mean	SD	Observed range	Possible range
DT fidelity	191	3.62	3.44	0–12	0–16
DT intensity	191	1.37	.99	0–3.5	0–4
DT accuracy	191	2.04	1.22	0–4	0–4
PRT fidelity	191	2.74	2.61	0–12	0–16
PRT intensity	191	1.47	.87	0–3	0–4
PRT accuracy	191	1.56	.96	0–4	0–4
FR fidelity	191	7.31	4.67	0–16	0–16
FR intensity	191	2.75	1.07	0–4	0–4
FR accuracy	191	2.36	1.07	0–4	0–4
Teacher experience in years	54	5.30	6.05	1–32.5	–
Student age in years	191	6.25	.87	5–8	–
DAS Time 1 standard score	191	62	21.57	30–111	–
DAS Time 2 standard score	191	66	20.90	30–122	–
Autism severity composite score	191	6.38	1.78	1–10	0–10



of years of teacher experience was 5.3 with a range of 1–32 years. Students' average cognitive score at the start of the academic year was 62.1. There was an average 4-point increase in cognitive ability following one academic year of intervention.

Table 3 presents the correlations between intervention intensity, intervention accuracy, and the composite fidelity score for each intervention component. Intervention intensity and accuracy were highly correlated for each intervention, and across interventions.

Table 4 presents the results of the regression analyses predicting improvements in cognitive ability. In unadjusted regression analyses the discrete trial training composite fidelity rating was associated with improvements in cognitive ability. However, discrete trial intensity and accuracy were not individually associated with improvements in cognitive ability. The pivotal response

training composite score and pivotal response training accuracy also were associated with outcome in unadjusted analyses. None of the remaining variables, including functional routines fidelity, were associated with improvements in cognitive ability in the unadjusted analyses. In adjusted regression analyses controlling for baseline cognitive ability, student age, autism severity, and years of teacher experience, the pivotal response training composite and intensity ratings were statistically significantly associated with improvements in cognitive ability. Pivotal response training accuracy was not associated with outcome. Discrete trial training and functional routines fidelity also were not associated with differences in student outcomes.

The association between the composite fidelity score for each intervention component and improvements in cognitive ability are visually displayed in Fig. 1.

**Table 3** Correlations between intervention intensity, accuracy, and composite fidelity scores for each intervention component

Component	DT composite	PRT composite	FR composite	DT intensity	PRT intensity	FR intensity	DT accuracy	PRT accuracy	FR accuracy
DT composite	–								
PRT composite	.83*	–							
FR composite	.62*	.55*	–						
DT intensity	.90*	.73*	.55*	–					
PRT intensity	.79*	.77*	.52*	.84*	–				
FR intensity	.47*	.40*	.88*	.52*	.55*	–			
DT accuracy	.78*	.66*	.75*	.67*	.60*	.59*	–		
PRT accuracy	.65*	.85*	.63*	.53	.53*	.48*	.73*	–	
FR accuracy	.66*	.60*	.92*	.60*	.53*	.73*	.83*	.66*	–

\*  $p < .05$

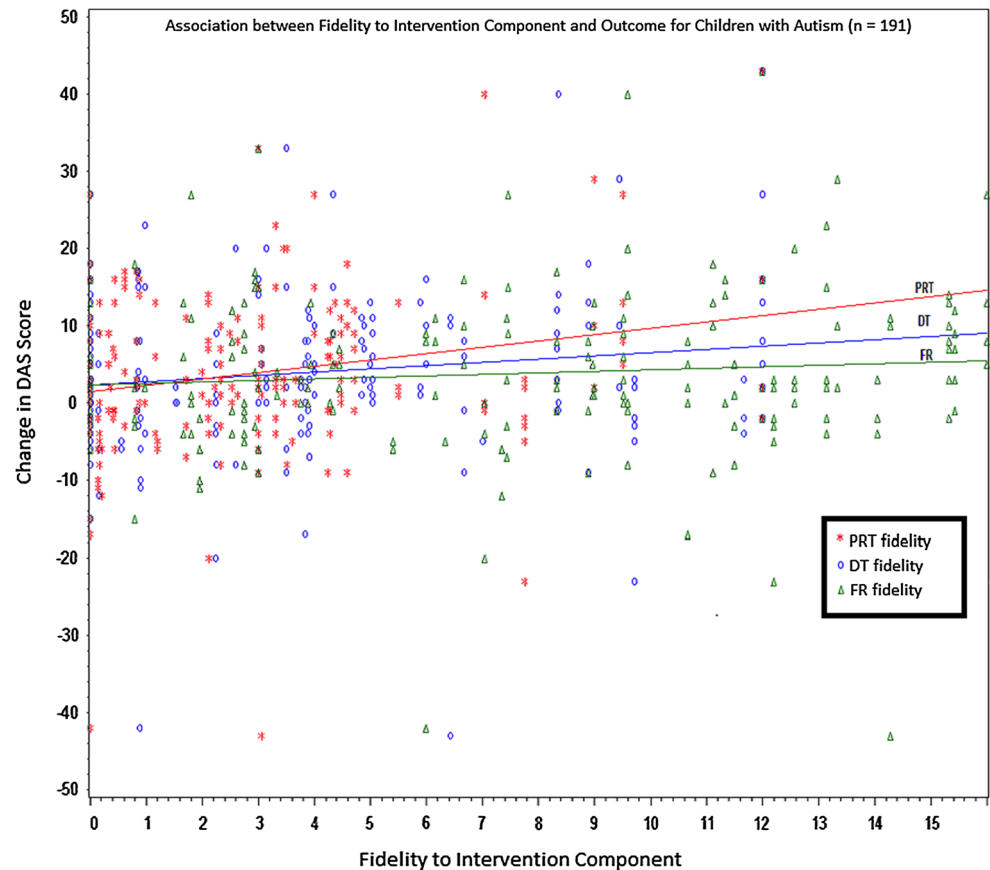
**Table 4** Linear regression predicting cognitive ability scores at follow-up in adjusted analyses

Variable	Unadjusted model		Adjusted model	
	<i>B</i>	95 % CI	<i>B</i>	95 % CI
DT fidelity	.37*	(−.06, .81)	.34	(−.10, .79)
DT intensity	1.31	(−.22, 2.85)	1.13	(−.48, 2.74)
DT accuracy	.83	(−.39, 2.05)	.84	(−.41, 2.08)
PRT fidelity	.63**	(.06, 1.19)	.61*	(.03, 1.20)
PRT intensity	2.23*	(.49, 3.97)	2.11*	(.31, 3.90)
PRT accuracy	1.28*	(−.22, 2.80)	1.31	(−.26, 2.88)
FR fidelity	.10	(−.21, .43)	.09	(−.24, .42)
FR intensity	.43	(−.97, 1.83)	.32	(−1.12, 1.76)
FR accuracy	.23	(−1.19, 1.65)	.14	(−1.31, 1.59)
Teacher experience	.07	(−2.5, .39)	–	–
Student age	−1.03	(−2.72, .64)	–	–
Autism severity	.07	(−.25, .39)	–	–

CI confidence interval. All adjusted models control for baseline cognitive ability scores, student age, autism severity score, and years of teacher experience

\*  $p < .05$ ; \*\*  $p < .01$

**Fig. 1** Association between teachers' fidelity to intervention components and outcome for children with autism (n = 191)



## Discussion

Children with autism often receive treatment in the form of comprehensive treatment packages, composed of multiple instructional strategies to remediate a wide-range of core deficits (Odom et al. 2010). These comprehensive treatment packages are typically time intensive, costly, and require significant training to implement with high fidelity (Arick et al. 2003; Dawson et al. 2010; Lopata et al. 2012). The complexity of these treatment packages, coupled with the resource challenges faced by many public schools, leads to substantial variation in how these models are implemented in community practice (Bibby et al. 2001; Boyd and Corley 2001; Stahmer et al. 2005). To date, little research has attempted to dismantle the core components of these comprehensive treatment packages to identify their active ingredients. Effectively identifying the core components of a treatment package for children with autism would facilitate the development of more cost-effective treatments that could be more feasibly implemented within public school settings, as has been demonstrated in the systematic dismantling of treatment packages designed for other populations (Chaffin et al. 2011; Damschroder et al. 2009; Micco et al. 2007; Van Brunt 2000). This study

represents the first attempt to systematically identify the active ingredients of a comprehensive treatment package for children with autism by examining the relative effectiveness of three different intervention strategies that comprise a comprehensive treatment model for children with autism. The results indicate that fidelity to pivotal response training is associated with significant increases in cognitive ability after one academic year. The magnitude of our findings for pivotal response training and student outcomes was small, but is consistent with those found in other community-based trials (Weisz et al. 2013). Previous research has suggested that the magnitude of findings may be diminished in samples with psychopathology severe enough to meet criteria for a DSM diagnosis, as was present in our sample (TADS 2004). Examination of the estimated increase in IQ score associated with increases in pivotal response training fidelity suggests that increasing fidelity to pivotal response training can lead to improvements in overall cognitive functioning. These findings are consistent with a recent randomized comparison of pivotal response training and discrete trial training, in which pivotal response training was more effective in improving outcomes for children with autism (Mohammadzahari et al. 2014).

Several additional findings are worth noting. First, there was substantial variability in outcome across students, with some students demonstrating significantly large increases or decreases in cognitive ability scores. This variability in outcome for children with autism is consistent with previous outcome studies of behavioral interventions for children with autism (Eikeseth et al. 2002; Lovaas 1987; Schreibman and Anderson 2011; Smith et al. 2000). The reasons for this variability are unclear, but may be related to changes in behavior or test-taking ability, rather than a true change in cognitive ability of 20–40 points over one academic year. To account for these potential outliers we conducted a sensitivity analysis in which we dropped outliers as defined by change scores greater than 30 points and found no difference in the association of the independent variables with student outcome. Another interesting finding was that increased student age was not associated with differences in student outcome, contrary to previous findings that indicate decreased gains associated with increased student age (Mandell et al. 2013). As previously mentioned, data used in this study were selected from the third year of a larger intervention study. It is possible that involvement in the intervention study during previous years affected the association between student age and outcome in that (1) teachers may have become more skilled at teaching older students with autism, (2) older students, who may have been receiving intensive intervention for several years, had reached their maximum response to such interventions, and (3) there may have been more higher functioning older students in the autism support classrooms in the third year of the intervention than in previous years. The third noteworthy finding was that differences in autism severity were not associated with differences in student outcome. This finding suggests that autism severity does not inhibit gains associated with behavioral interventions, and that these interventions are likely beneficial to children at varying levels of severity. It may also reflect the unstable relationship between cognitive ability and autism symptom severity, as evidenced by the heterogeneous presentation of autism associated with varying levels of autism severity and cognitive ability across the autism spectrum (Charman et al. 2011).

Several study limitations were present. First, the fidelity measures used in the study did not undergo rigorous assessments of reliability and validity. Measuring intervention fidelity is considered a benchmark of efficacy trials, but fidelity is more difficult to measure in effectiveness and implementation trials. The field of implementation science is a nascent field, and unfortunately a gold standard for measuring fidelity has not been established in this field; therefore we were not able to assess the validity of the fidelity measure used in this study. A second limitation is the outcome measure used in the study. Although changes

in overall cognitive ability are commonly used as outcomes measures for intervention studies with children with autism, this outcome measure may not have captured some gains achieved by the treatment package (Connell et al. 2014). For example, a measure of cognitive ability may not capture changes that result from an intervention that targets improvements in a child's ability to participate in daily functional routines (e.g., transitions between activities); thus student gains that resulted from fidelity to the functional routines portion of the STAR program may not have been adequately measured using the DAS-II. A third limitation is that many factors not measured in this study, such as organizational-level factors (i.e., administrative support, staffing ratios, team cohesion, and building resources) and teacher-level factors (i.e., teaching philosophy), have been associated with treatment fidelity in schools (Beets et al. 2008; Dingfelder and Mandell 2011). These additional variables were not evaluated in this study, but could also affect student outcomes, and therefore may have confounded the observed association between treatment fidelity and outcomes.

Despite these limitations, the findings have important implications for increasing the effectiveness of autism interventions in schools. Of particular concern, high fidelity was not achieved for any of the intervention components in these under-resourced urban classrooms, consistent with evaluations of treatment fidelity for evidence-based practices in many high-need, urban classroom settings (Mandell et al. 2013). Although fidelity varied across teachers, none of the classrooms achieved the highest possible fidelity score for pivotal response training or discrete trial training, and only four of the 54 classrooms achieved the highest possible fidelity score for functional routines. Treatment intensity for each of the intervention components was much lower than that usually recommended for children with autism (NRC 2001; Smith et al. 2000). There are several possible explanations for the observed variability in fidelity of implementation across teachers, despite consistent amounts of training and coaching provided to all teachers. For example, teachers' attitudes and beliefs about the intervention, level of administrative support, and student disruptive behavior may have been associated with both fidelity of implementation and student outcomes. Factors such as teachers' beliefs, administrative support, and student characteristics likely represent barriers to improving treatment fidelity within school settings and further support the need to identify treatments that can be feasibly implemented in schools and other community settings.

Despite these low levels of overall fidelity, increases in cognitive ability were observed overall, and were particularly associated with greater fidelity to the pivotal response training composite score and intensity, but were not associated with pivotal response training accuracy. Of the



three intervention components, pivotal response training had the lowest overall fidelity ratings, suggesting that it was the most difficult component to implement. If pivotal response training results in gains for children with autism, even when implemented at lower levels of intensity and accuracy, it may be a particularly effective intervention choice for adoption in under-resourced autism support classroom settings. One possible explanation for this finding is that the play-based teaching procedures employed in pivotal response training, which closely resemble naturalistic learning opportunities, have been shown to produce greater generalization of skills learned compared with more didactic teaching approaches for children with autism (Delprato 2001). This enhanced generalization associated with pivotal response training may facilitate a transfer of learned skills to increases in overall cognitive ability. Another possible explanation is that teachers who implement pivotal response training may be more philosophically aligned with the teaching approaches used in this treatment strategy, and thus may incorporate the procedures throughout their daily teaching procedures, leading to enhanced learning. Additionally, it is possible that higher levels of fidelity to pivotal response training indicate classroom teams that are better organized and demonstrate higher levels of classroom cohesion. These classroom environments likely facilitate greater individual instructional time and exposure to high quality teaching practices. Investigations of intervention fidelity to treatment packages for children with autism that also evaluate factors related to overall classroom climate will provide important information regarding classroom practices that are most likely to improve student outcomes.

Fidelity to discrete trial training was only marginally associated with gains in cognitive ability. It may be that the complexity and scripted teaching procedures found in this treatment approach require higher levels of fidelity to achieve positive outcomes. Given the lower levels of fidelity found in our study and in previous community-based evaluations (Mandell et al. 2013; Stahmer et al. 2005), it is possible that the levels of training and fidelity needed to implement discrete trial training effectively may be difficult for many public school teachers.

The findings related to teaching within functional routines may be related to the outcome measure used in our study. As discussed in the limitations, a measure of overall cognitive ability may not adequately capture outcomes associated with skills taught related to daily routines. Therefore, it is unclear whether high levels of fidelity to this treatment component would be associated with improved outcomes using a different outcome measure.

Two possible solutions to the problem of low treatment fidelity in public education and community settings are relevant to these findings and warrant future research. First,

the teachers in this study implemented the treatment package with low fidelity, despite considerable training and support, suggesting the need to develop new strategies to increase treatment fidelity. Performance feedback (PFB) is an example of a consultation approach that has been shown to increase treatment fidelity in high-need educational settings (Pellecchia et al. 2011), however it has yet to be applied to the systematic evaluation of a comprehensive treatment package. Future research evaluating the effectiveness of performance feedback to increase fidelity to comprehensive treatment packages may be an important next step in this line of research. Second, identifying the core components of these comprehensive treatments that produce maximum improvements in children's outcomes will enable under-resourced settings to implement effective interventions efficiently. The results of this study indicate that pivotal response training may be an example of this type of effective and efficient intervention that could be effectively embedded into current educational practices. Continued research that systematically dismantles the core components of comprehensive treatment packages will enable the widespread adoption of evidence-based interventions for children with autism, as well as other diagnoses, in public school and community-based settings.

**Acknowledgments** This research was conducted as partial fulfillment of the first author's doctoral dissertation research requirements. This work was supported by funding from the following Grants from the National Institute of Health: 1R01MH083717 and the Institute of Education Sciences: R324A080195 to Mandell. Additionally, the preparation of this article was supported in part by the following Grant from the National Institute of Health: K23 MH099179 to Beidas.

**Conflict of interest** The authors declare they have no conflict of interest.

**Ethical standard** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

**Informed consent** Informed consent was obtained from all individual participants included in the study.

## References

- Anderson, D. K., Lord, C., Risi, S., DiLavore, P. S., Shulman, C., Thurm, A., et al. (2007). Patterns of growth in verbal abilities among children with autism spectrum disorder. *Journal of Consulting and Clinical Psychology*, 75(4), 594–604.
- Arick, J., Loos, L., Falco, R., & Krug, D. (2004). *The STAR program: Strategies for teaching based on autism research*. Austin, TX: PRO-ED.
- Arick, J. R., Young, H. E., Falco, R. A., Loos, L. M., Krug, D. A., Gense, M. H., & Johnson, S. B. (2003). Designing an outcome study to monitor the progress of students with autism spectrum

- disorders. *Focus on Autism and Other Developmental Disabilities*, 18(2), 75–87.
- Beets, M. W., Flay, B. R., Vuchinich, S., Acock, A. C., Li, K. K., & Allred, C. (2008). School climate and teachers' beliefs and attitudes associated with implementation of the positive action program: A diffusion of innovations model. *Prevention Science*, 9(4), 264–275.
- Beidas, R. S., Cross, W., & Dorsey, S. (2014). Show me, don't tell me: Behavior rehearsal as a training and analog fidelity tool. *Cognitive and Behavioral Practice*, 21, 1–11.
- Bibby, P., Eikeseth, S., Martin, N. T., Mudford, O. C., & Reeves, D. (2001). Progress and outcomes for children with autism receiving parent-managed intensive interventions. *Research in Developmental Disabilities*, 22(6), 425–447.
- Boyd, R. D., & Corley, M. J. (2001). Outcome survey of early intensive behavioral intervention for young children with autism in a community setting. *Autism*, 5(4), 430–441.
- Chaffin, M., Funderburk, B., Bard, D., Valle, L. A., & Gurwitch, R. (2011). A combined motivation and parent-child interaction therapy package reduces child welfare recidivism in a randomized dismantling field trial. *Journal of Consulting and Clinical Psychology*, 79(1), 84–95.
- Charman, T., Pickles, A., Simonoff, E., Chandler, S., Loucas, T., & Baird, G. (2011). IQ in children with autism spectrum disorders: Data from the special needs and autism project (SNAP). *Psychological Medicine*, 41(3), 619–627.
- Connell, J. E., Pellecchia, M., & Vornträn, C. M. (2014). Classroom interventions for youth with pervasive developmental disorders/autism spectrum disorders. In M. Weist, N. Lever, & J. Owens (Eds.), *Handbook of school mental health, issues in clinical child psychology* (427.440) (2nd ed.). New York: Springer.
- Damschroder, L. J., Aaron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., & Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. *Implementation Science*, 4, 50.
- Dawson, G., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J., & Varley, J. (2010). Randomized, controlled trial of an intervention for toddlers with autism: The early start Denver model. *Pediatrics*, 125(1), e17–e23.
- Delprato, D. J. (2001). Comparisons of discrete-trial and normalized behavioral language intervention for young children with autism. *Journal of Autism and Developmental Disorders*, 31(3), 315–325.
- Detrich, R. (1999). Increasing treatment fidelity by matching interventions to contextual variables within the educational setting. *School Psychology Review*, 28(4), 608–620.
- Dingfelder, H. E., & Mandell, D. S. (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.
- Eikeseth, S., Smith, T., Jahr, E., & Eldevik, S. (2002). Intensive behavioral treatment at school for children with autism: A 1-year Comparison controlled study. *Behavior Modification*, 31, 264–278.
- Eikeseth, S., Smith, T., Jahr, E., & Eldevik, S. (2007). Outcome for children with autism who began intensive behavioral treatment between ages 4 and 7: A Comparisons controlled study. *Behavior Modification*, 31(3), 264–278.
- Elliott, C. D. (1990). *Differential ability scales* (2nd ed.). Pearson: San Antonio.
- Gotham, K., Pickles, A., & 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.
- Granpeesheh, D., Kenzer, A., & Tarbox, A. (2011). Comparison of two-year outcomes for children with autism receiving high and low intensity behavioral intervention. *European Psychiatry*, 26, 1839.
- Gresham, F. M., Gansle, K. A., Noell, G. H., Cohen, S., & Rosenblum, S. (1993). Treatment integrity of school-based behavioral intervention studies: 1980–1990. *School Psychology Review*, 22, 254–272.
- Hart, B., & Risley, T. R. (1980). In vivo language intervention: Unanticipated general effects. *Journal of Applied Behavior Analysis*, 13(3), 407–432.
- Kane, M., Connell, J. E., & Pellecchia, M. (2010). A quantitative analysis of language interventions for children with autism. *The Behavior Analyst Today*, 11, 128–144.
- Koegel, R. L., Koegel, L. K., & Carter, C. M. (1999a). Pivotal teaching interactions for children with autism. *School Psychology Review*, 28(4), 576–594.
- Koegel, L. K., Koegel, R. L., Harrower, J. K., & Carter, C. M. (1999b). Pivotal response intervention I: Overview of approach. *Journal of the Association for Persons with Severe Handicaps*, 24(3), 174–185.
- Lopata, C., Thomeer, M. L., Volker, M. A., Lee, G. K., Smith, T. H., Smith, R. A., & Toomey, J. A. (2012). Feasibility and initial efficacy of a comprehensive school-based intervention for high-functioning autism spectrum disorders. *Psychology in the Schools*, 49(10), 963–974.
- Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. (1999). *Autism diagnostic observation schedule manual*. Los Angeles, CA: Western Psychological Services.
- Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology*, 55(1), 3–9.
- 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(3), 281–295.
- Micco, J. A., Choate-Summers, M., Ehrenreich, J. T., Pincus, D. B., & Mattis, S. G. (2007). Efficacious treatment components of panic control treatment for adolescents: A preliminary examination. *Child & Family Behavior Therapy*, 29(4), 1–23.
- Mohammadzahari, F., Koegel, L. K., Rezaee, M., & Rafiee, S. M. (2014). A randomized clinical trial comparison between pivotal response treatment (PRT) and structured applied behavior analysis (ABA) intervention for children with autism. *Journal of Autism and Developmental Disorders*, 44(11), 2769–2777.
- National Autism Center. (2009). *The National Standards Report*. Retrieved December 1, 2013 from [http://www.nationalautismcenter.org/pdf/NAC%20NSP%20Report\\_FIN.pdf](http://www.nationalautismcenter.org/pdf/NAC%20NSP%20Report_FIN.pdf)
- National Research Council. (2001). Educating children with autism. Committee on educational interventions for children with autism. In C. Lord & J. P. McGee (Eds.), *Division of behavioral and social sciences and education*. Washington DC: National Academy Press.
- 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(4), 425–436.
- Pellecchia, M., Connell, J. E., Eisenhart, D., Kane, M., Schoener, C., Turkel, K., & Mandell, D. S. (2011). We're all in this together now: Group performance feedback to increase classroom team data collection. *Journal of School Psychology*, 49, 411–431.
- Perepletchikova, F., Treat, T. A., & Kazdin, A. E. (2007). Treatment integrity in psychotherapy research: Analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology*, 75(6), 829–841.

- Peters-Scheffer, N., Didden, R., Mulders, M., & Korzilius, H. (2010). Low intensity behavioral treatment supplementing preschool services for young children with autism spectrum disorders and severe to mild intellectual disability. *Research in Developmental Disabilities*, 31(6), 1678–1684.
- Sallows, G. O., & Graupner, T. D. (2005). Intensive behavioral treatment for children with autism: Four-year outcome and predictors. *American Journal on Mental Retardation*, 110, 417–438.
- Schreibman, L., & Anderson, A. (2011). Focus on integration: The future of the behavioral treatment of autism. *Behavior Therapy*, 32, 619–632.
- Smith, T. (1999). Outcome of early intervention for children with autism. *Clinical Psychology: Science and Practice*, 6(1), 33–49.
- Smith, T., Groen, A. D., & Wynn, J. W. (2000). Randomized trial of intensive early intervention for children with pervasive developmental disorder. *American Journal on Mental Retardation*, 105(4), 269–285.
- Stahmer, A. C., Collings, N. M., & Palinkas, L. A. (2005). Early intervention practices for children with autism: Descriptions from community providers. *Focus on Autism and Other Developmental Disabilities*, 20(2), 66–79.
- Strain, P. S., & Bovey, E. H. I. I. (2011). Randomized controlled trial of the LEAP model of early intervention for young children with autism spectrum disorders. *Topics in Early Childhood Special Education*, 31, 133–154.
- Thurm, A., Lord, C., Lee, L., & Newschaffer, C. (2007). Predictors of language acquisition in preschool children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 37(9), 1721–1734.
- Treatment for Adolescents with Depression Study (TADS). (2004). Fluoxetine, cognitive-behavioral therapy, and their combination for adolescents with depression: Randomized Controlled Trial. *Journal of the American Medical Association*, 292, 807–820.
- Tucci, V., Hursh, D., Laitinen, R., & Lambe, A. (2005). Competent learner model for individuals with Autism/PDD. *Exceptionality*, 13(1), 55–63.
- Turner, L. M., & Stone, W. L. (2007). Variability in outcome for children with an ASD diagnosis at age 2. *Journal of Child Psychology and Psychiatry*, 48, 793–802.
- Van Brunt, D. L. (2000). Modular cognitive-behavioral therapy: Dismantling validated treatment programs into self-standing treatment plan objectives. *Cognitive and Behavioral Practice*, 7(2), 156–165.
- Vismara, L. A., Colombi, C., & Rogers, S. J. (2009). Can one hour per week of therapy lead to lasting changes in young children with autism? *Autism*, 13(1), 93–115.
- Weisz, J. R., Kuppens, S., Eckshtain, D., Ugueto, D. M., Hawley, K. M., & Jensen-Doss, A. (2013). Do evidence-based youth psychotherapies outperform usual clinical care? A multilevel meta-analysis. *JAMA Psychiatry*, 70, 1–24.
- Weitlauf, A. S., McPheeters, M. L., Peters, B., Sathe, N., Travis, R., Aiello, R., Jerome R., & Warren, Z. (2014). Therapies for children with autism spectrum disorder: Behavioral interventions update. Comparative Effectiveness Review No. 137. (Prepared by the Vanderbilt Evidence-based Practice Center under Contract No. 290-2012-00009-I.) AHRQ Publication No. 14-EHC036-EF. Rockville, MD: Agency for Healthcare Research and Quality.