Title
The Effect of Implementation Climate on Program Fidelity and Student Outcomes in Autism Support Classrooms

Journal
Journal of Consulting and Clinical Psychology, 87(3)

ISSN
0022-006X

Authors
Kratz, Hilary E
Stahmer, Aubyn
Xie, Ming
et al.

Publication Date
2019-03-01

DOI
10.1037/ccp0000368

Peer reviewed

Running head: EFFECT OF IMPLEMENTATION CLIMATE

The Effect of Implementation Climate on Program Fidelity and Student Outcomes in Autism Support Classrooms

Hilary E. Kratz[1,2], Aubyn Stahmer[3], Ming Xie[2], Steven C. Marcus[2], Melanie Pellecchia[2], Jill Locke[4], Rinad Beidas[2], & David S. Mandell[2]

[1]La Salle University

[2] University of Pennsylvania

[3] University of California Davis MIND Institute

[4]University of Washington

Correspondence concerning this article should be addressed to:

Hilary E. Kratz, Ph.D.
1900 W. Olney Avenue
Philadelphia, PA 19141
Phone: 215-951-1379
Fax: 215-951-5140
Email: kratz@lasalle.edu

Author Note: This study was drawn from the Philadelphia Autism Instructional Methods Study, which was supported by NIMH grant 1R01MH083717-01A1 and IES grant R324A08195. This study also was funded by the NIMH F31-MH088172 grant awarded to the first author. We thank the children, parents, schools, teachers and classroom staff who participated, and the research associates who contributed countless hours of assessments, intervention, supervision, and data collection: Erica Blanch, Colleen Harker, Rukiya Wideman, Margaret Downey, Megan Carolan Tomkinson, Emily Bernabe, Laura MacMullen Freeman, Jessica Dodge, and Shawn Gilroy.

The Effect of Implementation Climate on Program Fidelity and Student Outcomes in Autism Support Classrooms

Re-submission date: September 20, 2018

Abstract

Objective: An organization's implementation climate, or the extent to which use of an intervention is expected, supported, and rewarded by colleagues and supervisors, has been identified as critical to successful intervention implementation and outcomes. The effect of implementation climate has not been well studied in special education settings. The present study examines the association between teachers' perceptions of implementation climate, teacher fidelity to a school-based program for students with autism, and student outcomes (measured as changes in IQ) over time. ***Method***: Participants included 158 students from 45 classrooms and their teachers. Teachers provided a measure of implementation climate at the beginning of the academic year; program fidelity was measured monthly throughout the year. The main and interaction effects of perceived implementation climate and fidelity on student outcomes were examined using longitudinal nested linear models with random effects for classroom and student, controlling for important covariates. ***Results***: On average, IQ scores improved 2.2 points (SD = 8.7). There were no main effects of perceived implementation climate or fidelity on student outcomes; however, the interaction between perceived implementation climate and fidelity was associated with student outcomes ($p < 0.05$; d = 0.54). Among classrooms with a strong perceived implementation climate, higher fidelity was associated with better student outcomes. ***Conclusions***: While preliminary and requiring replication, these findings suggest that perceived implementation climate and program fidelity each may be important but not sufficient for optimizing outcomes for students with autism.

Keywords: Autism; classroom climate; dissemination; implementation; special education

Public health significance: While further study is needed, this study suggests that implementation climate, in combination with intervention fidelity, may be an important component of improving student outcomes. Future directions include studying the benefit of supplementing teacher training and consultation with organizational strategies to increase perceptions that the use of the intervention is expected, supported, and rewarded.

The Effect of Implementation Climate on Program Fidelity and Student Outcomes in Autism Support Classrooms

Children with autism receive the bulk of their intervention in schools (National Research Council, 2001). Although a growing number of autism interventions have demonstrated efficacy in improving cognition, adaptive behavior, and social functioning (National Autism Center, 2009; Wong et al., 2015), they rarely are implemented in schools in the way they were designed (Stahmer, 2007; Stahmer, Collings, & Palinkas, 2005), even when teachers receive considerable support (Mandell et al., 2013; Young, Falco, & Hanita, 2016). Perhaps as a result, child outcomes in schools are much worse than what is observed in university-based trials (Dingfelder & Mandell, 2010; Mandell et al., 2013; Silverman & Kurtines, 2004; Weisz, Donenberg, Han, & Weiss, 1995).

To date, relatively little systematic research has examined the implementation of autism programs in schools and associated student outcomes (Mandell et al., 2013; Pellecchia et al., 2015; Stahmer et al., 2015; Suhrheinrich et al., 2013). Emerging research in this area, however, suggests that the relationship between program implementation and outcomes may be complex. Although at least two studies have shown that higher fidelity to some autism intervention components is associated with better student outcomes (Pellecchia et al., 2015; Strain & Bovey, 2011), another found no direct effect of program implementation on student outcomes (Mandell et al., 2013).

One possible explanation is that the organizational context may set the conditions necessary for implementation efforts to be effective in improving student or client outcomes. For example, Kam, Greenberg, and Walls (2003) found significant intervention effects only in schools where both principal support and program implementation was high; neither high implementation quality nor high principal support alone predicted intervention effectiveness. Similar findings have been noted in the child mental health field more generally. In a noteworthy demonstration of this concept, Glisson and colleagues (2010) found a significant interaction between training clinicians in the use of multi-systemic therapy and implementing an organizational intervention. The most significant clinical improvement was found among youth treated in agencies in which therapists were trained and the agency received the organizational intervention.

An organizational concept of great interest in implementation science is “climate for innovation implementation,” sometimes referred to as implementation climate. Implementation climate refers to employees’ perceptions of the extent to which use of an innovation, technology or intervention is feasible and expected, supported, and rewarded by colleagues and supervisors. In the first test of this construct, which was conducted in manufacturing plants, Klein, Conn, and Sorra (2001) found that implementation climate was strongly associated with effective use of new technologies. Further studies in business settings also found a strong association between organizational implementation climate and both innovation implementation and outcomes (Dong, Neufeld, & Higgins, 2008; Lesselroth et al., 2011; Osei-Bryson, Dong, & Ngwenyama, 2008; Pullig, Maxham, & Hair, 2002). More recent studies support the applicability of this construct in health and human services settings (Helfrich, Weiner, McKinney, & Minasian, 2007; Holahan, Aronson, Jurkat, & Schoorman, 2004; Jacobs et al., 2015; Knudsen & Studts, 2010).

Implementation climate has not yet been used to predict program implementation or student outcomes in schools. Validated measures of school climate tend to assess the totality of the organizational environment (Hoy & Fedman, 1987; Hoy, Tarter, & Kottkamp, 1991), rather than organizational characteristics specific to implementing new programs. While more general organizational context has been shown to affect implementation and outcomes in both mental health settings (Glisson & Hemmelgarn, 1998; Glisson, Hemmelgarn, Green, & Williams, 2013; Glisson & Schoenwald, 2005; Glisson et al., 2010; Schoenwald & Hoagwood, 2001) and schools (Beets et al., 2008; Gittelsohn et al., 2003; Gregory, Henry, & Schoeny, 2007), implementation climate may be a uniquely important organizational characteristic because it is more proximal to implementation (Aarons, Horowitz, Dlugosz, & Ehrhart, 2012; Ehrhart, Aarons, & Farahnak, 2014; Schneider, Ehrhart, & Macey, 2013).

In the present study, we estimate associations among implementation climate, fidelity of implementation, and student outcomes using data from a large randomized field trial that examined the effectiveness of the Strategies for Teaching based on Autism Research (STAR) program, a teacher-delivered, classroom-based intervention for children with autism (Mandell et al., 2013). The present prospective cohort study was conducted after the randomized trial phase had ended and all classrooms were receiving STAR. We examine how teacher perceptions of implementation climate at the beginning of the school year predicted their implementation of evidence-based practices over the course of the year. We also examine how teachers' perceptions of implementation climate and their implementation fidelity were associated with student outcome. We hypothesized that: 1) perceived implementation climate, measured at baseline, will predict teachers' implementation over the course of the academic year; 2) higher implementation climate and program fidelity each will be associated with better student outcomes; and 3) perceived implementation climate will moderate the impact of program fidelity on outcomes, such that better student outcomes will be achieved when both perceived implementation climate and program fidelity are high.

Method

This study was approved by the University of Pennsylvania's Institutional Review Board.

Participants and Procedure

Data were drawn from Year 2 of a three-year randomized field trial (Mandell et al., 2013) of the STAR program (Arick et al., 2003). During that year, teachers in the treatment and control groups received the same training in STAR. STAR requires that teachers use three instructional strategies based on applied behavior analysis – discrete trial training (DTT), pivotal response training (PRT), and functional routines – in combination with a curriculum, student assessments, ongoing data collection of student performance, and specific classroom set-up. DTT is implemented using an intensive one-to-one teaching session in a highly structured setting free from distractions and involves breaking down complex skills into small component parts, and teaching each component part individually. 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. Functional routines are predictable activities with an expected sequence of steps that occur naturally throughout the day, such as transitioning between activities and having a snack (Arick et al., 2003).

Eligible staff comprised those working in one of the District's 52 kindergarten-through-second-grade autism support classrooms. Teachers were recruited through the District's Office of Specialized Services. All teachers and support staff members were required as part of their professional development to participate in STAR training. Those who opted to participate in the study also were provided with in-classroom consultation (direct observation and performance feedback). Consultation was delivered by trained research staff with expertise in STAR and occurred approximately once a month.

Of these 52 classrooms, teachers in 49 agreed to participate in the present study and provided informed consent. One teacher declined to participate; two were unresponsive to multiple contacts by the study team. Students were recruited through a consent form and flyer describing the program that teachers sent home with the student. Inclusion criteria for students were that they have a classification of autism through the District and be enrolled at least half-time in a participating classroom.

Table 1 describes classroom and teacher characteristics. In four classrooms, data collection at baseline on participating students was incomplete, leaving 45 classrooms for our analyses. The four excluded classrooms were similar to those included in all respects except that the 4 excluded teachers were more likely to have had prior experience with STAR (75% vs. 18%). In the 45 remaining classrooms, assessments were collected on 158 consented students at both time points. The number of consented children in these classrooms ranged from 1 to 8. Children ranged from 4 to 8 years of age.

Measures

Perceived implementation climate was defined as teachers' perceptions of the extent to which the use of STAR is feasible and expected, supported, and rewarded by colleagues and supervisors. We adapted the measure of implementation climate from Klein, Conn, and Sorra (2001). As a first step, we, along with classroom consultants from the study, reviewed the scales and items and identified those that were relevant to STAR and schools. Some scales were not selected because they were not relevant, and some items from the original scales we did select were dropped because they did not apply to STAR. We obtained a final set of eight scales deemed appropriate for our population and intervention. We then conducted exploratory analyses on our sample of 122 classroom staff working in 49 classrooms. Because we were not powered to conduct a factor analysis of all items together, we conducted these factor analyses for each subscale to obtain our final set of items. The subscales included: 1) 5-item scale measuring program ease of use (e.g., “In general, STAR is easy to use”); 2) 6-item scale measuring stress associated with program use (e.g., “As a result of STAR, my work is more time-consuming”); 3) 4-item scale measuring training quality and accessibility (e.g., “STAR training taught me what I need to know about STAR”); 4) 6-item scale measuring ongoing user support (e.g., “If staff members have a problem when using STAR, they can easily find someone to help them”); 5) 5-item scale measuring upward communication (e.g., “Staff feel confident that their suggestions for improving STAR are seriously considered by those in charge of STAR”); 6) 3-item scale measuring communication to employees about STAR (e.g., “Staff understand the reasons why this classroom is implementing STAR”); 7) 2-item scale measuring rewards for program use (e.g., “Supervisors praise staff for using STAR properly”); and 8) a 6-item scale on global perceptions of implementation climate (e.g., “STAR is a top priority in this classroom”). All items were rated on five-point Likert scales ranging from not true (1) to true (5), with some items reverse-coded. To weight each of these domains equally, these scale scores were averaged to create an overall perceived implementation climate score, and its preliminary reliability and validity was examined. Cronbach’s alpha for the perceived implementation climate scales as rated by teachers was .76. Teachers completed these scales at the beginning of the academic year.

Program fidelity was coded through video and checklist. Each classroom was filmed for 30 minutes every month implementing STAR intervention strategies. Videoing was designed to allow for 10 minute of video of discrete trial training, pivotal response training, and functional routines; a brief overview of the classroom setup also was filmed. In addition, research assistants completed a monthly program checklist created by the STAR program developers (Arick et al., 2003), which captured other program elements, such as classroom set-up. Trained research assistants and undergraduate students coded on a five-point Likert scale (ranging from 1-low fidelity to 5-high fidelity). Coding used different criteria for each teaching technique. For example, for discrete trial training (DTT), 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 use error correction procedures.

The lead coder, a PhD-level board certified behavior analyst with considerable research and clinical experience in schools, recoded two tapes for each coder every other month to measure criterion validity. Less than 90% agreement resulted in additional training until 90% agreement was achieved. Inter-rater reliability, as measured by agreement within one Likert point, was calculated for each instructional strategy and each month of videos by having a second coder, blind to the initial codes, score one third of the videos per strategy for each month. The average overall percent agreement for each strategy was: 86% for DTT (range of 60-100%); 90% for PRT (range of 75-100%); and 90% for functional routines (range of 67-100%). A primary coder was assigned to each strategy and those codes were used in the analyses.

Overall program fidelity was measured by averaging these four dimensions (DTT, PRT, functional routines, checklist) over the course of the academic year. All monthly fidelity data then were averaged to create a cumulative measure of STAR program fidelity throughout the academic year. There was good internal consistency for all four program components: 0.76 for the DTT coding scale, 0.95 for the PRT coding scale, 0.92 for the functional routines coding scale, and 0.80 for the checklist coding scale.

Child outcome was measured with the Differential Ability Scales, Second Edition (DAS-II (Elliott, 1990), which is used to assess cognitive abilities in children ages 2 years 6 months through 17 years 11 months across a broad range of developmental areas. Subscales include verbal, nonverbal, and spatial reasoning, which are combined to form a General Conceptual Ability (GCA) scale. Psychologists trained to research reliability administered the DAS-II at baseline (September) and at the end of the school year (May). Outcome was measured as each child's change in DAS-II General Conceptual Ability score from baseline to the end of the academic year. That score was selected as an outcome measure because STAR is a comprehensive program designed to target overall cognitive functioning. Reporting child outcomes as change in overall cognitive ability is common in autism intervention studies, especially outcome studies of comprehensive behavioral interventions (e.g., Eikeseth, Smith, Jahr, & Eldevik, 2007; Lovaas, 1987; Smith, 1999).

Other important classroom, teacher and student characteristics were included based on their possible association with climate, fidelity, or outcomes (Baker-Ericzén, Stahmer, & Burns, 2007; Beidas, Edmunds, Marcus, & Kendall, 2012; Fixsen, Blase, Naoom, & Wallace, 2009; Pellecchia et al., 2016; Perry et al., 2011; Rohrbach, Graham, & Hansen, 1993; Szatmari, Bryson, Boyle, Streiner, & Duku, 2003). They included:

C*lassroom characteristics:* baseline GCA score from the DAS-II administered in September and averaged across children in the classroom; mean age of children in the classroom; and number of students and support staff members in the classroom.

Teacher characteristics: years teaching children with autism; years working with the STAR program; and hours attending training and receiving consultation in the classroom (measured by attendance sheets and reports from the consultants).

Student characteristics: age and DAS-II GCA score at baseline. These student-level covariates were only included in the second set of analyses that estimated the association among climate, fidelity, and student outcomes.

Data Analysis

Data management and analyses were conducted using SPSS software version 24.0 and SAS. First, we examined descriptive data on the classrooms participating in the study. To estimate the association between climate and fidelity, we examined the association between the combined implementation climate score and the mean score for all fidelity components across the academic year. In order to present the most parsimonious regression model while still accounting for potentially confounding variables, we first estimated bivariate associations between each independent variable, including the perceived implementation climate score, and fidelity using linear regression. Next, variables with a bivariate association with fidelity significant at $p < .20$ were entered together into a multiple linear regression predicting overall fidelity. This entry criterion is based on Hosmer & Lemeshow's (2001) finding that covariates associated with the dependent variable at a statistical significance of $p<.20$ can potentially confound the association between the independent and dependent variables.

We then examined whether climate moderated the association between fidelity and outcomes. Predictors of outcome were estimated with two-level longitudinal linear models with random effects for classroom and student, accounting for time by using DAS-II change score for student outcomes (Donner & Klar, 2000; Murray, Hannan, Wolfinger, Baker, & Dwyer, 1998; Sashegyi, Brown, & Farrell, 2000). Similar to the approach detailed above, the unadjusted association between each variable and outcome first was estimated. Second, variables significant at $p < .20$ were included in an adjusted model as covariates. Third, we included the interaction of perceived implementation climate and program fidelity in the model predicting outcome. To examine this interaction further, we created dichotomous variables for implementation climate and program fidelity and examined descriptive data to illustrate the students and classrooms that comprise these four classroom types. Finally, to examine the stability of these findings, we conducted sensitivity analyses: 1) limiting our dataset to classrooms where we had 3 or more students (n = 33 classrooms and 137 students); 2) using robust standard errors, with classroom as a cluster variable; 3) using bootstrapping with 1,000 replications; and 4) using the classroom average for all student-level independent and dependent variables, rather than nested data.

Results

Description of Sample

Table 1 presents descriptive data on teachers and classrooms participating in the study. Teachers averaged 3.6 years of experience working with children with autism; 18 percent had prior experience with the STAR program. The number of total students in each classroom ranged from 4 to 12 and the number of support staff ranged from 1 to 7. There was considerable variation in perceived implementation climate and program fidelity.

Twenty percent of the students had experience with the STAR program from the prior academic year. Students were 88% male and 45% African American, 18% white, 9% Hispanic/Latino, 6% other race/ethnicity, and 21% unknown race/ethnicity. These data were missing because those parents did not submit the requested demographic information form. Students averaged 87.0 months old (SD = 10.8; range 61-107) and had an average baseline DAS-II GCA score of 62.5 (SD = 21.8; range: 20 – 108). On average, DAS scores improved 2.2 points

(SD = 8.7), with considerable variability, ranging from a decrease of 24 points to an increase of 35 points.

Climate and Fidelity

Table 2 presents the results of analyses examining the association between the combined perceived implementation climate score and the mean score for all fidelity components across the academic year. The first set of columns comprises unadjusted bivariate analyses of each covariate predicting fidelity. The second includes only those covariates that were significant at $p < .20$ in the unadjusted analyses. Only prior experience with the STAR program was statistically significantly associated with fidelity in these unadjusted bivariate analyses. In addition, two covariates were associated with fidelity at $p < .20$ and therefore met entry criteria for the multiple regression: perceived implementation climate score and number of consultation hours received. In the adjusted model (second set of columns), these three variables explained 16% of the variance in overall STAR fidelity; however, no variable remained statistically significantly associated with fidelity in this adjusted model at $p < .05$.

Climate, Fidelity, and Student Outcomes

Table 3 presents the results of two regression models predicting change in DAS-II score. The first set of columns comprises unadjusted bivariate analyses of each covariate predicting change in DAS-II score. The second presents the results of the multiple regression including only those covariates that were significant at $p < .20$ in the unadjusted bivariate analyses. In bivariate analysis, each additional child in the classroom was associated with a 0.91-point average increase in change in DAS score. Each additional month of child age at baseline was associated with a .12 decrease in change in DAS score ($p < .10$). These findings persisted after adjustment at $p < .05$.

There was a statistically significant interaction between perceived implementation climate and program fidelity (Figure 1; B = 6.09; $p < .01$ in unadjusted analysis; B = 4.75; $p < .05$ after adjusting for child age and number of children in classroom). The effect size for the interaction term, measured as Cohen's D, was 0.70 for the unadjusted model and 0.54 for the adjusted model, suggesting a medium effect after adjusting for student age and the number of students in the classroom (Cohen, 1988). This effect size was derived by dividing the unstandardized regression coefficient for the interaction term by the standard deviation for the outcome (change in DAS score; SD = 8.74; Feingold, 2009). Among classrooms with low perceived implementation climate, students whose classrooms implemented STAR with low fidelity experienced a greater gain in DAS-II score than those whose classrooms implemented STAR with high fidelity. Among classrooms with high perceived implementation climate, students whose classrooms implemented STAR with high fidelity experienced a greater gain than those whose classrooms implemented STAR with low fidelity. Table 4 illustrates this finding using descriptive data for the students and classrooms that comprise the four classroom types (low/high implementation climate, low/high program fidelity). The four groups were roughly equal in size when using a median split for both climate and fidelity. In our sensitivity analyses, the magnitude, direction, and (with the exception of the bootstrapping and classroom average approaches in the adjusted models) statistical significance of our findings was consistent across all five approaches, suggesting stability of these results. The two sensitivity analyses that did not yield statistically significant findings at $p < .05$ resulted in p values $< .10$.

Discussion

This study found that, while the average IQ gain across the sample was small (2.2 points) and neither perceived implementation climate nor program fidelity independently was associated with student outcomes, their interaction was, such that higher program fidelity was associated with greater student gains when implementation climate was high.

Only one variable showed a consistent association with program fidelity – prior experience with the program. Teachers with more STAR experience demonstrated the greatest fidelity to STAR. Given STAR's complexity, it is not surprising that increased practice with the program would increase the likelihood of stronger implementation.

Although the finding that climate did not predict fidelity after adjusting for covariates was contrary to our hypotheses and prior research, it was consistent with a recent study that found no association between implementation climate and implementation of evidence-based practices (Beidas et al., 2015). These researchers suggest that more complex interactive or mediational processes may account for how implementation climate works together with other measures of culture or climate to predict implementation.

The interaction between climate and fidelity in predicting outcomes suggests that perceived implementation climate and program fidelity both may be important but that neither is sufficient to improve outcomes. The finding that students in the high-climate, high-fidelity classrooms had positive outcomes is consistent with the few previous studies in this area, including the studies noted in the introduction by Kam, Greenberg, and Walls (2003) and Glisson et al. (2010).

One explanation for the lack of association between fidelity and outcome in low-climate classrooms is that the high-fidelity, low-climate group consisted of teachers that persisted in implementing the intervention without necessary supports and despite substantial barriers, leading to high implementation at the cost of other important classroom factors (e.g., general positive teaching behaviors, teacher burnout, staff morale). Another interpretation, based on observations in psychotherapy studies, is that therapists deliver interventions with greater fidelity to those patients who are not improving (Webb, DeRubeis, & Barber, 2010), resulting in the observation that fidelity is not associated with outcomes.

More surprising are those teachers whose students achieved good outcomes despite low fidelity and low perceived implementation climate. In these classrooms, it may be that teachers correctly decided that STAR was not a good fit for their classrooms, either because they did not view it as effective for their students or did not perceive that its use was feasible or supported. This hypothesis is consistent with an accumulating body of literature emphasizing the importance of innovation-environment fit (Aarons, 2005; Lyon et al., 2014). Anecdotal evidence suggests that this group included some teachers who agreed to participate in the study but were not interested in changing practice. Their teaching did not correspond to the requirements for STAR, but may have been equally effective in that particular context. For example, Mandell et al. (2013) note that these teachers may systematically or intuitively abstract program components and apply them flexibly but appropriately to students based on individual need. This aligns with findings that more flexible, modular treatments have demonstrated greater effectiveness in community settings compared to less flexible manualized treatments (Chorpita et al., 2013; Santucci, Thomassin, Petrovic, & Weisz, 2015; Weisz et al., 2012), and has the potential to inform research about selectively applying intervention components based on context (Green, 2008).

Other than the climate-by-fidelity interaction, only student age (with younger children demonstrating the greatest gains) and number of students in the classroom (with children in larger classrooms demonstrating the greatest gains) were associated with outcomes. The finding regarding student age is consistent with prior research (Fenske, Zalenski, Krantz, & McClannahan, 1985; Harris & Handleman, 2000) and with the widespread agreement in the autism community that earlier intervention results in better outcome (National Research Council, 2001). The finding that larger class size was associated with greater gains is unexpected. One explanation is that students in these larger classrooms had more opportunities for socialization, resulting in larger cognitive gains through peer modeling (Nahmias, Kase, & Mandell, 2014; National Research Council, 2001). Another possibility is that larger classroom size is a proxy for positive school or classroom functioning. The district and principals may make informed decisions to move more children into them based on positive progress; children who have greater learning needs may be more likely placed in smaller classrooms.

Several study limitations should be considered. First, while our perceived implementation climate scale demonstrated good internal consistency, it is an investigator-created adaption of an existing measure and has gone through only limited psychometric testing. Second, the original implementation climate scale, as with other climate scales (e.g., Glisson & Schoenwald, 2005; Schoenwald & Hoagwood, 2001), was developed in settings in which there were a large number of group members within each unit. Because our classrooms were composed of individuals who were less homogenous in their role (teacher and classroom assistant), data could not be treated in the same fashion as recommended by many climate researchers (Bliese, 2000; Klein et al., 2000; Klein et al., 2001), in which individual-level data are examined for within-unit agreement, with the goal of being aggregated to the higher unit level. Instead, we relied on only the teachers' perceptions of implementation climate, making our measure analogous to measures of psychological climate, rather than true implementation climate. This approach is consistent with how some other researchers have examined implementation climate (e.g., Dong et al., 2008; Jacobs et al., 2015; Osei-Bryson et al., 2008).

Second, the unique setting of the special education classroom made it infeasible to use better-tested scales such as the Implementation Climate Scale (ICS; Ehrhart et al., 2014): while many of the items overlap since both were adapted from Klein’s measure, some of the other items on the ICS are not appropriate for this setting (for example, some items on rewards for use were less applicable as the public education system does not allow for flexibility in this area).

Third, while we used well-validated measures of program fidelity developed specifically for the STAR program, we only rated videos of classrooms from a thirty-minute sample each month. While these data demonstrated whether classroom staff *could* implement the program when prompted, it may not fully capture the consistency and intensity of implementation. Given the complexity of STAR, however, it would be difficult to achieve high fidelity without practice, and it would be difficult for teachers to alter their classrooms and schedule for the fidelity visits alone. Our measurement strategy is still likely more accurate than self-report, which tends to overestimate program fidelity (Emshoff et al., 1987; Fixsen, Naoom, Blase, Friedman, & Wallace, 2005; O'Donnell, 2008).

Finally, we relied on the DAS-II as an outcome measure. While measures of cognitive ability are the most commonly used outcome measure for comprehensive ABA-based programs such as STAR, they may miss important targets of intervention, such as academic achievement or autism-specific symptoms.

Despite these limitations, our study findings may have important implications if replicated. Among children in classrooms with high implementation climate, IQ increased more than 11 points from the lowest to highest fidelity classrooms – almost one standard deviation on the DAS-II. This approaches the gains found in university-based trials of early intensive behavioral intervention (Cohen, Amerine-Dickens, & Smith, 2006; Eikeseth et al., 2007; Howard, Sparkman, Cohen, Green, & Stanislaw, 2005; Sallows & Graupner, 2005) and is a greater magnitude than gains found in a study of lower-intensity (12 hours per week) early behavioral intervention (Eldevik, Eikeseth, Jahr, & Smith, 2006). Increasing program fidelity and implementation climate concurrently may hold promise for creating gains in children served in the community that approach gains observed in university-based trials.

Despite our focus on autism intervention in schools, our findings have the potential to generalize beyond the school setting. While STAR is an autism-specific teacher-delivered intervention, its strategies are based on principles of operant conditioning, consistent with the majority of evidence-based intervention strategies for youth (e.g., Weisz, Donenberg, Han, & Kauneckis, 1995). Similarly, while we focused on the classroom implementation climate specifically, Klein et al.'s (2001) measures required only minimal adaption and have been studied in other health and human service contexts.

Researchers have suggested that multi-component implementation strategies are needed to improve program implementation and outcomes (Aarons, Hurlburt, & Horwitz, 2011; Ferlie & Shortell, 2001; Fixsen et al., 2009; Glisson & Schoenwald, 2005; Glisson et al., 2010). Despite this general consensus, an emergent theme across proposed implementation models is the lack of evidence about which variables are likely to play key roles in any given implementation effort or context (Aarons et al., 2011). Most research has focused on directly improving implementation (e.g., through teacher training and consultation) to maximize client or student outcomes. Our findings provide preliminary evidence that implementation climate also may be important, and point to the potential benefit of additional strategies to maximize the impact of implementation on outcomes. While preliminary and requiring replication, our results suggest that it may be helpful for training and consultation to be accompanied by an organizational assessment to evaluate whether the use of the intervention is expected, supported, and rewarded, as well as interventions to improve implementation climate where it is needed.

References

Aarons, G. A. (2005). Measuring provider attitudes toward evidence-based practice: Consideration of organizational context and individual differences. *Child and Adolescent Psychiatric Clinics of North America, 14*(2), 255-271.

Aarons, G. A., Horowitz, J., Dlugosz, L., & Ehrhart, M. (2012). The role of organizational processes in dissemination and implementation research. *Dissemination and Implementation Research in Health: Translating Science to Practice, 128*, 153.

Aarons, G. A., Hurlburt, M., & Horwitz, S. M. (2011). Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Administration and Policy in Mental Health, 38*, 4-23.

Arick, J., Young, H., Falco, R., Loos, L., Krug, D., Gense, M., 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.

Baker-Ericzén, M. J., Stahmer, A. C., & Burns, A. (2007). Child demographics associated with outcomes in a community-based pivotal response training program. *Journal of Positive Behavior Interventions, 9*(1), 52-60.

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., Edmunds, J. M., Marcus, S. C., & Kendall, P. C. (2012). Training and consultation to promote implementation of an empirically supported treatment: A randomized trial. *Psychiatric Services, 63*(7), 660-665.

Beidas, R. S., Marcus, S., Aarons, G. A., Hoagwood, K. E., Schoenwald, S., Evans, A. C., et al. (2015). Predictors of community therapists' use of therapy techniques in a large public mental health system. *JAMA Pediatrics, 169*(4), 374-382.

Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 349-381). San Francisco, CA: Jossey-Bass.

Chorpita, B. F., Weisz, J. R., Daleiden, E. L., Schoenwald, S. K., Palinkas, L. A., Miranda, J., et al. (2013). Long-term outcomes for the Child STEPs randomized effectiveness trial: A comparison of modular and standard treatment designs with usual care. *Journal of Consulting and Clinical Psychology, 81*(6), 999-1009.

Cohen, H., Amerine-Dickens, M., & Smith, T. (2006). Early intensive behavioral treatment: Replication of the UCLA model in a community setting. *Journal of Developmental and Behavioral Pediatrics, 27*(2), S145-S155, Suppl. S.

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

Dingfelder, H., & Mandell, D. (2010). 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.

Dong, L., Neufeld, D. J., & Higgins, C. (2008). Testing Klein and Sorra's innovation implementation model: An empirical examination. *Journal of Engineering and Technology Management, 25*, 237-255.

Donner, A., & Klar, N. (2000). *Cluster Randomization Trials in Health Research*. London: Arnold.

Ehrhart, M. G., Aarons, G. A., & Farahnak, L. R. (2014). Assessing the organizational context for EBP implementation: the development and validity testing of the Implementation Climate Scale (ICS). *Implementation Science, 9*(1), 157.

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 comparison controlled study. *Behavior Modification*, *31*(3), 264-278.

Eldevik, S., Eikeseth, S., Jahr, E., & Smith, T. (2006). Effects of low-intensity behavioral treatment for children with autism and mental retardation. *Journal of Autism and Developmental Disorders, 36*, 211-224.

Elliott, C. (1990). *Differential Ability Scales: Administration and Scoring Manual*. San Antonio, TX: Psychological Corp.

Emshoff, J. G., Blakely, C., Gottschalk, R., Mayer, J., Davidson, W. S., & Erickson, S. (1987). Innovation in education and criminal justice: measuring fidelity of implementation and program effectiveness. *Educational Evaluation and Policy Analysis*, *9*(4), 300-311.

Feingold, A. (2009). Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. *Psychological Methods*, *14*(1), 43-53.

Fenske, E. C., Zalenski, S., Krantz, P. J., & McClannahan, L. E. (1985). Age at intervention and treatment outcome for autistic children in a comprehensive intervention program. *Analysis and Intervention in Developmental Disabilities, 5*, 49-58.

Ferlie, E. B., & Shortell, S. M. (2001). Improving the quality of heath care in the United Kingdom and the United States: a framework for change. *Milbank Quarterly, 79*, 281-315.

Fixsen, D. L., Blase, K. A., Naoom, S. F., & Wallace, F. (2009). Core implementation components. *Research on Social Work Practice, 19*, 531-540.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: a synthesis of the literature*. Tampa: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network.

Gittelsohn, J., Merkle, S., Story, M., Stone, E. J., Steckler, A., Noel, J., et al. (2003). School climate and implementation of the Pathways study. *Preventive Medicine, 37*, S97-S106.

Glisson, C., & Hemmelgarn, A. (1998). The effects of organizational climate and interorganizational coordination on the quality and outcomes of children's service systems. *Child Abuse and Neglect, 22*(5), 401-421.

Glisson, C., Hemmelgarn, A., Green, P., & Williams, N. J. (2013). Randomized trial of the availability, responsiveness and continuity (ARC) organizational intervention for improving youth outcomes in community mental health programs. *Journal of the American Academy of Child and Adolescent Psychiatry, 52*(5), 493-500.

Glisson, C., & Schoenwald, S. K. (2005). The ARC organizational and community intervention strategy for implementing evidence-based children's mental health treatments. *Mental Health Services Research, 7*(4), 243-259.

Glisson, C., Schoenwald, S. K., Hemmelgarn, A., Green, P., Dukes, D., Armstrong, K. S., 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*, 537-550.

Green, L. W. (2008). Making research relevant: if it is an evidence-based practice, where's the practice-based evidence? *Family Practice, 25*(suppl_1), i20-i24.

Gregory, A., Henry, D. B., & Schoeny, M. E. (2007). School climate and implementation of a preventive intervention. *American Journal of Community Psychology, 40*, 250-260.

Harris, S. L., & Handleman, J. S. (2000). Age and IQ at intake as predictors of placement for young children with autism: a four- to six-year follow-up. *Journal of Autism and Developmental Disorders, 30*(2), 137-142.

Helfrich, C. D., Weiner, B. J., McKinney, M. M., & Minasian, L. (2007). Determinants of implementation effectiveness: adapting a framework for complex innovations. *Medical Care Research and Review, 64*, 279-303.

Holahan, P., Aronson, Z. H., Jurkat, M. P., & Schoorman, F. D. (2004). Implementing computer technology: a multiorganizational test of Klein and Sorra's model. *Journal of Engineering and Technology Management, 21*, 31-50.

Hosmer, D., & Lemeshow, S. (2001). *Applied logistic regression, textbook and solutions manual*. Hoboken, NJ: Wiley-Interscience.

Howard, J., Sparkman, C., Cohen, H., Green, G., & Stanislaw, H. (2005). A comparison of intensive behavior analytic and eclectic treatments for young children with autism. *Research in Developmental Disabilities, 26*(4), 359-383.

Hoy, W., & Fedman, J. (1987). Organizational health: The concept and its measure. *Journal of Research and Development in Education, 20*(4), 30-37.

Hoy, W., Tarter, C., & Kottkamp, R. (1991). *Open schools/healthy schools: Measuring organizational climate*. Beverly Hills, CA: Sage.

Jacobs, S. R., Weiner, B. J., Reeve, B. B., Hofmann, D. A., Christian, M., & Weinberger, M. (2015). Determining the predictors of innovation implementation in healthcare: A quantitative analysis of implementation effectiveness. *BMC Health Services Research, 15*(1), 6.

Kam, C.-M., Greenberg, M. T., & Walls, C. T. (2003). Examining the role of implementation quality in school-based prevention using the PATHS curriculum. *Prevention Science, 4*(1), 55-63.

Klein, K. J., Bliese, P. D., Kozolowski, S. W., Dansereau, F., Gavin, M. B., Griffin, M. A., et al. (2000). Multilevel analytical techniques: Commonalities, differences, and continuing questions. In K. J. Klein & S. W. Kozolowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 512-553). San Francisco, CA: Jossey-Bass.

Klein, K. J., Conn, A. B., & Sorra, J. S. (2001). Implementing computerized technology: an organizational analysis. *Journal of Applied Psychology, 86*(5), 811-824.

Knudsen, H. K., & Studts, J. L. (2010). The implementation of tobacco-related brief interventions in substance abuse treatment: a national study of counselors. *Journal of Substance Abuse Treatment, 38*, 212-219.

Lesselroth, B. J., Holahan, P. J., Adams, K., Sullivan, Z. Z., Church, V. L., Woods, S., et al. (2011). Primary care provider perceptions and use of a novel medication reconciliation technology. *Informatics in Primary Care, 19*, 105-118.

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.

Lyon, A. R., Ludwig, K., Romano, E., Koltracht, J., Vander Stoep, A., & McCauley, E. (2014). Using modular psychotherapy in school mental health: Provider perspectives on intervention-setting fit. *Journal of Clinical Child and Adolescent Psychology, 43*(6), 890-901.

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.

Murray, D., Hannan, P., Wolfinger, R., Baker, W., & Dwyer, J. (1998). Analysis of data from group-randomized trials with repeat observations on the same groups. *Statistics in Medicine, 17*, 1581-1600.

Nahmias, A. S., Kase, C., & Mandell, D. S. (2014). Comparing cognitive outcomes among children with autism spectrum disorders receiving community-based early intervention in one of three placements. *Autism, 18*(3), 311-320.

National Autism Center. (2009). National Standards Report: The National Standards Project: Addressing the need for evidence-based practice guidelines for autism spectrum disorders: National Autism Center.

National Research Council (Ed.). (2001). *Educating children with autism*. Washington, DC: National Academy Press.

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research, 78*(1), 33-84.

Osei-Bryson, K. M., Dong, L., & Ngwenyama, O. (2008). Exploring managerial factors affecting ERP implementation: an investigation of the Klein-Sorra model using regression splines. *Information Systems Journal, 18*, 499-527.

Pellecchia, M., Connell, J. E., Beidas, R. S., Xie, M., Marcus, S. C., & Mandell, D. S. (2015). Dismantling the active ingredients of an intervention for children with autism. *Journal of Autism and Developmental Disorders, 45*(9), 2917-2927.

Pellecchia, M., Connell, J. E., Kerns, C. M., Xie, M., Marcus, S. C., & Mandell, D. S. (2016). Child characteristics associated with outcome for children with autism in a school-based behavioral intervention. *Autism, 20*(3), 321-329.

Perry, A., Cummings, A., Geier, J. D., Freeman, N. L., Hughes, S., Managhan, T., et al. (2011). Predictors of outcome for children receiving intensive behavioral intervention in a large, community-based program. *Research in Autism Spectrum Disorders, 5*(1), 592-603.

Pullig, C., Maxham, J. G., & Hair, J. F. (2002). Salesforce automation systems: An exploratory examination of organizational factors associated with effective implementation and sales force productivity. *Journal of Business Research, 55*, 401-415.

Rohrbach, L. A., Graham, J. W., & Hansen, W. B. (1993). Diffusion of a school-based substance abuse prevention program: predictors of program implementation. *Preventive Medicine, 22*, 237-260.

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.

Santucci, L. C., Thomassin, K., Petrovic, L., & Weisz, J. R. (2015). Building evidence-based interventions for the youth, providers, and contexts of real-world mental-health care. *Child Development Perspectives, 9*(2), 67-73.

Sashegyi, A., Brown, S., & 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.

Schneider, B., Ehrhart, M. G., & Macey, W. H. (2013). Organizational climate and culture. *Annual Review of Psychology, 64*, 361-388.

Schoenwald, S. K., & Hoagwood, K. (2001). Effectiveness, transportability, and dissemination of interventions: what matters when? *Psychiatric Services, 52*(9), 1190-1197.

Silverman, W. K., & Kurtines, W. M. (2004). Research progress on effectiveness, transportability, and dissemination of empirically supported treatments: Integrating theory and research. *Clinical Psychology: Science and Practice, 11*(3), 295-299.

Smith, T. (1999). Outcome of early intervention for children with autism. *Clinical Psychology: Science and Practice*, *6*(1), 33-49.

Stahmer, A. C. (2007). The basic structure of community early intervention programs for children with autism: provider descriptions. *Journal of Autism and Developmental Disorders, 37*(7), 1344-1354.

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*, 66-79.

Stahmer, A. C., Rieth, S., Lee, E., Reisinger, E. M., Mandell, D. S., & Connell, J. E. (2015). Training teachers to use evidence-based practices for autism: Examining procedural implementation fidelity. *Psychology in the Schools, 52*(2), 181-195.

Strain, P. S., & Bovey, E. H. (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*(3), 133-154.

Suhrheinrich, J., Stahmer, A. C., Reed, S., Schreibman, L., Reisinger, E., & Mandell, D. (2013). Implementation challenges in translating pivotal response training into community settings. *Journal of Autism and Developmental Disorders, 43*(12), 2970-2976.

Szatmari, P., Bryson, S., Boyle, M., Streiner, D., & Duku, E. (2003). Predictors of outcome among high functioning children with autism and Asperger syndrome. *Journal of Child Psychology and Psychiatry, 44*(4), 520-528.

Webb, C. A., DeRubeis, R. J., & Barber, J. P. (2010). Therapist adherence/competence and treatment outcome: a meta-analytic review. *Journal of Consulting & Clinical Psychology, 78*, 200-211.

Weisz, J. R., Chorpita, B. F., Palinkas, L. A., Schoenwald, S. K., Miranda, J., Bearman, S. K., et al. (2012). Testing standard and modular designs for psychotherapy treating depression, anxiety, and conduct problems in youth: A randomized effectiveness trial. *Archives of General Psychiatry, 69*(3), 274-282.

Weisz, J. R., Donenberg, G. R., Han, S. S., & Kauneckis, D. (1995). Child and adolescent psychotherapy outcomes in experiments versus clinics: Why the disparity? *Journal of Abnormal Child Psychology, 23*(1), 83-106.

Weisz, J. R., Donenberg, G. R., Han, S. S., & Weiss, B. (1995). Bridging the gap between laboratory and clinic in child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology, 63*(5), 688.

Wong, C., Odom, S. L., Hume, K. A., Cox, A. W., Fettig, A., Kucharczyk, S., et al. (2015). Evidence-based practices for children, youth, and young adults with autism spectrum disorder: A comprehensive review. *Journal of Autism and Developmental Disorders, 45*(7), 1951-1966.

Young, H., Falco, R., & Hanita, M. (2016). Randomized, controlled trial of a comprehensive program for young students with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 46*(2), 544-560.

Appendix. *Items from Program Implementation Climate Scales*

	not true	slightly true	somewhat true	mostly true	true
PICS-E					
1. STAR is cumbersome to use.	1	2	3	4	5
2. People can easily access necessary materials to use STAR.	1	2	3	4	5
3. In general, STAR is easy to use.	1	2	3	4	5
4. It is easy to use STAR forms and data sheets.	1	2	3	4	5
5. STAR is "user-friendly."	1	2	3	4	5
PICS-S					
1. As a result of STAR, my work is more time-consuming.*	1	2	3	4	5
2. STAR has made it more uncomfortable for me to perform my work tasks.*	1	2	3	4	5
3. Because of STAR, I do not have enough time to get my work done.*	1	2	3	4	5
4. STAR has made my job a lot more frustrating than it was before STAR.*	1	2	3	4	5
5. People in this classroom feel that they have enough time to do their work *and* learn new skills associated with the shift to STAR.	1	2	3	4	5
6. Most people are so busy that they have little time to devote to the implementation of STAR.*	1	2	3	4	5
PICS-T					
1. I was given enough information during STAR training.	1	2	3	4	5
2. STAR training taught me what I need to know about STAR.	1	2	3	4	5
3. I learned a lot in STAR training.	1	2	3	4	5
4. The quality of STAR training I received was very good.	1	2	3	4	5
PICS-US					
1. Training is readily available to staff who want to learn more about STAR.	1	2	3	4	5
2. If staff members have a problem when using STAR, they can easily find someone to help them.	1	2	3	4	5
3. Helpful books and manuals are available when staff members have problems with STAR.	1	2	3	4	5
4. Staff members find it hard to get help when they run into problems using STAR.*	1	2	3	4	5
5. It takes a long time to get questions about STAR	1	2	3	4	5

answered.*					
6. Someone is available to help when employees get stuck on a problem when using STAR.	1	2	3	4	5
PICS-UC					
1. The people in charge of STAR ignore staff members' suggestions for improving STAR.*	1	2	3	4	5
2. If staff members have ideas about how STAR should be used, they are able to influence the people who make the final decisions.	1	2	3	4	5
3. Staff are encouraged to make suggestions about how to improve the use of STAR in the classroom.	1	2	3	4	5
4. When staff members have criticisms of STAR, they tell the people who are in charge of STAR.	1	2	3	4	5
5. Staff feel confident that their suggestions for improving STAR are seriously considered by those in charge of STAR.	1	2	3	4	5
PICS-R					
1. Supervisors praise staff for using STAR properly.	1	2	3	4	5
2. Staff get a "pat on the back" when they go out of their way to learn more about STAR.	1	2	3	4	5
PICS-CE					
1. Staff here are well informed about STAR.	1	2	3	4	5
2. Staff understand the reasons why this classroom is implementing STAR.	1	2	3	4	5
3. Staff know the specific goals that this classroom hopes to achieve by implementing STAR.	1	2	3	4	5
PICS-GC					
1. STAR is a top priority in this classroom.	1	2	3	4	5
2. In this classroom, STAR takes a back seat to other projects or activities.*	1	2	3	4	5
3. People put a lot of effort into making STAR a success here.	1	2	3	4	5
4. People in this classroom think that the implementation of STAR is important.	1	2	3	4	5
5. One of this classroom's main goals is to use STAR effectively.	1	2	3	4	5
6. In this classroom, there is a big push for people to make the most of STAR.	1	2	3	4	5

NOTES:
PICS = Program Implementation Climate Scales
PICS-E = ease of STAR use
PICS-S = stress associated with using STAR
PICS-T = training quality and accessibility
PICS-US = user support from STAR trainers/coaches
PICS-UC = upward communication about STAR

PICS-CE = communication to employees about STAR
PICS-R = rewards for STAR use
PICS-GC = global climate

SCORING:
* = reverse scored items

Figure 1. *Association Between Program Fidelity and Student Outcomes Among Low and High Implementation Climate Classrooms*

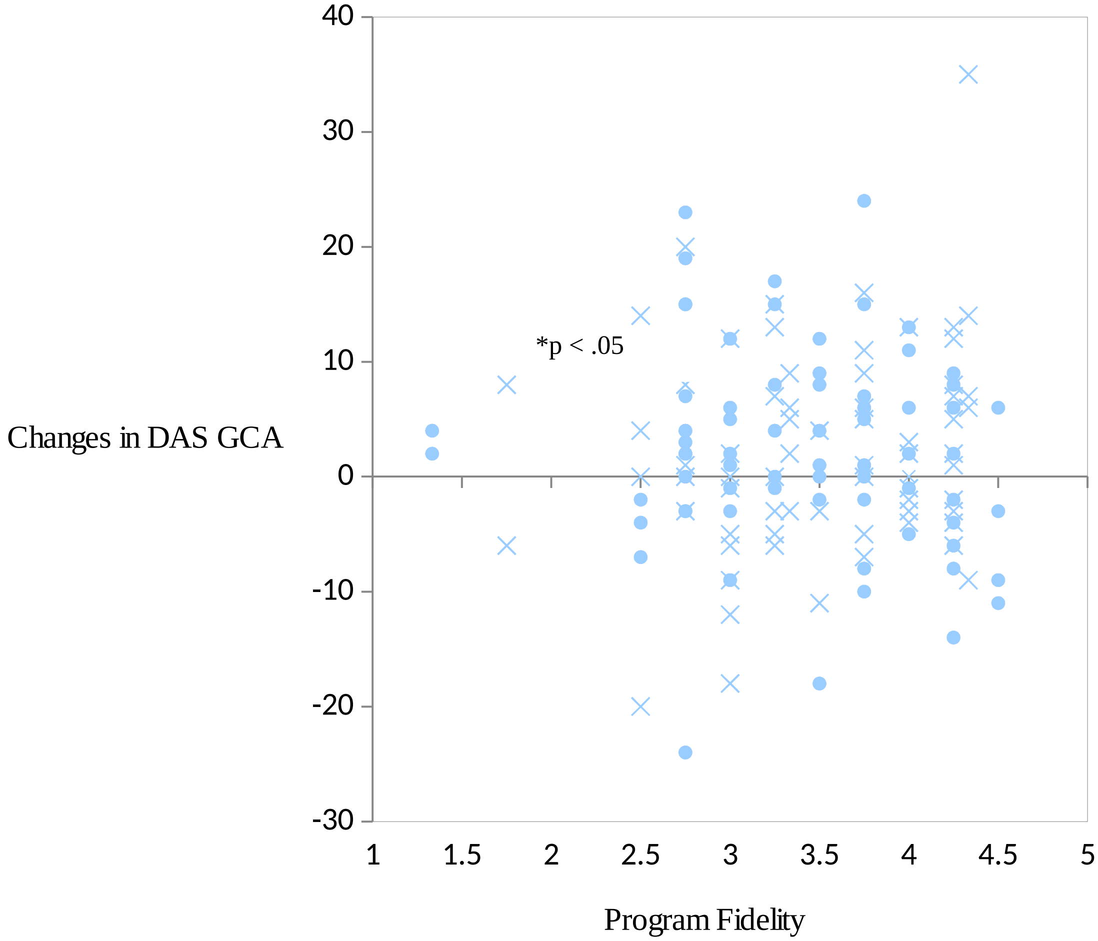

Table 1. *Teacher and Classroom Characteristics*

Variable	Mean	SD	Range
Classroom Characteristics (n = 45)			
Implementation climate	4.0	0.5	2.5-4.8
Overall program fidelity	3.3	0.7	1.3-4.6
Number of enrolled children in classroom	4.2	1.7	1-8
Number of support staff in classroom	3.8	1.3	1-7
Number of children in classroom	7.3	1.6	4-12
Average age of students (years)	6.2	0.7	4.7-8.0
Average student baseline cognitive functioning	58.5	14.3	27.0-86.3
Teacher Characteristics (n = 45)			
Number of workshop hours teacher attended	32.2	6.8	17-46.5
Number of consultation hours received by teacher	68.5	14.2	30-95.5
Teacher's years of autism experience	3.6	5.0	0-21
Teacher's years of experience teaching STAR	0.2	0.4	0-1

Table 2. *Multiple Regression of Baseline Variables Predicting Overall Fidelity by Year-End*

Variable	**Unadjusted Analyses**		**Adjusted Analysis**	
	B	**P**	***B***	**p**
Implementation climate	0.35	0.08	0.23	0.25
Number of consultation hours received	0.01	0.16	0.01	0.16
Years of experience teaching STAR	0.55	0.05	0.45	0.12
Years of autism teaching experience	-0.02	0.50	-	-
Number of workshop hours attended	0.01	0.62	-	-
Number of support staff in classroom	0.01	0.94	-	-
Number of children in classroom	0.04	0.51	-	-
Average age of students (years)	0.03	0.87	-	-
Average baseline cognitive functioning of students	0.00	0.94	-	-
R^2				0.16

Note. Unadjusted analyses consisted of each variable alone predicting fidelity. Adjusted analyses consisted of multiple linear regression with all listed variables included.
n = 45 for all analyses.

Table 3. *Multiple Regression of Teacher, Student, and Classroom Characteristics Predicting Student Change in DAS Score*

Variable	**Unadjusted Analyses**		**Adjusted Analysis**	
	Estimate	**P-Value**	**Estimate**	**P-Value**
Number of children in classroom	0.91	0.04	1.00	0.04
Student age (months)	-0.12	0.09	-0.14	0.03
Implementation climate	1.46	0.22	-	-
Overall program fidelity	0.40	0.71	-	-
Teacher's years of autism experience	0.05	0.77	-	-
Teacher's years of experience teaching STAR	-0.03	0.99	-	-
Number of workshop hours teacher attended	0.00	0.97	-	-
Number of consultation hours teacher received	-0.02	0.69	-	-
Number of support staff in classroom	0.47	0.38	-	-
Student baseline DAS-II General Conceptual Ability Score	-0.03	0.40	-	-
R^2			0.06	

Note. Unadjusted analyses consisted of each variable alone predicting student change in DAS GCA score. Adjusted analysis consisted of multiple linear regression with both listed variables included.
n = 158.

Table 4. *Student, teacher, and classroom characteristics by classroom classification*

			PROGRAM FIDELITY					
			LOW			**HIGH**		
PERCEIVED IMPLEMENTATION CLIMATE	**LOW**		N = 39			N = 38		
		Measure	Mean	SD	Range	Mean	SD	Range
		Baseline implementation climate	3.4	0.4	2.5-4.0	3.7	0.2	3.3-4.1
		Overall program fidelity	2.8	0.4	1.3-3.3	3.9	0.3	3.5-4.5
		Student DAS GCA change	3.8	9.3	-24.0-23.0	1.6	8.8	-18.0-24.0
		Student age (months)	83.0	11.7	61.0-107.0	92.2	10.5	74.0-107.0
		Student baseline DAS GCA	70.1	23.0	21.0-108.0	53.0	18.4	23.0-84.0
		Teacher's years of autism experience	3.2	4.8	0-20.0	1.9	1.1	0-4.0
		Teacher's years of experience with STAR	0.0	0.0	0-0	0.0	0.2	0-1.0
		Number of workshop hours teacher attended	33.3	8.2	19.5-46.5	32.6	7.4	17.0-38.5
		Number of consultation hours teacher received	71.1	12.6	52.0-93.0	66.9	20.0	30.0-92.0
		Number of support staff in classroom	4.3	1.1	2.0-6.0	4.5	1.2	1.0-6.0
		Number of children in classroom	7.5	1.7	4.0-11.0	7.5	1.7	5.0-10.0
	HIGH		N = 39			N = 42		
		Measure	Mean	SD	Range	Mean	SD	Range
		Baseline implementation climate	4.4	0.3	4.1-4.8	4.5	0.2	4.1-4.8
		Overall program fidelity	2.9	0.4	1.8-3.3	4.0	0.3	3.5-4.3
		Student DAS GCA change	0.2	8.5	-20.0-20.0	3.1	8.2	-11.0-35.0
		Student age (months)	87.2	10.4	68.0-107.0	85.7	9.0	71.0-106.0
		Student baseline DAS GCA	58.7	21.1	20.0-91.0	67.5	20.7	21.0-94.0
		Teacher's years of autism experience	3.8	5.8	0-17.0	4.9	5.2	0-21.0
		Teacher's years of experience with STAR	0.2	0.4	0-1.0	0.4	0.5	0-1.0
		Number of workshop hours teacher attended	30.4	6.6	17.0-36.5	33.0	6.4	19.5-38.5
		Number of consultation hours teacher received	64.2	5.9	51.0-76.0	73.9	13.9	46.0-96.0
		Number of support staff in classroom	3.6	1.3	1.0-7.0	3.3	1.0	2.0-5.0
		Number of children in	7.5	1.6	4.0-9.0	7.6	1.7	6.0-12.0

		classroom						