ORIGINAL PAPER

Teacher Ratings of Acceptability of a Daily Report Card Intervention Prior to and During Implementation: Relations to Implementation Integrity and Student Outcomes

Erin Girio-Herrera[1] · Theresa E. Egan[2] · Julie Sarno Owens[3] · Steven W. Evans[3] · Erika K. Coles[4] · Alex S. Holdaway[2] · Clifton S. Mixon[5] · Hannah D. Kassab[3]

Abstract

The goals of the study were to (a) examine teacher-reported acceptability of a daily report card (DRC) intervention for a student in their classroom prior to and during implementation; (b) examine factors that predict acceptability; and (c) explore the relations between teacher-reported acceptability, student and teacher characteristics prior to implementation, implementation integrity (treatment dose, adherence, and teacher competence), and student outcomes. Participants were 39 elementary school teachers and 39 students with or at risk of attention-deficit hyperactivity disorder (ADHD). Teachers were asked to implement the DRC for up to 16 weeks with consultation support provided by research team staff every other week. Teachers completed acceptability ratings about the DRC prior to and after two months of implementation. This multi-method assessment using correlation and regression analyses revealed that although acceptability ratings prior to implementation were related to teacher knowledge of ADHD, they were was not related to acceptability ratings during implementation, integrity, or student outcomes. Student's initial positive response to the intervention (i.e., the magnitude of improvement in DRC target behaviors) was associated with higher acceptability ratings during the intervention. Greater increases in acceptability over time were associated with greater DRC dose (i.e., teacher compliance to procedures and longer DRC duration). Greater duration of implementation and responding appropriately to rule violations were associated with greater student achievement of DRC goals. Implications for interpreting acceptability ratings and for understanding factors related to implementation and outcomes are discussed.

Keywords Acceptability · Teacher · Implementation · Attention-deficit/hyperactivity disorder (ADHD) · Intervention · Daily report card

Introduction

Attention-deficit hyperactivity disorder (ADHD) affects 5–11% of school-aged children (Merikangas et al., 2010; Wolraich et al., 2014). Thus, most elementary classrooms contain one to two students with ADHD (Fabiano et al., 2013a, b). Relative to non-ADHD peers, students with ADHD often experience lower academic achievement (Volpe et al., 2006), higher rates of absenteeism (Classi, Milton, Ward, Sarsour, & Johnston, 2012) and grade retention (Loe & Feldman, 2007), and more conflict with peers and teachers (Greene, Beszterczey, Katzenstein, Park, & Goring, 2002; Hoza, 2007). Students with ADHD are more likely to use special education services and receive more disciplinary referrals than their non-ADHD peers (Robb et al., 2011). Given the negative effects of ADHD on the student and the

✉ Erin Girio-Herrera
  egirioherrera@towson.edu

1 Department of Psychology, Towson University, Towson, MD 21252, USA

2 Department of Child and Adolescent Psychiatry and Behavioral Sciences, Children's Hospital of Philadelphia, Philadelphia, PA, USA

3 Department of Psychology, Ohio University, Athens, OH, USA

4 Department of Psychology, Florida International University, Miami, FL, USA

5 Department of Pediatrics, Ochsner Hospital for Children, Ochsner Health System, New Orleans, LA, USA

school (e.g., financial costs, teacher stress), it is important for educators to be equipped with effective interventions to address the needs of these youth and for these interventions to be implemented with integrity.

Classroom-based behavior management interventions are classified as well-established interventions for elementary school students with ADHD (Evans, Owens, Wymbs, & Ray, 2018). The most widely studied of these is the daily report card (DRC) intervention. When using a DRC, teachers identify and define two to three target behaviors that cause impairment for the student, track those behaviors for one week, and use these baseline data to establish the initial, achievable goals for each target behavior (e.g., completes 75% of daily math problems, raises hand before speaking with four or fewer violations). During the day, the teacher tracks each behavior and provides feedback to the student about progress toward each goal. At the end of the day, DRC performance is reviewed with the student by teachers and parents and contingent privileges are provided. Using shaping procedures, each behavior is modified until it moves into the typical range for the student's age.

The DRC is effective for students with ADHD in general education (e.g., Owens et al., 2012) and special education (e.g., Fabiano et al., 2010) classrooms, is feasible for use over several months (e.g., Owens et al., 2012; Owens, Murphy, Richerson, Girio, & Himawan, 2008), and is effective in modifying academic and behavioral problems (e.g., Pyle & Fabiano, 2017; Vannest, Davis, Davis, Mason, & Burke, 2010). Further, there are incremental benefits of the DRC with each month of the intervention over four months (Holdaway et al., 2020; Owens et al., 2012). Despite this evidence, there is variability in the integrity with which teachers implement this (and other) classroom interventions which has implications for intervention effectiveness. For example, two studies found that, on average, teachers adhered to recommended DRC procedures on 77% of school days; yet the range in both studies was wide (0–98% in Fabiano et al.; 10–100% in Owens et al.). Given this variability, and the well-established link between intervention integrity and student outcomes (e.g., Conroy et al., 2015; Noell, Gresham, & Gansle, 2002), there is a need to better understand factors that predict integrity.

Treatment integrity is a critically important methodological consideration in ensuring validity in treatment outcome research and implementation science. It is a multifaceted concept with several definitions and models and is often overlooked in treatment outcome research (Cox, Martinez, & Southam-Gerow, 2019). A description and exploration of all factors associated with treatment integrity is beyond the scope of this manuscript; however, the complexity of integrity is noteworthy. Hagermoser Sanetti and Kratochwill (2009; Table 2) highlight several factors proposed to influence treatment integrity across four levels within the context of the school environment: the external environment, organization, intervention, and interventionist. Listed among the factors related to the interventionist are many that involve the interventionists' perceptions. For teachers involved in the implementation of the DRC, this could include their perceived need for intervention, effectiveness of intervention, role in implementation, and self-efficacy.

One factor that may affect teachers' implementation integrity is their perception of intervention acceptability, that is, the extent to which the intervention is viewed as appropriate, fair, and reasonable for the problem (Kazdin, 1981). Witt and Elliott (1985) proposed a model containing four elements (i.e., intervention acceptability, use, integrity, and effectiveness) that are sequential and reciprocal. They suggest that high intervention acceptability is associated with high intervention use, which is subsequently associated with high intervention integrity (i.e., adhering to recommended procedures). High intervention integrity is proposed to be related to high intervention effectiveness, which in turn facilitates high intervention acceptability, thus creating a dynamic interactive cycle. In this way, acceptability can be conceptualized as both a predictor of use and integrity, and an outcome of intervention effectiveness. Reimers, Wacker, and Koeppl (1987) expanded this model to include treatment knowledge and environmental disruption (i.e., interruption to typical activities) as additional factors influencing treatment acceptability. These models are foundational in the school-based literature and are echoed in the implementation science literature (see Proctor et al., 2009 for review).

Often treatment acceptability has not been evaluated in the context of environmental disruption. However, when teachers are asked to report on their acceptability of an intervention, it is valuable to capture their perceptions against the backgroup of their countless other responsibilities. One study showed that among teachers implementing student interventions (classroom and individual), those who demonstrated acceptable integrity (between 80 and 100% on an integrity checklist) were more likely to evaluate the interventions to be suitable and convenient; whereas those who implemented the interventions with very high (100%) or low integrity (less than 80%) rated the interventions less suitable and convenient (Harrison, State, Evans, & Schamberg, 2016). These findings suggest that the extent to which teachers pushed themselves (intensively, moderately, or slightly) to implement the intervention in light of competing priorities, may have an impact on the teachers' perceptions of treatment acceptability. Examining acceptability within the framework of these conditions offers a more realistic picture of teachers' true perceptions of interventions.

Previous studies have examined teachers' acceptability of a DRC; however, the conclusions drawn are limited due to the use of vignettes or hypothetical scenarios and the lack of measurement of acceptability during implementation (e.g.,

Gresham & Lopez, 1996). Further, despite calls to action in the general acceptability literature (e.g., Nastasi & Truscott, 2000), no studies have examined intervention acceptability of the DRC in relation to factors existing prior to implementation, implementation integrity, and outcomes. Measuring acceptability and teacher and student factors at multiple timepoints expands on previous studies by conceptualizing acceptability as a dynamic factor that is potentially related to prior factors and may be associated with subsequent implementation and student outcomes. The current study addresses these limitations by examining teachers' ratings of acceptability of a DRC prior to and during DRC implementation, and the relationships between DRC acceptability, teacher characteristics and student characteristics, integrity, and outcomes.

Intervention Acceptability Research

Studies measuring teachers' reports of acceptability of non-DRC treatments suggest that teachers demonstrate higher integrity and observe greater change in disruptive student behavior when they implement interventions they find more acceptable (Andersen & Daly, 2013). Johnson et al. (2014) demonstrated that teachers who implemented a *preferred or acceptable* intervention adopted the intervention more quickly, sustained higher implementation quality independent of coaching, and were more likely to continue implementing the intervention following study completion, as compared to teachers who implemented an *assigned* intervention. Thus, improving teachers' acceptability of interventions and/or designing interventions in collaboration with teachers (so that they are acceptable) may be effective strategies to enhance intervention adoption and to improve integrity and related student outcomes.

Previous studies examining teachers' treatment acceptability for behavioral interventions document that teachers generally find most evidence-based and promising interventions are acceptable (Briesch, Briesch, & Chafouleas, 2015; Elliott, Witt, Galvin, & Peterson, 1984; Girio & Owens, 2009; Power, Hess, & Bennett, 1995). Specific to the DRC, Girio and Owens (2009) reported that elementary school teachers endorsed the DRC as acceptable, and more acceptable than other interventions for addressing disruptive behavior (e.g., time out, medication, social skills), which is consistent with prior work (Pisecco, Huzinec, & Curtis, 2001; Power et al., 1995). However, most studies have examined teachers' acceptability under hypothetical conditions (e.g., rating acceptability after reading vignettes) rather than during implementation of an actual intervention (Gresham & Lopez, 1996). Among the few studies that have examined teacher report of acceptability of the DRC before or after actual implementation (e.g., Murray, Rabiner, Schulte, & Newitt, 2008; Williams, Noell, Jones, & Gansle, 2012),

ratings of acceptability remain high. However, none of these studies examined the association between the acceptability and integrity or student outcomes. Thus, it is not clear whether high acceptability is a predictor of teachers' integrity or an outcome of teachers' experiences with the DRC, or both, as hypothesized by Witt and Elliott (1985). Indeed, Gresham and Lopez (1996) once argued that "while informative… pretreatment acceptability may not correspond to what consumers might tell us about the acceptability of treatments after they have tried them" (p. 213). It may be more important to determine acceptability after the use of an intervention, as this experience and observation of child response likely influences continued use and perceptions of the intervention.

Factors Affecting Acceptability and Implementation

Given that there is some variability in teachers' acceptability ratings for a given intervention, prior research has investigated characteristics of the teacher and student prior to implementation as potential predictors of intervention acceptability. First, consistent with Reimers and colleagues' (1987) model, two reviews highlight studies that found a positive relation between teacher knowledge (e.g., of student problems or behavioral principles) and intervention acceptability (e.g., see Elliott, 1988; Han & Weiss, 2005 for review). Thus, we included a measure of knowledge of ADHD and hypothesized that it would be positively related to higher acceptability ratings and higher integrity and use (Vereb & DiPerna, 2004).

Second, some studies have found that teachers' years of experience and highest degree earned are positively related to acceptability of some classroom interventions (Girio & Owens, 2009; Vereb & DiPerna, 2004), whereas others have found no relation (Pisecco et al., 2001; Power et al., 1995). However, all of these studies occurred in the context of hypothetical vignette methods. Further, with other classroom interventions, years of teaching and highest degree earned were unrelated to the number of times the strategy was used (Domitrovich et al., 2015). Given this variability across studies, years of experience and degree earned were included but directional hypotheses were not made.

Lastly, in addition to teacher-level factors, the most salient student-level factor associated with teacher ratings of acceptability is severity of student problems. Multiple studies have found that teachers view behavioral interventions to be more acceptable and reasonable when student problems are severe (as compared to when student problems are mild or moderate), presumably because milder problems may not warrant the effort required to implement the intervention (Elliott et al., 1984; Martens, Witt, Elliott, & Darveaux, 1985). However, most of these studies used vignettes and there are a few studies with contrary findings (see Elliott,

1988 for review). Thus, examination of student severity under implementation conditions is warranted. We assessed student impairment in multiple domains of functioning and hypothesized that greater severity would be associated with higher acceptability.

Current Study

The first aim examines teacher-reported acceptability of a DRC intervention prior to and during implementation of a DRC for a student in their classroom. It was hypothesized that teacher-reported acceptability of the DRC would be high (Girio & Owens, 2009). The second aim examines factors that predict acceptability. The third aim explores the relation between teacher report of acceptability, baseline student and teacher characteristics, implementation integrity (dose, adherence, and teacher competence), and student outcomes. It was expected that acceptability both prior to and during treatment would be related to implementation integrity and student outcomes. Further, consistent with the interactive, dynamic models (Witt & Elliott, 1985; Reimers et al., 1987), it was hypothesized that initial positive student outcomes would be positively related to later acceptability and implementation integrity.

Method

Participants

Eight schools participated across two sites. In Ohio, the five participating schools had an average of 377 students and 16 general education teachers per school, with 12–29% of students receiving special education services and 35–75% receiving free or reduced lunch services. In Florida, the three participating schools had an average of 1024 students and 50 general education teachers with 4–11% receiving special education services and 76–95% receiving free or reduced lunch services. The racial makeup of schools was primarily Caucasian (range 90–98%) in Ohio and predominantly Latinx in Florida (range 94–98%).

Teacher participants were 39 elementary school teachers (19 from Ohio, 20 from Florida) teaching grades K through 5 who were participating in a multi-site consultation study designed to facilitate teachers' implementation of effective classroom management strategies and a DRC intervention (masked for review). Teachers represented in the current study are those who completed all measures prior to and during DRC implementation. Nineteen teachers (not part of the 39) were excluded because they had not completed the acceptability measure at one of the two time points (in five cases because the student moved before the second time point; in three cases because the teacher withdrew from the

study). Analyses of available data revealed that teachers included in the current sample did not differ from those excluded with regard to teacher or student gender, site, years of teaching experience, student impairment, DRC acceptability ratings, DRC adherence, competence ratings, or student initial response to the DRC. Excluded teachers were more likely (than included teachers) to be in third or fifth grade and, on average, implemented the DRC for fewer days. Included teachers were women (94.9%) and identified as Non-Hispanic White (53.8%) or Hispanic (any race; 43.6%). Included teachers had an average of 14.88 years (SD = 8.72) of teaching experience. Most (64%) had a master's degree or higher.

Target students were 39 elementary school students (76.9% male; 51.3% Hispanic). Most target students (92.3%) met criteria for ADHD (71.8% combined presentation; 17.9% inattentive presentation; 2.6% hyperactive/impulsive presentation) and the remaining 7.7% were at risk of ADHD (at least four symptoms plus teacher-rated impairment). The DRC is effective for a variety of presenting problems; thus, we allowed subclinical symptoms as long as there was impairment. Students had an average IQ estimate of 98.24 (SD = 13.43), as assessed by the Wechsler Abbreviated Scales of Intelligence, Second Edition (Wechsler, 2011). Family socioeconomic status was low to middle class (15.4% had a household income of under $15,000, 58.9% had an income between $15,000 and 49,999; and 18% were above $50,000; 7.7% did not report income). Per parent report at intake, 2.6% had been diagnosed with a learning disability and 23.1% had a medication prescription for ADHD.

Procedures

The study was conducted at two universities and procedures were approved by the Institutional Review Boards at both and within all participating school districts. A complete description of procedures can be found in Owens et al. (2017). See Fig. 1 for a list of study constructs and the timeline for data collection. All general education teachers in each elementary school were invited to a 3-h workshop conducted by the investigators that focused on best practices in general classroom management strategies and the DRC. At the end of the workshop, teachers completed the questionnaires described below. Teachers interested in participating in consultation were required to identify one student with or at risk of ADHD; consent was required by teacher and parent, and assent was required by the student. Inclusion criteria for being a target student were the following: (a) enrolled in a general education classroom (K-5) for at least 50% of the day, (b) IQ estimate that fell in or above the 90th percentile confidence interval for a score of 80, and (c) met diagnostic criteria for DSM-IV ADHD or were at risk of ADHD. ADHD was defined as the presence of six or more
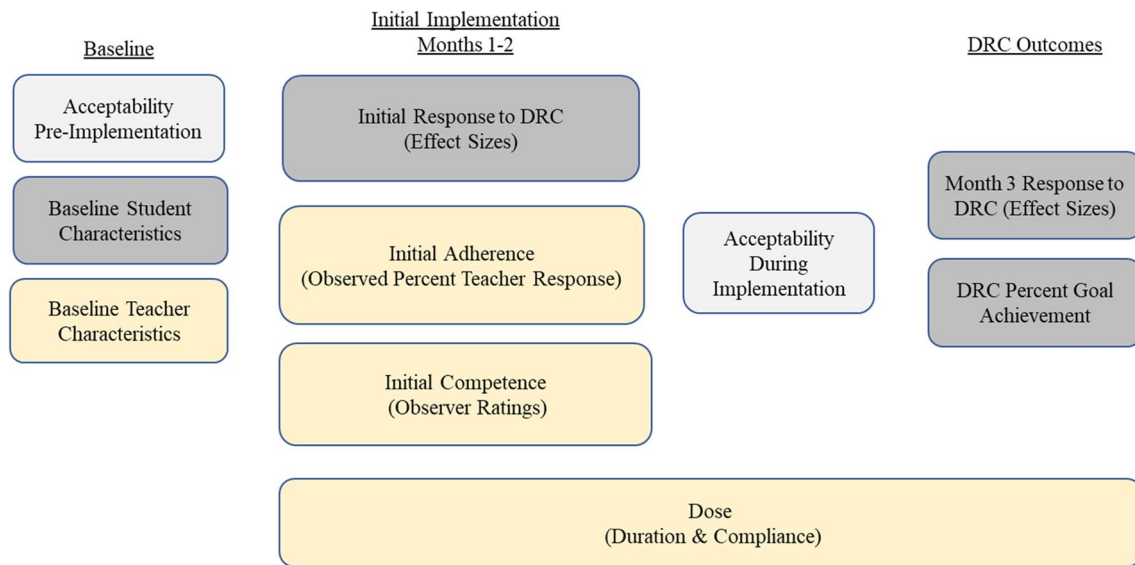
Fig. 1 Temporal representation of study constructs

symptoms of inattention and/or hyperactivity/impulsivity as reported by parents on the *Children's Interview for Psychiatric Syndromes-Parent Version* (P-ChIPS; Fristad, Teare, Weller, Weller, & Salmon, 1998) or the parent or teacher version of the *Disruptive Behavior Disorders Rating Scale* (Pelham, Gnagy, Greenslade, & Milich, 1992), and teacher-rated impairment as defined by a rating of at least 3 on the *Impairment Rating Scale* (Fabiano et al., 2006). Information obtained from the P-ChIPs helped to rule out other disorders as sources of ADHD symptoms and to assess symptom chronicity. At-risk status was defined as four or more symptoms and teacher-rated impairment. Students were excluded if a previous diagnosis of autism spectrum disorder, bipolar disorder, or intellectual disability was reported by the parent.

Once a target student was identified, teachers were asked to implement the DRC for 16 weeks. Teachers completed the questionnaires described below a second time after two months of DRC intervention implementation. Teachers were paid for attending the inservice and completing questionnaires, but did not receive compensation for participating in consultation sessions or for implementation of any classroom management practices.

Consultation Procedures

For the purposes of a clinical trial, stratified random sorting was used to assign teachers to two consultation conditions (Owens et al., 2017). The standard condition was designed to represent best practices in school psychology. It followed a problem-solving process (Frank & Kratochwill, 2014) and included brief performance feedback that mirrored best practice procedures reported in previous research (Gilbertson,

Witt, Singletary, & VanDerHeyden, 2007; Noell, Witt, Gilbertson, Ranier, & Freeland, 1997). In the multi-component condition, consultants followed the problem-solving process described above, but also assessed and attempted to address possible barriers to integrity using the knowledge, skills, and beliefs components (Owens et al., 2017). Both conditions focused on creation of the DRC, use of the DRC, general classroom management strategies, receipt of performance feedback, discussion of implementation, and problem-solving. Sessions were 30 min to 1 h and conducted before, during, or after school. Although adequate differentiation of conditions and an equal number of consultation sessions across conditions was achieved (Owens et al., 2017), for the current study, teachers were combined across conditions. Teachers in the two conditions did not differ on acceptability prior to the intervention implementation (condition 1: $M = 5.18$, $SD = .62$; condition 2: $M = 5.26$, $SD = 1.02$; $t(37) = -.275$, $p = .79$).

Measures

Student Impairment

The teacher version of the Impairment Rating Scale (IRS; Fabiano et al., 2006) assesses teacher perceptions of student functioning on a 7-point scale that ranges from 0 (*No problem, Definitely does not need treatment*) to 6 (*Extreme problem, Definitely needs treatment or special services*). This was completed at the point of student referral to the project. The domains assessed include relationship with peers, relationship with teacher, academic progress, the classroom in general, self-esteem, and overall. With elementary school

samples, the measure has respectable cross-informant reliability, convergent and divergent validity with other impairment scales, and predictive validity in identifying students with ADHD diagnoses (Fabiano et al., 2006) when scores are three or higher.

Teacher Demographic Information

Teachers were asked to provide the number of years they have taught and their highest degree earned.

Teacher Knowledge of ADHD

The teacher knowledge of ADHD questionnaire is a 24-item true/false/don't know measure that assesses teacher knowledge of ADHD and best practices in the treatment of ADHD. Responses were coded as correct or incorrect. Don't know was coded as incorrect. A total percent correct was calculated. The measure was developed by the authors of Owens et al. (2017) clinical trial and was inspired by the measure developed by Jones and Chronis-Tuscano (2008). The measure has demonstrated sensitivity to change as a function of participating in a workshop focused on ADHD and classroom management (Owens, Coles, & Evans, 2014). Because this measure was completed after the workshop, scores reflect that all teachers were given an equal opportunity to have foundational knowledge about ADHD.

Intervention Acceptability

Intervention Rating Profile-10 Item Version (IRP-10; Power et al., 1995) was used to assess teacher perceptions of DRC acceptability prior to implementation and two months after implementing the DRC. Items are rated on a 6-point scale that ranges from 1 (*Strongly Disagree*) to 6 (*Strongly Agree*). Ratings for each item are mean-averaged to yield a total score reflecting a single dimension of acceptability. Higher scores indicate higher acceptability of that treatment. The measure includes a few general questions such as, "I liked the procedure used in this intervention" and "I would suggest the use of this intervention to other teachers." However, the majority of items are specific to the student such as, "This would be an acceptable intervention for [Child's] school difficulties;" "Overall, this intervention would be beneficial for [Child];" "This intervention would not result in negative side effects for [Child]:" and "[Child's] school problems are severe enough to warrant use of this intervention." The IRP-10 has excellent internal consistency with alpha coefficients ranging from .95 to .97 (Power et al., 1995) and has evidence of discriminant validity, as it can identify interventions of varying acceptability among teachers

(Girio & Owens, 2009; Martens et al., 1985). In the current sample, internal consistency estimates were .96 prior to and .91 during implementation.

Initial DRC Response

To examine the student's initial response to the DRC, we calculated individual effect sizes (standard mean difference; SMD) that represented the cumulative benefit of the DRC at the end of Month 1 and the end of Month 2. This 2-month time frame was selected based on a previous study that demonstrated that (a) large effects could be detected after one month of implementation and (b) responders to the intervention were highly distinguishable from non-responders (Owens et al., 2012) after two months of implementation. For the cumulative effect size (ES), the SMD represents the difference between the mean of a follow-up period (i.e., Month 1, Month 2) and the mean of the period prior to implementation (i.e., baseline tracking) divided by the standard deviation of the period prior to implementation. By this definition, SMD can be interpreted as the mean improvement of a participant at a follow-up period, compared to the baseline period, as a function of the variability during the baseline period. This procedure is consistent with other studies that examined the incremental benefit of the DRC (Holdaway et al., 2020; Owens et al., 2012) and different medication doses for students with ADHD (e.g., Evans et al., 2001). For this study, the average ES from Month 1 and Month 2 for the student's first two DRC targets was calculated, as all students had at least two DRC targets and these typically represented the teachers top concerns.

DRC Dose

DRC implementation requires the teacher to give student feedback when a rule violation occurs (e.g., *Carlos, that's an interruption*) and make a tally for that rule violation on the DRC. Teachers were asked to either insert these data into a website that produced graphs of performance or give these data to the consultant who assisted with data entry. Dose was calculated in two ways (compliance and duration) based on this data. Compliance was defined as the number of days in which DRC data were submitted divided by the number of days data could have been submitted (i.e., all days for which the student was present at school during the 2-month DRC implementation period). This was calculated for the first two targets on the student's DRC, then averaged. For duration, the number of days between DRC initiation and the last day of DRC data submission was calculated. Data on compliance and duration were gathered across the academic year and serve as dependent variables.

DRC Adherence

Teachers were observed using the Student Behavior-Teacher Response Observation Rating System (SBTR; Pelham, Greiner, & Gnagy, 2008). This is a systematic class-wide observation system developed using a behavior theory framework and intended to capture discrete student–teacher interactions in preschool and elementary classrooms. Previous studies have shown SBTR to have adequate interrater reliability and convergent validity (Fabiano et al., 2013; Vujnovic, Holdaway, Owens, & Fabiano, 2014), as well as sensitivity to change as a function of intervention (Fabiano et al., 2010; Owens et al., 2017). The SBTR observation manual includes definitions, and inclusion and exclusion examples for coding student violations of seven common classroom rules (i.e., be respectful, obey adults, work quietly, use materials appropriately, remain in seat, raise hand to speak, stay on task), and the teacher's response to each violation (i.e., coded as appropriate, inappropriate, or no response). All definitions are available upon request from the first author. SBTR observers obtained frequency counts of (a) all rule violations by the target student, (b) how the teacher responded to each of those types of violations, (c) all DRC violations, and (d) how teachers responded to each of those types of violations (i.e., responded by labeling the DRC behavior or not, *Carlos, that's an interruption*).

Two variables from the SBTR observations were used to represent adherence for the current study: (a) teacher percent appropriate response to student rule violations and (b) teacher percent appropriate response to DRC violations. To calculate these variables, the total number of appropriate teacher responses to each violation for a given observation was divided by the total number of the respective violations by the student for that observation period. These percentages were captured for Month 1 and Month 2 and then averaged. Both variables serve as a predictor variable for the analyses.

Observers were trained to reliability on the SBTR. They attended an initial training, were required to pass (100% accurate) a written definitions test, (100% accuracy) coding of 2–5-min video clips, and achieve at least 80% reliability across all coded behaviors in a classroom with a master observer. Maintenance of reliability was checked across the year. Interobserver assessments were conducted for 24% of all observations in the clinical trial. To assess the interrater reliability, calculations were completed for Intraclass Correlations (ICC) of type 1 for average of k raters (that is ICC(1,k)). Across all frequency count variables, the ICC(1,k) ranged from .88 to .94 with an average of .93.

DRC Competence

Once a target child was identified and teacher consent was obtained, at least two classroom observations (using SBTR) were conducted to obtain a baseline assessment of each teacher's competence in classroom management. Competence was evaluated in several domains (e.g.,. response to all inappropriate behaviors, response to DRC-related behavior, and global competence). Observers considered facets within each domain (e.g., timing, specificity, tone of voice, consistency) and assigned a competence rating on a 10-point scale. For the current study, competence in relation to teachers' response to inappropriate behavior and DRC-related behaviors (averaged across Month 1 and Month 2 observations) were used as predictor variables. The ICC(1,k) were $\geq$ .88, respectively.

DRC Outcomes

We examined DRC outcomes in two ways. First, we examined the incremental SMD effect size for Month 3 (i.e., the month following completion of acceptability ratings). The incremental SMD for Month 3 represents the difference between the mean of the data during the last two weeks of Month 3 and the mean of the data during the last two weeks of Month 2 divided by the standard deviation of the period prior to implementation. The SMD can be interpreted as the average improvement during Month 3 compared to Month 2, as a function of the variability at baseline. This is an indicator of student improvement after acceptability ratings were given. Second, we examined a global indicator of the student's success with the DRC, namely the student's overall achievement of all DRC goals over the duration of implementation. We included all goals, so that this represented the overall achievement with this intervention. It was defined as the number of days the student met their DRC goal divided by the number of days DRC data was implemented. For example, if a student had a goal of "five or fewer interruptions" and the student achieved five or less interruptions on 30 of 50 days DRC data were submitted, the student's achievement of this DRC target would be 60%. Then, this percentage was averaged across all DRC target behaviors.

Results

Aim 1: Acceptability Prior to and During Implementation

On average, teacher-reported acceptability prior to implementation was 5.23 (range 1–6) and during implementation was 5.20 (range 3–6). Descriptive statistics for all other study variables can be found in Table 1. As shown in Table 1, all variables had adequate variability for inclusion in the regression models.

Table 1 Descriptive statistics for study variables

Variable	M (SD) ($N=39$)	Range
Acceptability		
IRP pre-intervention	5.23 (.86)	1–6
IRP during intervention	5.20 (.74)	3–6
Baseline variables		
Years teaching	14.88 (8.72)	1–36
% with Master's degree or higher	64.0	
ADHD knowledge (% correct)	83.9 (10.9)	63–100
IRS-Tch–student relationship	3.62 (1.44)	0–6
IRS-Academic	4.54 (1.55)	0–6
IRS-Classroom	4.08 (1.61)	0–6
Initial DRC response		
Avg Cum ES Targets 1 & 2 @ Mos 1 & 2	.67 (1.07)	− 2.77–2.90
DRC dose		
Compliance for targets 1 & 2 (%)	88.7 (10.5)	54–100
Duration (number of days)	55.61 (19.34)	18–87
DRC adherence		
Avg respond appropriately Mos 1 & 2 (%)	37.36 (26.82)	0–100
Avg label DRC Mos 1 & 2 (%)	55.61 (33.55)	0–100
DRC competence		
Avg respond Inappr Behav Mos 1 and 2	5.83 (2.70)	3.50–9.25
Avg respond DRC Behav Mos 1 and 2	6.99 (1.25)	1–10
DRC outcome		
Avg incremental ES targets 1 and 2 @ Mo 3	.45 (.97)	− 1.5–2.45
Overall achievement of DRC goals (%)	65.79 (18.38)	19–97

N = 39

IRP Intervention Rating Profile-10, *IRS* Impairment Rating Scale, *Tch* Teacher, *DRC* daily report card, *ES* standard mean difference effect size, *Avg* average, *Cum* cumulative

Aim 2: Factors that Predict Acceptability During Implementation

First, we examined correlations between all theoretically relevant variables, i.e., teacher report of acceptability, student and teacher characteristics prior to implementation, and integrity variables (dose, adherence, and competence and student outcomes). See Table 2. Variables that are significant ($p < .05$) were included in each regression model and are described below.

The first hierarchical linear regression model examined factors that predict DRC acceptability during implementation. The following baseline and early implementation factors correlated significantly with DRC acceptability during implementation and were included as possible predictors: teacher knowledge of ADHD, teacher-reported student impairment in classroom functioning, and students' initial

response to treatment (average effect size for behavioral targets 1 and 2 across Month 1 and 2). Pre-implementation acceptability was entered hierarchically into the first step, and the other predictors were entered together in the second step (see Table 3). The first step (pre-implementation DRC acceptability) was not significant. The second step was significant, $F(4)=5.83$, $p < .01$, accounting for 41% of the variability in DRC acceptability during implementation. Notably, student's initial response to treatment (effect sizes) was a significant predictor ($p < .01$), and the beta (.44) indicated that larger effect sizes (greater improvement over Month 1 and 2) were associated with higher acceptability ratings at Month 2 (see Table 3). ADHD knowledge and teacher-rated student impairment in classroom functioning (prior to implementation) were marginally significant ($p < .07$).

Of note, although the dose variables (compliance and duration) correlated with acceptability during intervention, they were not included in the model as these variables were measured across the academic year. Further, one adherence variable (teacher percentage of responding appropriately) was correlated with acceptability. However, adherence and competence in Months 1 and 2 could be considered either a predictor of acceptability during implementation or a response to acceptability. Including the responding appropriately variable as a predictor did not modify the results above; thus, the simplest model was presented.

Aim 3: Relationship between Acceptability, Implementation, and Student Outcomes

This aim was examined by exploring student outcomes (the student's initial and overall improvement and achievement of DRC goals) and the relation between teacher report of acceptability, baseline teacher characteristics (e.g., knowledge, years of experience, highest degree earned), baseline student characteristics (e.g., initial severity of impairment), and implementation integrity (dose, adherence, and teacher competence).

Two variables had significant correlations with incremental improvement in Month 3: the number of years teaching and DRC compliance (see Table 2). This model, $F(2)=5.50$, $p < .01$, accounted for 31% of the variance in the incremental improvement in Month 3. The number of years teaching was a significant predictor (Beta $= − .40$, $p < .05$), indicating less years of experience was associated with more improvement in Month 3 (see Table 4).

The following variables had significant correlations with overall achievement of DRC goals: DRC acceptability during treatment, student initial response to treatment (effect sizes), compliance, duration, and adherence (i.e., responding appropriately to rule violations). Because overall achievement of DRC goals is partially a function of initial response to treatment, this variable was not included

Table 2 Correlations between study variables

	1. IRP Pre	2. IRP During	3. Yrs Teach	4. Highest Degree	5. ADHD Know	6. IRS Tch-Stud	7. IRS Acad	8. IRS Class	9. Avg ES Mos 1 & 2	10. DRC Comp	11. DRC Dur	12. Avg Resp Appr	13. Avg Lab DRC	14. Res Inapp Beh	15. Res DRC Beh	16. Avg ES Mo 3
Acceptability																
1. IRP Pre	–															
2. IRP During	– .02	–														
Baseline variables																
3. Years Teaching	.01	.04	–													
4. Highest Degree	.13	.11	.47**	–												
5. ADHD Knowledge	.38*	.33*	– .02	.10	–											
6. IRS-Tch-student	– .10	– .10	.01	– .15	.13	–										
7. IRS-Academic	– .09	– .10	– .06	.09	– .03	.48**	–									
8. IRS-Classroom	.03	.41**	.11	.12	.34*	.25	.12	–								
Initial DRC response																
9. Avg ES Target 1 & 2 M1 & M2	.01	.46**	.09	– .24	– .03	.01	– .12	.10	–							
DRC dose																
10. Compliance	– .07	.40*	– .02	.08	.09	– .15	.03	.16	.46**	–						
11. Duration	.12	.51**	– .16	– .27	.11	– .22	– .20	.06	.37*	.58**	–					
DRC adherence																
12. Avg % Resp Appr M1 & M2	.07	.34*	.09	– .00	.01	– .02	– .23	.17	.46**	.16	.20	–				

Table 2 (continued)

	1. IRP Pre	2. IRP During	3. Yrs Teach	4. Highest Degree	5. ADHD Know	6. IRS Tch-Stud	7. IRS Acad	8. IRS Class	9. Avg ES Mos 1 & 2	10. DRC Comp	11. DRC Dur	12. Avg Resp Appr	13. Avg Lab DRC	14. Res Inapp Beh	15. Res DRC Beh	16. Avg ES Mo 3
13. Avg % Label DRC M1 & M2	−.07	.06	−.07	−.11	−.00	.28	.04	−.08	.05	−.02	−.30	−.06	–			
DRC competence																
14. Avg Resp Inappr Behav M1 & M2	.30	.21	−.18	.23	.39*	.17	.04	−.07	.26	.22	.00	.26	.41*	–		
15. Avg Resp DRC Behav M1 & M2	.13	.24	−.16	.30	.14	.04	−.04	−.06	.34*	.34*	.05	.33*	.14	.71**	–	
DRC outcome																
16. Avg ES Target 1 & 2 M3	−.11	.06	−.50**	−.22	.08	.13	−.04	.35	−.15	.41*	.16	.06	−.06	.27	.06	–
17. DRC Achieve	−.16	.44**	−.20	−.30	−.03	−.12	−.38*	.16	.60**	.35*	.58**	.46**	−.30	.03	.18	.29

N = 39

IRP Intervention Rating Profile-10, *IRS* Impairment Rating Scale, *Know* knowledge, *Tch* teacher, *Stud* student, *Acad* academic, *Class* classroom, *DRC* daily report card, *ES* standard mean difference effect size, *M1* Month 1, *M2* Month 2, *M3* Month 3

$^+ p < .07$; $^* p < .05$; $^{**} p < .01$

Table 3 Factors predicting DRC acceptability during intervention

Variable	F	ΔR^2	β	B
Step 1	.01	.00		
IRP—pre-implementation				− .02
Step 2	5.83**	.41**		
IRP—pre-implementation			− .14	− .12
Teacher ADHD knowledge			.30[+]	.02
Student classroom impairment (IRS)			.27[+]	.12
Initial DRC response (effect sizes)			.44**	.31

Impairment ratings are teacher-rated prior to implementation

IRP Intervention Rating Profile, *IRS* Impairment Rating Scale, *DRC* daily report card

[+]$p < .07$; *$p < .05$; **$p < .01$

Table 4 Factors predicting DRC outcomes

Variable	F	R^2	β	B
Improvement in Month 3	5.50*	.31		
Years teaching			− .40*	− .05
DRC compliance			.26	.04
Overall DRC target achieve @ M3	6.82**	.46		
IRP during implementation			.03	.87
DRC compliance			.33	.10
DRC duration			.46*	.43
			.35*	
Avg responds appropriately M1&2				24.04

IRP Intervention Rating Profile, *DRC* daily report card, *M1&2* Month 1 and 2

[+]$p < .07$; *$p < .05$; **$p < .01$

in the model. Results showed this model to be significant, $F(4) = 6.82$, $p < .001$, accounting for 46% of the variability in overall achievement (see Table 4). Specifically, duration was a significant predictor (Beta = .46, $p < .05$), such that greater implementation duration was associated with greater achievement on DRC goals. The percentage of teachers' appropriate responding was also predictive (Beta = .35, $p < .05$) with higher percentages of appropriate responding associated with greater achievement on the DRC goals (see Table 4).

Post Hoc Analyses

In most previous studies (Girio & Owens, 2009; Murray et al., 2008; Power et al., 1995; Williams et al., 2012) teacher-rated acceptability of the DRC has been quite high (average scores of 4.5 or higher on a 6-point scales; with standard deviations less than 1). Given that ratings in the current study at both times were similarly high, but differentially predictive of other constructs, we conducted post hoc analyses to better understand this finding. Namely, we created difference scores (acceptability scores during implementation minus pre-implementation acceptability scores) and examined the patterns of change. There were 22 teachers (56%) who had stable or increasing scores (i.e., DRC became more acceptable over time) and 17 (44%) teachers who had lower scores during implementation than prior to implementation (i.e., DRC became less acceptable over time). Difference scores were not correlated with years of teaching experience, highest degree obtained, ADHD knowledge, teacher-rated student impairment prior to implementation, or change in student impairment. However, difference scores were related to DRC compliance ($r = .43$, $p < .01$) and duration ($r = .49$, $p < .01$), such that greater increases in acceptability over time were associated with greater DRC compliance and longer DRC duration.

Discussion

The goal for this study was to evaluate teachers' acceptability of a DRC intervention both before and during implementation and in relation to student and teacher characteristics, initial DRC success, intervention integrity, and student outcomes. Our results indicate that levels of initial intervention acceptability were not predictive of initial DRC response, dose, adherence, and competence, or DRC outcomes. However, in support of the dynamic models of implementation (Reimers et al., 1987; Witt & Elliott, 1985), the student's initial positive response to the intervention predicted higher acceptability ratings during the intervention, which was correlated with longer use of the intervention, and longer duration then predicted greater DRC achievement. Further, fewer years of teaching experience was predictive of greater student improvement later in the school year. Lastly, longer DRC duration and more appropriate teacher responding predicted greater overall student success with this intervention. Implications for interpreting are discussed.

We found that teachers' knowledge of ADHD prior to implementation was the only characteristic that was significantly associated with initial acceptability. This suggests that prior knowledge of ADHD or the knowledge provided at the inservice may have positively impacted teachers' perceptions of the utility of the DRC and may have provided some rationale for implementing it. However, teacher report of acceptability prior to DRC implementation was not related to any implementation variable or student outcomes. Although educating teachers about students' behavior problems and effective interventions may help with initial uptake of an intervention, it is likely insufficient to impact use over time. This is a unique contribution to the literature and suggests that research on intervention acceptability obtained via vignette methods or prior to the teacher implementing

the intervention should be interpreted with caution. Future studies should examine the relation between acceptability ratings after the decision to initiate a DRC and subsequence implementation.

In contrast, teacher report of acceptability that follows experience with the intervention, appears to play an important role in implementation and student outcomes. In support of Witt and Elliott's (1985) dynamic model, we found that greater initial intervention effectiveness (student's initial response to the DRC) was associated with higher intervention acceptability at Month 2, which in turn, was associated with higher intervention dose (i.e., longer duration). This suggests that the student's positive initial response to the DRC may have been reinforcing to the teacher, thus motivating continued use. In further support of the model, greater duration and responding appropriately by the teacher to rule violations (adherence) were associated with greater improvement on DRC goals later in the school year, which may have fed back into perceptions of intervention acceptability. Indeed, our post hoc analyses showed teachers whose acceptability increased over time implemented with greater integrity and for a longer period of time. This support for the dynamic cycle is promising and may suggest that teachers need little support or consultation when interventions are producing positive outcomes. However, these relations also suggest that poor response from one student and/or limited implementation integrity could derail the intervention and teachers' acceptability of using it. This is consistent with literature suggesting that teachers benefit from ongoing consultation (Owens et al., 2020). It also highlights the importance of consultant awareness of these relations so they can monitor student response, implementation integrity, and the teacher's report of acceptability, as initial deterioration in any of these may begin an undesirable negative spiral in reducing intervention effectiveness.

Our results raise some concerns related to the suggestion to allow teachers to "test-drive" and choose interventions to determine their acceptability (Dart, Cook, Collins, Gresham, & Chenier, 2012). Some research shows that allowing teachers' choice as to which intervention to use has resulted in higher levels of integrity to treatment, longer sustained implementation, and subsequent better outcomes when compared to teachers who were assigned an intervention (Andersen & Daly, 2013; Johnson et al., 2014). However, it may be important to postpone that choice until after at least 1 month of implementation, when a more fully informed decision can be made. Indeed, there is some evidence in the consultation literature (Fischer et al., 2016), that when teachers gain experience with specific intervention procedures (video conference technology, in this study), their ratings of acceptability for those procedures increases. Thus, encouraging teachers to test-drive the DRC and then consider whether to continue use may be an effective approach, as long as

the test-drive process includes consideration of treatment integrity that would be sufficient enough to lead to a positive student response. Ultimately, more research is needed to better understand integrity within this context and offer specific criteria for what may be considered "sufficient."

Lastly, our analyses suggest that some teacher and student characteristics may play a role in this model as well. Although no teacher factors were significantly associated with acceptability in the current study after controlling for pre-acceptability ratings, teacher knowledge of ADHD was marginally significantly associated with mid-treatment ratings of acceptability. Although knowledge of ADHD and its relation to acceptability have been examined in the past (e.g., Elliott, 1988 and Han & Weiss, 2005), this is the first study that has considered the role of teacher knowledge of ADHD as it relates to acceptability during intervention. The marginally significant results suggest that this factor is worth exploring in more detail and in further studies related to teachers' acceptability following the start of intervention.

Our results did show that fewer years of teaching was associated with greater student improvement later in the school year. We speculate that perhaps younger teachers or less experienced teachers found more incremental benefit of the DRC above the strategies they were already using or are more willing to comply with solutions offered by or generated with consultants. However, across studies, there has been inconsistency in findings related to years of teaching and acceptability (e.g., Girio & Owens, 2009; Power et al., 1995; Vereb & DiPerna, 2004). This study provided the opportunity to examine the relation when teachers consider a student with whom they are implementing the DRC (rather than a vignette) and within the context of acceptability during the academic year (rather than only prior to intervention). Despite these improvements in external validity, a relation between years of experience and acceptability was still not found. One reason for the lack of relation and inconsistent findings across the literature may be the possibility of a curvilinear relations between these variables. Unfortunately, due to sample size, this study was unable to power analyses to examine a curvilinear relation or conduct mediation analyses.

To better understand the relation between teacher experience and acceptability, perhaps future research could examine the role of teacher experience further by (a) continuing to examine acceptability beyond the start of intervention; (b) determining if a third, unmeasured variable could explain the relation between years of teaching and acceptability; (c) exploring the possibility of a curvilinear relation; and (d) examining whether teachers with less experience are more amenable to consultation or are more readily able to modify their behavior to obtain student improvement, as teacher receptivity to consultation serves as a conduit to treatment integrity of evidence-based practices (Owens et al., 2017).

This study identified that one student factor, student classroom impairment, may contribute to the model. Consistent with prior research (Elliott et al., 1984; Martens et al., 1985; Witt, Moe, Gutkin, & Andrews, 1984), more severe student classroom impairment was marginally associated with higher ratings of intervention acceptability during implementation after controlling for pre-implementation acceptability. Again, although these relations have been identified, they have done so in the context of early acceptability before intervention begins. Further research to explore the role of student impairment during intervention is warranted.

Treatment outcome researchers must continue to expand the trend toward improvement in the quality and extent to which integrity is examined (e.g., Cox et al., 2019). Pintello (2020) discusses these advances and provides commentary and a table that offers a glance at a framework and themes that correspond to core implementation science principles and future research. Future research in this area could benefit from advances in implementation science, particularly as it may relate to the complexity of measurement. Teachers' perceptions of the acceptability of the DRC is multifaceted and complex. Thus, it is challenging to measure, as well as account for other factors that could influence teachers' integrity that were not examined in this study (e.g., self-efficacy, teacher expected outcomes). Chafouleas and colleagues developed the Usage Rating Profile-Intervention (URP-I) to evaluate factors that influence implementation of innovations (Chafouleas, Briesch, Riley-Tillman, & McCoach, 2009). Importantly, the URP-I includes items to assess dimensions such as acceptability, but other constructs such as feasibility and collaboration. The URP was developed for school-based contexts and is modified for use across assessment and intervention by range of stakeholders (e.g., interventionists, parents). This team offers access to this profile, as well as information on history and data on the URP-I, including articles and examples of its application via their website. Further examination using the URP-I or other tools to assess acceptability and related dimension is needed to understand teacher implementation of interventions in the classroom.

This study's strengths must be interpreted in light of its limitations. First, the sample size of this study limited the types of analyses we could conduct and, thus, restricted the research questions we could address (e.g., mediation analyses) and the number of variables able to be explored in our regression analyses. Second, there are many student characteristics we did not measure that could influence the relations we studied. For example, students' sensitivity to reward could influence response and thus, also affect teacher report of acceptability. In addition, the consistency with which parents provided additional contingent privileges or other rewards based on student response to the DRC was not measured and may also be an important

factor in this process. Third, we did not assess possible contextual or organizational factors or other constructs identified in the implementation science literature (e.g., self-efficacy; outcome expectancy) that may have influenced acceptability ratings. Lastly, we did not assess teacher preferences relative to other treatment options. Realistically, teachers are engaging in a cost–benefit analysis when determining how much they like an intervention or will implement the approach. Thus, more advanced methods such as conjoint analysis (Cunningham et al., 2014; Egan et al., 2019) used in the context of real implementation choices may shed additional light on which factors may be most salient in determining acceptability.

In conclusion, to date, most research on teachers' acceptability of classroom interventions for youth with or at risk of ADHD has been conducted in the context of vignettes and no studies have examined intervention acceptability of the DRC in relation to factors existing prior to implementation, implementation integrity, and outcomes. The current study builds upon previous work by capturing acceptability prior to and during DRC implementation and examining the relations between acceptability, implementation, and outcomes in the context of intervention delivery. The findings suggest that researchers and consultants should be cautious about interpreting teachers' acceptability prior to the use of an intervention. In contrast, acceptability ratings after gaining experience with an intervention may be valuable in predicting long-term use of the intervention. Lastly, consultants must be aware on the cyclical nature of these variables, engage in ongoing assessment of these variables over time, and be prepared to collaboratively address initial deterioration in any of these areas, as they may predict an undesirable negative spiral in student outcomes.

Compliance with Ethical Standards

Conflict of interest All authors declare that they have no conflict of interest.

Ethical Standards Procedures were approved by the Institutional Review Boards at both universities and within all school districts. All procedures were performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

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

References

Andersen, M., & Daly, E. J. (2013). An experimental examination of the impact of choice of treatment components on treatment integrity. *Journal of Educational and Psychological Consultation, 23,* 231–263.

Briesch, A. M., Briesch, J. M., & Chafouleas, S. M. (2015). Investigating the usability of classroom management strategies among elementary schoolteachers. *Journal of Positive Behavior Interventions, 17*(1), 5–14.

Chafouleas, S. M., Briesch, A. M., Riley-Tillman, T. C., & McCoach, D. B. (2009). Moving beyond assessment of treatment acceptability: An examination of the factor structure of the Usage Rating Profile-Intervention (URP-I). *School Psychology Quarterly, 24*(1), 36.

Classi, P., Milton, D., Ward, S., Sarsour, K., & Johnston, J. (2012). Social and emotional difficulties in children with ADHD and the impact on school attendance and healthcare utilization. *Child and Adolescent Psychiatry and Mental Health, 6,* 1–8.

Conroy, M. A., Sutherland, K. S., Algina, J. J., Wilson, R. E., Martinez, J. R., & Whalon, K. J. (2015). Measuring teacher implementation of the BEST in CLASS intervention program and corollary child outcomes. *Journal of Emotional and Behavioral Disorders, 23,* 144–155.

Cox, J. R., Martinez, R. G., & Southam-Gerow, M. A. (2019). Treatment integrity in psychotherapy research and implications for the delivery of quality mental health services. *Journal of Consulting and Clinical Psychology, 87,* 221–233.

Cunningham, C. E., Barwick, M., Short, K., Chen, Y., Rimas, H., Ratcliffe, J., et al. (2014). Modeling the mental health practice change preferences of educators: A discrete choice conjoint experiment. *School Mental Health, 6,* 1–14.

Dart, E. H., Cook, C. R., Collins, T. A., Gresham, F. M., & Chenier, J. S. (2012). Test driving interventions to increase treatment integrity and student outcomes. *School Psychology Review, 41*(4), 467.

Domitrovich, C. E., Pas, E. T., Bradshaw, C. P., Becker, K. D., Keperling, J. P., Embry, D. D., et al. (2015). Individual and school organizational factors that influence implementation of the PAX good behavior game intervention. *Prevention Science, 16*(8), 1064–1074.

Egan, T. E., Wymbs, F. A., Owens, J. S., Evans, S. W., Hustus, C., & Allan, D. M. (2019). Elementary school teachers' preferences for school-based interventions for students with emotional and behavioral problems. *Psychology in the Schools, 56,* 1633–1653.

Elliott, S. N. (1988). Acceptability of behavioral treatments: Review of variables that influence treatment selection. *Professional Psychology: Research and Practice, 19,* 68–80.

Elliott, S. N., Witt, J. C., Galvin, R., & Peterson, M. (1984). Acceptability of positive and reductive interventions: Factors that influence teachers' decisions. *Journal of School Psychology, 22,* 353–360.

Evans, S. W., Owens, J. S., Wymbs, B., & Ray, R. (2018). Evidence-based psychosocial treatments for children and adolescents with attention-deficit/hyperactivity disorder. *Journal of Clinical Child and Adolescent Psychology, 47,* 157–198.

Evans, S. W., Pelham, W. E., Smith, B. H., Bukstein, O., Gnagy, E. M., Greiner, A. R., et al. (2001). Dose–response effects of methylphenidate on ecologically valid measures of academic performance and classroom behavior in adolescents with ADHD. *Experimental and Clinical Psychopharmacology, 9,* 163–175.

Fabiano, G. A., Pelham, W. E., Majumdar, A., Evans, S. W., Manos, M. J., Caserta, D., et al. (2013a). Elementary and middle school teacher perceptions of attention-deficit/hyperactivity disorder incidence. *Child & Youth Care Forum, 42,* 87–99.

Fabiano, G. A., Pelham, W. E., Jr., Waschbusch, D. A., Gnagy, E. M., Lahey, B. B., Chronis, A. M., et al. (2006). A practical measure of impairment: Psychometric properties of the impairment rating scale in samples of children with attention deficit hyperactivity disorder and two school-based samples. *Journal of Clinical Child and Adolescent Psychology, 35*(3), 369–385.

Fabiano, G. A., Vujnovic, R. K., Pelham, W. E., Waschbusch, D. A., Massetti, G. M., Pariseau, M. E., et al. (2010). Enhancing the effectiveness of special educational programming for children with attention deficit hyperactivity disorder using a daily report card. *School Psychology Review, 39,* 219–239.

Fabiano, G. A., Vujnovic, R. K., Waschbusch, D. A., Yu, J., Mashtare, T., Pariseau, M. E., et al. (2013b). A comparison of workshop training versus intensive, experiential training for improving behavior support skills in early educators. *Early Childhood Research Quarterly, 28,* 450–460.

Fischer, A. J., Dart, E. H., LeBlanc, H., Hartman, K., Steeves, R. O., & Gresham, F. M. (2016). An investigation of the acceptability of videoconferencing within a school-based behavioral consultation framework. *Psychology in the Schools, 53,* 240–252.

Frank, J. L., & Kratochwill, T. R. (2014). School-based problem-solving consultation. In W. P. Erchul & S. M. Sheridan (Eds.), *Handbook of research in school consultation* (2nd ed., pp. 18–39). New York, NY: Routledge.

Fristad, M. A., Teare, M., Weller, E. B., Weller, R. A., & Salmon, P. (1998). Study III: Development and concurrent validity of the Children's Interview for Psychiatric Syndromes-Parent version (P-ChIPS). *Journal of Child and Adolescent Psychopharmacology, 8,* 221–226.

Gilbertson, D., Witt, J. C., Singletary, L. L., & VanDerHeyden, A. (2007). Supporting teacher use of interventions: Effects of response dependent performance feedback on teacher implementation of a math intervention. *Journal of Behavioral Education, 16*(4), 311–326.

Girio, E. L., & Owens, J. S. (2009). Teacher acceptability of evidence-based and promising treatments for children with attention-deficit/hyperactivity disorder. *School Mental Health, 1,* 16–25.

Greene, R., Beszterczey, S. K., Katzenstein, T., Park, L., & Goring, J. (2002). Are students with ADHD more stressful to teach? Patterns of teacher stress in an elementary school sample. *Journal of Emotional and Behavioral Disorders, 10,* 79–89.

Gresham, F. M., & Lopez, M. F. (1996). Social validation: A unifying concept for school-based consultation research and practice. *School Psychology Quarterly, 11*(3), 204.

Hagermoser Sanetti, L. M., & Kratochwill, T. R. (2009). Toward developing a science of treatment integrity: Introduction to the special series. *School Psychology Review, 38,* 445–459.

Han, S. S., & Weiss, B. (2005). Sustainability of teacher implementation of school-based mental health programs. *Journal of Abnormal Child Psychology, 33,* 655–679.

Harrison, J. R., State, T. M., Evans, S. W., & Schamberg, T. (2016). Construct and predictive validity of social acceptability: Scores from high school teacher ratings on the school intervention rating form. *Journal of Positive Behavior Interventions, 18*(2), 111–123.

Holdaway, A. S., Hustus, C. L., Owens, J. S., Evans, S. W., Coles, E. K., Himawan, L., Zoromski, A. K., Dawson, A. E., & Mixon, C. S. (2020). Incremental benefits of a daily report card over time for youth with disruptive behavior: Replication and Extension. *School Mental Health.*

Hoza, B. (2007). Peer functioning in children with ADHD. *Academic Pediatrics, 7*(1), 101–106.

Johnson, L. D., Wehby, J. H., Symons, F. J., Moore, T. C., Maggin, D. M., & Sutherland, K. S. (2014). An analysis of preference relative to teacher implementation of intervention. *Journal of Special Education, 48,* 214–224.

Jones, H. A., & Chronis-Tuscano, A. (2008). Efficacy of teacher in-service training for attention-deficit/hyperactivity disorder. *Psychology in the Schools, 45*(10), 918–929.

Kazdin, A. E. (1981). Acceptability of child treatment techniques: The influence of treatment efficacy and adverse side effects. *Behavior Therapy, 12,* 493–506.

Loe, I. M., & Feldman, H. M. (2007). Academic and educational outcomes of children with ADHD. *Journal of Pediatric Psychology, 32,* 643–654.

Martens, B. K., Witt, J. C., Elliott, S. N., & Darveaux, D. (1985). Teacher judgments concerning the acceptability of school-based interventions. *Professional Psychology: Research and Practice, 46,* 191–198.

Merikangas, K. R., He, J., Brody, D., Fisher, P. W., Bourdon, K., & Koretz, D. S. (2010). Prevalence and treatment of mental health disorders among U.S. children in 2001–2004 NHANES. *Pediatrics, 125,* 75–81.

Murray, D. W., Rabiner, D., Schulte, A., & Newitt, K. (2008). Feasibility and integrity of a parent–teacher consultation intervention for ADHD students. *Child & Youth Care Forum, 37,* 111–126.

Nastasi, B. K., & Truscott, S. D. (2000). Acceptability research in school psychology: Current trends and future directions. *School Psychology Quarterly, 15*(2), 117.

Noell, G. H., Gresham, F. M., & Gansle, K. A. (2002). Does treatment integrity matter? A preliminary investigation of instructional implementation and mathematics performance. *Journal of Behavioral Education, 11,* 51–67.

Noell, G., Witt, J., Gilbertson, D., Ranier, D., & Freeland, J. (1997). Increasing teacher intervention implementation in general education settings through consultation and performance feedback. *School Psychology Quarterly, 12,* 77–88.

Owens, J. S., Coles, E. K., & Evans, S. W. (2014, September). *The role of teacher knowledge, skills, and beliefs in the implementation of classroom management skills.* Presentation at the Annual School Mental Health Research Summit. Pittsburgh, PA.

Owens, J. S., Coles, E. K., Evans, S. W., Himawan, L. K., Girio-Herrara, E. L., Holdaway, A. S., et al. (2017). Using multi-component consultation to increase the integrity with which teachers implement behavioral classroom interventions: A pilot study. *School Mental Health, 9,* 218–234.

Owens, J. S., Evans, S. W., Coles, E. K., Himawan, L. K., Holdaway, A. S., et al. (2020). Consultation for classroom management and targeted interventions: Examining benchmarks for teacher practices that produce desired change in student behavior. *Journal of Emotional and Behavioral Disorders, 28*(1), 52–64.

Owens, J. S., Holdaway, A. S., Zoromski, A. K., Evans, S. W., Himawan, L. K., Girio-Herrera, E., et al. (2012). Incremental benefits of a daily report card intervention over time for youth with disruptive behavior. *Behavior Therapy, 43,* 848–861.

Owens, J. S., Murphy, C. E., Richerson, L., Girio, E. L., & Himawan, L. K. (2008). Science to practice in underserved communities: The effectiveness of school mental health programming. *Journal of Clinical Child and Adolescent Psychology, 37,* 434–447.

Pelham, W. E., Gnagy, E. M., Greenslade, K. E., & Milich, R. (1992). Teacher ratings of DSM-III-R symptoms for the disruptive behavior disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 31,* 210–218.

Pelham, W. E., Greiner, A. R., & Gnagy, E. M. (2008). Student behavior teacher response observation code manual (Unpublished manual).

Pintello, D. (2020). Commentary: Establishing scientific rigor and excellence in implementation science training to improve the deployment of evidence-based mental health services. *Administration and Policy in Mental Health and Mental Health Services Research, 47,* 265–271.

Pisecco, S., Huzinec, C., & Curtis, D. (2001). The effect of child characteristics on teachers' acceptability of classroom-based behavioral strategies and psychostimulant medication for the treatment of ADHD. *Journal of Clinical Child Psychology, 30*(3), 413–421.

Power, T. J., Hess, L. E., & Bennett, D. S. (1995). The acceptability of interventions for attention-deficit hyperactivity disorder among elementary and middle school teachers. *Developmental and Behavioral Pediatrics, 16,* 238–243.

Proctor, E. K., Landsverk, J., Aarons, G., Chambers, D., Glisson, C., & Mittman, B. (2009). Implementation research in mental health services: An emerging science with conceptual, methodological, and training challenges. *Administration and Policy in Mental Health and Mental Health Services Research, 36*(1), 24–34.

Pyle, K., & Fabiano, G. A. (2017). Daily report card intervention and attention deficit hyperactivity disorder: A meta-analysis of single-case studies. *Exceptional Children, 83*(4), 378–395.

Reimers, T. M., Wacker, D. P., & Koeppl, G. (1987). Acceptability of behavioral interventions: A review of the literature. *School Psychology Review, 16,* 212–227.

Robb, J. A., Sibley, M. H., Pelham, W. E., Foster, E. M., Molina, B. S., Gnagy, E. M., et al. (2011). The estimated cost of ADHD to the US education system. *School Mental Health, 3*(3), 169–177.

Vannest, K. J., Davis, J. L., Davis, C. R., Mason, B. A., & Burke, M. D. (2010). Effective intervention for behavior with a daily behavior report card: A meta-analysis. *School Psychology Review, 39*(4), 654–672.

Vereb, R. L., & DiPerna, J. C. (2004). Teachers' knowledge of ADHD, treatments for ADHD, and treatment acceptability: An initial investigation. *School Psychology Review, 33*(3), 421.

Volpe, R. J., DuPaul, G. J., DiPerna, J. C., Jitendra, A. K., Lutz, G., Tresco, K., et al. (2006). Attention deficit hyperactivity disorder and scholastic achievement: A model of mediation via academic enablers. *School Psychology Review, 35,* 47–61.

Vujnovic, R. K., Holdaway, A. S., Owens, J. S., & Fabiano, G. A. (2014). Response to intervention for youth with attention-deficit/hyperactivity disorder: Incorporating an evidence-based intervention within a multi-tiered framework. In Weist, M. D., Lever, N. A., Bradshaw, C. P., & Sarno Owens, J. (Eds.), *Issues in clinical child psychology. Handbook of school mental health: Research, training, practice, and policy* (pp. 399–411).

Wechsler, D. (2011). *Wechsler abbreviated scale of intelligence-second edition (WASI-II).* San Antonio, TX: NCS Pearson.

Williams, K., Noell, G., Jones, B. A., & Gansle, K. A. (2012). Modifying students' classroom behaviors using an electronic daily behavior report card. *Child and Family Therapy, 34,* 269–289.

Witt, J. C., & Elliott, S. N. (1985). Acceptability of classroom intervention strategies. In T. R. Kratochwill (Ed.), *Advances in school psychology* (Vol. 4, pp. 251–288). Hillsdale, NJ: Erlbaum.

Witt, J. C., Moe, G., Gutkin, T. B., & Andrews, L. (1984). The effects of saying the same thing in different ways: The problem of language and jargon in school-based consultation. *Journal of School Psychology, 22,* 361–367.

Wolraich, M. L., McKeown, R. E., Visser, S. N., Bard, D., Cuffe, S., Neas, B., et al. (2014). The prevalence of ADHD: its diagnosis and treatment in four school districts across two states. *Journal of Attention Disorders, 18,* 563–575.