

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
1–14

© Hammill Institute on Disabilities 2018

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/1063426618795440

jebd.sagepub.com



Julie Sarno Owens, PhD¹, Steven W. Evans, PhD¹, Erika K. Coles, PhD², Alex S. Holdaway, MS¹, Lina K. Himawan, MS¹, Clifton S. Mixon, MS¹, and Theresa E. Egan, MA¹

Abstract

In the context of best practices consultation with high needs teachers, we examined (a) relations between teachers' appropriate response to student rule violations and rates of rule violations, and (b) rates of student misbehavior among teachers who do and do not achieve various benchmarks of integrity and/or growth in skills. Participants were 48 teachers, 48 target students with or at risk for attention deficit hyperactivity disorder (ADHD; one per teacher), and remaining students in each classroom. Teachers received up to eight consultation sessions on classroom management and implementation of a daily report card (DRC) with the target student. We observed classwide rule violations, target student rule violations, and DRC violations, as well as the percentage of rule violations to which the teacher provided an appropriate response. Teachers who responded to a higher percentage of rule violations had fewer classwide rule violations ($r_s = -.32$ to $-.53$) and target student rule violations ($r_s = -.22$ to $-.51$) at baseline, Months 1 to 2, and Months 3 to 4 of consultation. Teachers who reached the minimum benchmark of 51% appropriate response and who demonstrated greater growth in appropriate responding witnessed fewer rule violations than teachers who did not achieve these benchmarks. Implications for preservice training, professional development, and consultation are discussed.

Keywords

consultation, integrity, daily report card, classroom management, benchmark

About 10% to 20% of children exhibit disruptive behaviors that are difficult for teachers to manage (Fabiano et al., 2013; Visser et al., 2014). If not addressed, disruptive student behavior detracts from instruction time (Robb et al., 2011) and contributes to teacher stress and occupational attrition (Greene, Beszterczey, Katzenstein, Park, & Goring, 2002; Ingersoll, 2001). When teachers have difficulty managing challenging behaviors, they often seek assistance from an experienced peer colleague, school psychologist, or building-level problem-solving team. Problem-solving consultation (Frank & Kratochwill, 2014), coupled with observation and performance feedback (Solomon, Klein, & Politylo, 2012) from a trained school professional, represents current best practices for facilitating teachers' use of universal or targeted classroom management strategies that address challenging student behavior.

Despite the large literature on consultation and coaching (see Stormont, Reinke, Newcomer, Marchese, & Lewis,

2015, for review), there are few studies that document the magnitude of growth in teacher behavior or the level of intervention integrity needed to produce change in student behavior. Our knowledge of this relationship is limited for two reasons. First, most teacher consultation studies have reported impacts on proximal outcomes (i.e., teacher knowledge, skills, or efficacy), with fewer reporting distal outcomes like student behavior or achievement (see Pas, Bradshaw, & Cash, 2014, for review). Second, most teacher consultation studies that report student outcomes are single-case designs (e.g., DiGennaro, Martens, &

¹Ohio University, Athens, USA

²Florida International University, Miami, USA

Corresponding Author:

Julie Sarno Owens, Ohio University, Porter Hall 200, Athens, OH 45701, USA.

Email: owensj@ohio.edu

Kleinmann, 2007; Reinke, Lewis-Palmer, & Merrell, 2008). These studies allow us to draw conclusions about the functional relationship between change in teacher behavior and change in student behavior; however, they are limited in sample size and scope (e.g., restricted to a few weeks). Both of these factors limit the extent to which the findings generalize to typical classroom conditions.

Unless the consultation-related change in teacher behavior is linked to change in student behavior, even the most effective consultation may represent a squandering of expensive resources. In addition, school mental health professionals (SMHPs) have many competing demands for their attention, leaving limited time for consultation to teachers. The identification of minimum benchmarks for integrity could offer guidelines to determine which teachers may benefit most from consultation (e.g., high need teachers). Furthermore, a foundational tenet of multitiered systems of support is that the level of intervention intensity can be reduced or intensified based on the student's response to a given level of intervention. If we can develop benchmarks for integrity for universal and targeted classroom intervention, SMHPs could have greater confidence in knowing whether a lack of response to an intervention is a function of the student or inadequate intervention implementation. Finally, benchmarks could offer guidelines for training teachers in classroom management practices and policies related to teacher evaluation.

The aims of this study were to (1) examine the relations between teachers' appropriate response to rule violations and rates of student rule violations (classwide and by a target student) in the context of best practices consultation over 4 months and (2) explore rates of student rule violations (classwide and by a target student) among teachers who do and do not achieve (and maintain) various benchmarks of integrity or growth in the skill of responding to rule violations.

Intervention Integrity in Consultation Research

Intervention integrity refers to the consultee's implementation of a given strategy (universal or targeted) in a natural setting following consultation (Wilkinson, 2007). This is considered not only a proximal outcome of consultation but also the process through which distal outcomes (e.g., change in student behavior) are achieved. However, little is known about the magnitude of growth in teacher behavior or the level of integrity needed to produce change in student behavior. Although a functional relationship between intervention integrity and student outcomes following consultation has been documented in some studies, Noell and Gansle (2014) in their review of studies examining this relationship conclude "that the literature is too diverse and fractured at present to yield broad conclusions" (p. 400).

Thus, additional study of the relationship with specific attention to benchmarks for integrity and growth in skills is warranted.

Connections Between Change in Teacher Behavior and Change in Student Behavior

Several single-case, multiple-baseline design studies demonstrate a functional relationship between change in teacher behavior (e.g., intervention integrity) and change in observed student behavior. Sanetti, Collier-Meek, Long, Kim, and Kratochwill (2014) found a connection between teachers' implementation of a behavior support plan and change in observed student disruptive behavior and academic engagement for two of three cases in the context of a consultation program that included implementation planning as a strategy to promote integrity. Similarly, Reinke et al. (2008) found that consultation that included motivational enhancement strategies and visual performance feedback increased teachers' use of praise and simultaneously decreased observed student disruptions for three of four student-teacher dyads. Furthermore, in a multiple-baseline design study, DiGennaro et al. (2007) examined the correlations between teacher integrity and target student behavior. They found significant correlations for three of four cases (r s ranged from $-.45$ to $-.78$) indicating higher integrity was associated with lower frequency of problem behavior. Finally, in a case study design, we found (Coles et al., 2015) a visual relationship between specific strategies used in consultation (skills practice, beliefs modification strategy), observed teacher integrity (increase in labeled praise, appropriate response to student rule violations), and observed student outcomes (decrease in off-task rule violations and noncompliance, increase in work completion.).

These studies provide promising evidence that integrity is associated with change in student behavior; however, the finding may not be representative of the duration of strategy implementation often required over the course of the school year. Furthermore, these studies did not test minimum benchmarks of integrity needed to produce change in student behavior. Indeed, in some studies, a small proportion of students did not respond to the intervention, possibly suggesting that the level of intervention integrity was not sufficient.

There are several group design studies examining the impact of consultation or coaching on teacher behavior (e.g., Becker, Bradshaw, Domitrovich, & Ialongo, 2013; Bradshaw, Pas, Goldweber, Rosenberg, & Leaf, 2012; Motoca et al., 2014); however, there are very few group design consultation studies that examine change in observed teacher practices and change in observed student behavior (Reinke et al., 2014). Conroy and colleagues'

studies provide the most relevant evidence regarding this relationship. Across two studies, Conroy and colleagues (Conroy et al., 2015; Conroy, Sutherland, Vo, Carr, & Ogston, 2014) examined the impact of coaching with performance feedback over 14 weeks on teachers' use of effective universal classroom management strategies and on concomitant change in student behavior. In both studies, they found that, relative to the baseline phase, teachers improved meaningfully in the use of most strategies following the coaching phase. They also observed concomitant changes in observed student behaviors (increases in engagement and decreases in defiance, aggression, and disruptive behavior) that were not observed in the control condition (Conroy et al., 2015).

There are noteworthy patterns in their data. First, growth in teacher behavior was significant (e.g., use of praise and corrective feedback increased from 1% to 7% of observed intervals), yet there remained room for growth in the use of the strategies (Conroy et al., 2015). Second, neither these authors nor others (e.g., Reinke et al., 2014) tested for minimum benchmarks of integrity. Third, Conroy et al. (2015) did not consider variability in the teacher sample to determine whether the magnitude of teacher growth was related to the magnitude of improvements in student behavior. Both benchmarks and growth may be important, particularly among the teachers with the lowest levels of integrity at baseline (e.g., high need teachers).

Variability in Teacher Profiles

There is emerging evidence of variability in teacher's knowledge, beliefs, and skills prior to consultation (Owens, Coles et al., 2017; Owens, Holdaway et al., 2017; Reddy, Fabiano, Dudek, & Hsu, 2013), receptivity to consultation (Owens, Schwartz et al., 2017), and intervention implementation (Domitrovich et al., 2015; Owens, Coles et al., 2017). For example, we (Owens, Holdaway et al., 2017) examined teachers' use of praise and appropriate response to rule violations prior to engaging in consultation. Teachers' rates of praise per hour ranged from 11 to 38 and the percentage of rule violations to which teachers responded appropriately ranged from 27% to 47% for classwide rule violations and 11% to 31% for target student rule violations (a student with or at risk for attention deficit hyperactivity disorder [ADHD]). Theoretically, this variability can be leveraged to determine whether there are minimum benchmarks of teacher behaviors that produce desired levels of student behavior, as well as which teachers may have the greatest need for consultation.

Similarly, there is variability in the extent to which teachers respond to coaching (e.g., Becker, Darney, Domitrovich, Keperling, & Ialongo, 2013; Cappella et al., 2012; Owens, Coles et al., 2017). We (2017a) evaluated the effectiveness of an individually tailored, multicomponent consultation package (designed to address barriers to integrity) relative to

a comparison condition designed to represent best practices (problem solving with performance feedback) with elementary school teachers. Teachers in both conditions showed significant improvements in universal and targeted classroom management strategies. However, the group of high need teachers (i.e., those with lower baseline levels of knowledge, skills, and intervention-supportive beliefs) demonstrated more improvement in response to the multicomponent consultation than in response to the comparison consultation (Cohen's d ranged from 0.33–1.12). Similar to the findings of Conroy et al., despite demonstrating meaningful growth in strategy use, there was ample room for improvement after consultation. On average, high need teachers who received the multicomponent consultation improved from responding appropriately to 14% of target student rule and 28% of classwide rule violations at baseline, to responding to 40% and 68% of rule violations, respectively, at the end of consultation, with wide variability across teachers.

These findings show that individually tailored consultation that directly addresses barriers to integrity can result in improved teacher outcomes relative to current best practices. However, even individually tailored consultation results in variable teacher change over time, with many teachers showing room for continued growth after consultation. Thus, perhaps aiming for 100% integrity (or even 80%) is unrealistic, yet also unnecessary. Additional research is needed to identify student outcomes associated with various benchmarks of integrity and various degrees of growth in teacher skills.

Possible Benchmarks for Integrity

Three studies provide evidence that a modest benchmark of integrity (i.e., between 51% and 66%) may be sufficient to achieve change in student behavior. First, using a multiple-baseline design, Noell, Gresham, and Gansle (2002) examined the impact of three levels of intervention integrity (prompts were provided for 100%, 66%, or 33% of problems) to remind a student to use a strategy. The pattern of results indicated that there was a clear distinction in student outcomes (digits correct) between baseline levels (no prompts) and the 100% condition, as well as between baseline levels (no prompts) and the 66% condition. Furthermore, the 100% condition produced better outcomes than the 33% condition. However, the distinction between 66% and 100% was negligible and described as idiosyncratic. Thus, perhaps a benchmark lower than 100% is adequate for changing student behavior.

Second, in our (2017b) study, we observed the percentage of student rule violations (classwide) to which teachers provided an appropriate response. Teachers were sorted into a variety of groups based on their percentage of appropriate responding to student rule violations, and

student disruptive behavior was compared across groups. When the percentage of rule violations to which the teacher responded appropriately was less than 30%, rule violations were high (70 or 80 per hr); however, when a benchmark of 51% appropriate response to rule violations was reached, rule violations dropped to about 35 per hr. Furthermore, there was little incremental benefit at higher levels of appropriate responding.

Third, Sanetti et al. (2014) found that their consultation procedures improved two of three teachers' adherence rates (i.e., percentage of intervention steps implemented) from below 50% at baseline to at least 80%, and this change was associated with concomitant improvement in student behavior. Yet, the third teacher only improved adherence from 44% to 55% and her student showed minimal change in observed outcomes. Together, these data suggest that there may be a minimum (e.g., 51% or 55%) needed to produce a change in student behavior.

Several benchmarks could be examined, as the previous studies suggest 51%, 55%, and 66% may be possible minimum benchmarks (Owens, Holdaway et al., 2017; Noell et al., 2002; Sanetti et al., 2014). However, we argue that examination should begin at 51% integrity. Namely, as long as teachers are applying a strategy "more often than not," they are creating predictability in the classroom and following through on their expectations. Furthermore, we are not aware of any study that has tested a minimum benchmark; thus, examining the lowest possible, theoretically defensible minimum seems prudent. Clearly, it is also worth examining if there are incremental benefits with higher levels. We also hypothesize that the minimum benchmark of integrity required to modify the behavior of students with severe behavior problems may be higher than 51% (as seen in the Noell et al. and Sanetti et al. studies). These students are more likely to demonstrate unpleasant responses in reaction to teacher demands than typical students (Carr, Taylor, & Robinson, 1991). The principles of negative reinforcement predict that these unpleasant responses lead teachers to withdraw from using strategies, thereby negatively reinforcing both student and teacher behavior. Thus, for these students, teachers may need to demonstrate even greater consistency than what is needed for typical students, to extinguish this negative student response.

Possible Benchmarks for Growth in Teacher Skills

In addition to a minimum benchmark of integrity, there may also be a minimum benchmark for *percent growth* in teacher strategy use for achieving improved student behavior. For example, Conroy et al. observed that an average of 6% growth in teachers' use of praise (from 1%–7%) was associated with concomitant change in student behavior. However,

because we are not aware of any other study that has examined percent growth in teacher skills in relation to student behavior and because percent growth in a given strategy may vary by strategy (e.g., rates of praise, response to rule violations), we do not make hypotheses about benchmarks for growth.

Current Study

In our previous study (Owens, Coles et al., 2017), we reported the teacher outcomes of a randomized trial comparing an individually tailored multicomponent consultation condition designed to address possible barriers to implementation (i.e., low knowledge, skills, and/or intervention-supportive beliefs) with a consultation condition designed to mirror best practices. Teachers in both conditions received an equal dose of consultation (up to eight biweekly sessions) focused on universal classroom management strategies and implementation of a targeted daily report card (DRC; Owens et al., 2012; Vannest, Davis, Davis, Mason, & Burke, 2010) with one student demonstrating significant disruptive behavior. Because teachers in both conditions showed significant improvements in classroom management skills, we combined teachers in our current analyses. In addition, our previous analyses identified a group of teachers with low levels of knowledge, intervention-supportive beliefs, and/or skills at baseline (i.e., high need teachers; see details in "Method" section). In the current study, we conducted our analyses on this sample of teachers, as they are likely most in need of consultation, have the most room for growth, and allow us to examine the benchmark of 51% integrity.

In the current study, we used the above-described data set to examine (Aim 1) relations between teachers' appropriate response to student rule violations and rates of rule violations, and to examine (Aim 2) rates of student misbehavior among teachers who do and do not achieve and/or maintain 51% integrity, and who demonstrate various levels of growth in this strategy. We examine integrity benchmarks and growth in strategy use for both classwide and targeted strategies. We expected to find that higher percentages of appropriate teacher response to rule violations would be associated with lower rates of student rule violations. We also hypothesized that teachers who reached a minimum benchmark (i.e., an average of appropriately responding to 51% of rule violations) during the first half of consultation would observe less disruptive behavior (in the class and the target student) than teachers who did not achieve this benchmark. Furthermore, we expected that teachers who maintained this benchmark during the first and second half of the consultation would observe the lowest levels of disruptive student behavior. Although we examined student outcomes at other benchmarks, we did not make specific hypotheses about these other benchmarks. Similarly, we examined

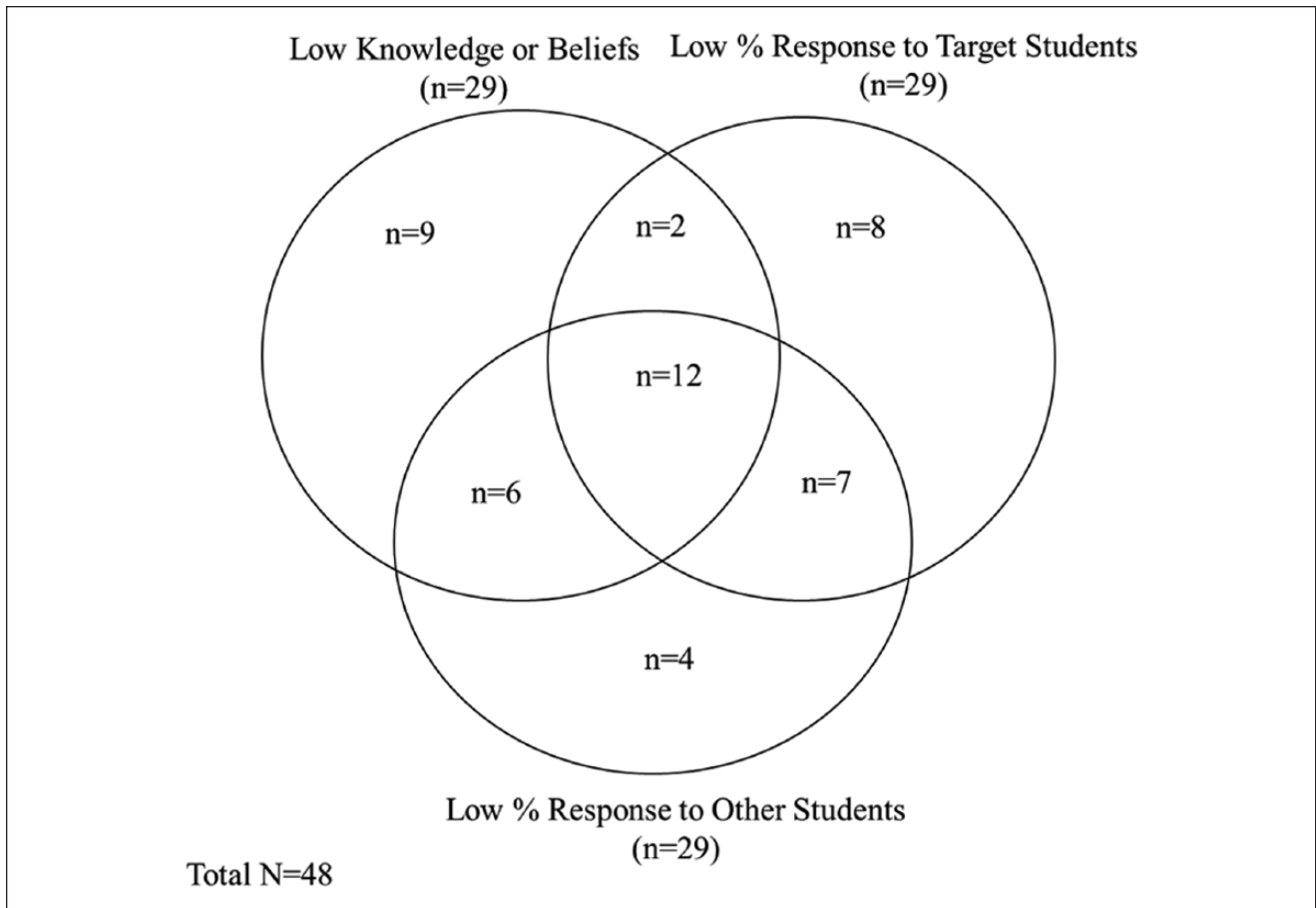


Figure 1. Venn diagram representing the distribution of sample with knowledge, belief, and skills barriers at baseline.

various rates of growth in skills, but given the lack of literature, we did not make specific hypotheses.

Method

Participants

Participants were 48 general education teachers (K–5 grade; 21 from Ohio, 27 from Florida) from our previous trial (Owens, Coles, et al., 2017) who presented with either (a) low baseline levels of knowledge and intervention-supportive beliefs (defined below), or (b) low classroom management skills at baseline with target student, or (c) low classroom management skills at baseline with whole class (defined below; see Figure 1 for distribution of sample across these criteria). These teachers had an average of 14.33 years ($SD = 8.52$) of teaching experience. Most (63%) had obtained a master's degree. Teachers identified as non-Hispanic White (45.8%) and Hispanic (any race; 50%), women (93.8%). The five Ohio schools had an average of 377 students and 16 general education teachers per school, with 12% to 29% of students receiving special education services

and 35% to 75% receiving free or reduced lunch services. The three Florida schools had an average of 1,024 students and 50 general education teachers, with 4% to 11% receiving special education services and 76% to 95% receiving free or reduced lunch services. Class size across sites ranged from 19 to 25 and the teacher was the sole educator in the room. Consultants were postdoctoral fellows ($n = 2$), master's level clinicians ($n = 2$), or graduate students in psychology ($n = 5$). Six identified as Caucasian, one identified as African American, and two identified as Hispanic.

Target students were 48 elementary school students (77.1% male) referred because they demonstrated inattentive or disruptive behavior. Students identified as Hispanic, any race (60%), or non-Hispanic White (39%) and non-Hispanic Black (1%). Most (91.6%) met criteria for ADHD (66.7% combined presentation; 20.8% inattentive presentation; 4.2% hyperactive/impulsive presentation) and 8.3% were at risk for ADHD (elevated symptoms plus impairment). The sample had an average IQ estimate of 98.57 ($SD = 13.01$), as assessed by the *Wechsler Abbreviated Scales of Intelligence, Second Edition* (WASI-II; Wechsler, 2011). The socioeconomic status of their families was low to

middle class (15.9% had a household income below \$15,000, 54.2% had an income between \$15,000 and \$49,999; 22.9% were above \$50,000; 8.3% did not report income). Per parent report, 10.4% of the students had been diagnosed with a learning disability, 25% had a prescription for a psychiatric medication, and 18.8% had repeated a grade.

As expected, teachers in the current sample differed from teachers excluded from the current sample on baseline levels of skills, knowledge, and beliefs (all $ps < .05$). Teachers in the excluded sample, on average, completed about one more consultation session ($M = 7.30$, $SD = 1.25$) than teachers in the included sample ($M = 5.98$, $SD = 2.37$) and were more likely to be kindergarten or second grade teachers. Teachers in the two samples did not differ on number of years in the profession, gender, or ethnicity, and their target students did not differ on gender, ethnicity, severity of ADHD symptoms or impairment, or on IQ estimates (all $ps > .05$).

Measures

Measures to identify high needs teachers. To test our hypotheses, it was important to have a sample of teachers in need of consultation and with ample room for growth. Thus, we only included teachers from our previous study who were in the latent class comprised of teachers with low baseline levels of knowledge and intervention-supportive beliefs (see Owens, Coles et al., 2017, for detail), and teacher who demonstrated a low percentage of appropriate response to student rule violations (i.e., below the sample median), as evidenced by our baseline observations. The measures used to assess knowledge, beliefs, and observed teacher are described below.

Tests of teacher knowledge and beliefs. We assessed teachers' knowledge of ADHD (prevalence, etiology, treatment) using a 24-item true/false/don't know test, inspired by Jones and Chronis-Tuscano (2008). We assessed teachers' knowledge of behavioral principles using a 16-item multiple-choice test, inspired by the Behavior Modification Test (Kratowill, Elliott, & Busse, 1995). For both measures, total percent correct was calculated. Both measures demonstrate sensitivity to change as a function of participating in a workshop focused on ADHD and classroom management (Owens, Coles, & Evans, 2014). We assessed teacher beliefs using the 25-item *Teacher Locus of Control* measure, which assesses teachers' perceptions of personal control and responsibility for student academic and behavioral outcomes. The measure is reliable (Kuder-Richardson formula 20 [KR20] reliability scores were .81 for failure and .71 for success in a previous study), and scores are predictive of teachers' use of techniques learned during an in-service training (Rose & Medway, 1981). We viewed higher scores (i.e., internal locus of control) to be associated with intervention-supportive beliefs. The KR20 scores with this sample were .71 for failure and .46

for success. Given the low score for success (and that removing items did not improve the reliability scores), this subscale was not used. In our previous study (Owens, Coles et al., 2017), the above scores were subjected to a latent class analyses which resulted in two classes; teachers in the high need class were less knowledgeable and took less credit for their students' failure than teachers in the low need class (effect sizes [d] between classes for the above measures ranged from 1.55–2.89; Owens, Coles et al., 2017).

Observations of student rule violations and teachers' response to rule violations. Student rule violations (classwide and by the target student) and teachers' response to these rule violations were the primary variables of interest for the study and were obtained via a modified version of the Student Behavior-Teacher Response Observation Rating System (SBTR; Pelham, Greiner, & Gnagy, 2008). Data from the baseline observations (two to four per teacher) were used to identify high need teachers; namely, any teacher who fell below the sample median on percent appropriate response to classwide or target student rule violations was included (see Figure 1).

The SBTR has adequate interrater reliability, convergent validity, and sensitivity to change when used in elementary classrooms (Fabiano et al., 2010; Owens, Holdaway et al., 2017). Using this system, observers obtained frequency counts of (a) classroom rule violations (RVs) by the target student, (b) violations of DRC target behaviors (DRCRVs) by the target student, (c) classroom RVs by all other students, and (d) teacher's appropriate responses to each RV (ARRV). The observation manual includes definitions for the violation of seven common classroom rules (i.e., be respectful, obey adults, work quietly, remain in seat, raise hand to speak, use materials appropriately, stay on task), and for coding how the teacher responded to each rule violation (i.e., appropriately, inappropriately, or no response). In the manual, an appropriate response is defined as any verbal or nonverbal action that follows a rule violation to provide a response to the behavior. Appropriate responses contain appropriate content and are delivered with appropriate affect, with a neutral tone of voice of normal pitch and intensity, and without including any behavior included in the Inappropriate Response definition (i.e., verbal or nonverbal behavior that is antagonistic, accompanied by excessive or inappropriate gestures, or delivered with inappropriate affect or an inappropriate tone of voice). Prior to conducting observations, teachers were informed of the rules to be coded. Although we acknowledge there are differences in rules posted and enforced in elementary classrooms, all seven rules were coded for all teachers to maintain consistency in the data across teachers.

Consultants and research assistants (unaware of teacher condition) were trained to reliability on the SBTR (see details in Owens, Coles et al., 2017). Interobserver

assessments were conducted for 24% of all observations in the trial. Across all frequency variables, the intraclass correlations (ICC) of Type 1 for average of k raters (ICC(1,k)) as outlined in Shrout and Fleiss (1979) ranged from .78 to .98 with an average of .90. Observation durations ranged from 15 to 45 min. To standardize the variables across observations, frequency counts of target student classroom RVs, target student DRCRVs, and other student RVs were transformed into rates per hour, and teacher responses are presented as percentage of ARRVs for the target student RVs, DRCRVs, and other student RVs. To maximize our ability to present overall trends, we also created averages of the above variables to represent the baseline period, observations that occurred during the first half of consultation (Months 1–2), and observations that occurred during the second half of consultation (Months 3–4).

Study Procedures

See Owens, Coles et al., (2017) for a complete description of procedures. All general education teachers in each elementary school were invited to a 3-hr workshop conducted by the investigators that focused on best practices in general classroom management strategies and the DRC. At this time, teachers completed a battery of questionnaires, including those used to determine their status in the current study.

The larger study from which the data were drawn was focused on students with or at risk for ADHD. Thus, teachers interested in participating in consultation were required to identify one student with or at risk for 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) had an IQ estimate that fell in or above the 90% confidence interval for a score of 80, and (c) met diagnostic criteria for *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; *DSM-IV*; American Psychiatric Association, 1994) ADHD or were at-risk for ADHD. ADHD was defined as the presence of six or more 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 impairment in the school setting 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 the chronicity of symptoms. At-risk status was defined as four or more symptoms and impairment in school. Children were excluded if they had a previous diagnosis of an autism spectrum disorder, bipolar disorder, or intellectual disability per parent report.

Once a target student was identified and teacher consent obtained, at least two baseline classroom observations using the SBTR were conducted, and weekly observations were scheduled for the duration of participation in consultation. Teachers were paid for attending the in-service training and completing questionnaires, but did not receive compensation for participating in consultation sessions or for implementation of any classroom management practices.

Consultation Procedures

In the previous trial, we used stratified random sorting to assign teachers to one of two consultation conditions so that teachers in each condition did not differ in baseline competence ratings (Owens, Coles et al., 2017). Furthermore, teachers in each condition received an equal number of consultation sessions and observations (Owens, Coles et al., 2017). Thus, for the current study, teachers were combined across conditions as the distinction between the conditions was not relevant to the research questions for this study. Consultation in both conditions focused on general classroom management strategies (i.e., labeled praise, use of rules, effective instructions, and appropriate response to RVs) and the use of a DRC intervention. In both conditions, teachers participated in meetings focused on the creation of a DRC. Once the DRC was launched, teachers met with consultants every other week to receive performance feedback from observations, discuss high quality implementation, and problem solve challenges that arose. Sessions ranged from 30 min to 1 hr, and were conducted during, before, or after school.

Results

Missing Data

It is important to note that the sample sizes change across analyses because only teachers with low baseline scores for that given variable are included in a given analyses. For example, the correlation between teachers' ARRV of target students and target students' RVs only includes those teachers who had low baseline scores in either knowledge and beliefs or ARRV for target student RVs (i.e., this correlation does not include the four teachers who were low baseline for other student RVs; see Figure 1). In addition, the sample size declines between baseline and Months 3 to 4 for a variety of reasons: Three teachers withdrew before Month 3, seven students moved out of the teacher's classroom before Month 3, four teachers referred students to the program late in the year so they did not start biweekly sessions until late winter and thus did not receive 3 months of consultation, and in several instances (e.g., applies to four teachers for the target student variables), there were 0 rule violations observed in the Months 3 to 4 time period; thus, the variable of teacher ARRV is not possible and is missing. Because the

Table 1. Correlations Between Rates of Student RVs and Teacher Percent ARRV for Target Student and Other Students by Time.

Student variables	Teacher variables		
	Baseline %ARRV	Months 1–2 %ARRV	Months 3–4 %ARRV
Target student RVs ^a			
Baseline RVs	-.22		
Months 1–2	-.21	-.27 [†]	
Months 3–4	-.27	-.26	-.51**
Target student DRC violations ^b			
Months 1–2	NA	-.09	
Months 3–4	NA	-.22	-.21
Other student RVs ^c			
Baseline RVs	-.34*		
Months 1–2	-.29 [†]	-.32*	
Months 3–4	-.17	-.29	-.53**

Note. RV = rule violation; ARRV = appropriate response to rule violations; DRC = daily report card.

^aFor target student data, $n = 44$ at baseline, $n = 42$ at Months 1–2, and $n = 25$ at Months 3–4.

^bFor DRC data, $n = 36$ at Months 1–2 and $n = 21$ at Months 3–4.

^cFor Other Student data, $n = 40$ at baseline and Months 1–2 and $n = 27$ at Months 3–4.

[†] $p < .09$. * $p < .05$. ** $p < .01$.

reasons for missing data represent typical school practice (i.e., students moving, natural course of teacher referrals) and are not related to the research process, and because the patterns detected at Months 1 to 2 are generally similar to those found at Months 3 to 4 (see below), we view the data at all time points to be valid.

Aim 1: Association Between Teacher Behavior and Student Behavior

We examined correlations between teacher response to rule violations and rates of rule violations using data averaged during the baseline period, Months 1 to 2, and Months 3 to 4 (see Table 1). Within any given time point, these variables are negatively related. For other student RV, the r s range from $-.32$ to $-.53$ (all p s $< .05$), with the strongest relation emerging at Months 3 to 4. For target student RVs, the r s range from $-.22$ to $-.51$, with the strongest relation emerging at Months 3 to 4. For DRCRVs, r s range from $-.09$ to $-.21$ (all nonsignificant).

Aim 2: Student Outcomes Relative to Achievement of Benchmarks and Teacher Growth

Achieving the benchmark: Months 1 to 2. First, we identified the portion of the sample who had achieved an average of 51% ARRV to target student ($n = 8$; 16.6%) and other student RVs ($n = 23$; 47.9%) during Months 1 to 2. On average, those meeting the benchmark for target

students responded appropriately to 78% ($SD = 21\%$) of target student RVs during Months 1 to 2, whereas those not meeting the benchmark responded to 24% ($SD = 17\%$) of target student RVs during Months 1 to 2, $t(40) = 7.65$, $p < .001$. Similarly, on average, those teachers meeting the benchmark for other students responded appropriately to 74% ($SD = 16\%$) of other student RVs during Months 1 to 2, whereas those not meeting the benchmark responded to 28% ($SD = 11\%$) of other student RVs during Months 1 to 2, $t(34) = 10.08$, $p < .001$.

Independent-samples t tests were conducted on the average rate of target student RVs, DRC violations, and other student RVs between teachers who did and did not achieve the 51% benchmark during Months 1 to 2 (see the top half of Table 2). Teachers who met the benchmark experienced about half of the RVs than teachers who did not meet this benchmark. Hedges's g effect sizes are moderate to large for these comparisons (significant p values ranged from .07 to .001).

Maintaining the benchmark. We also grouped teachers based on whether they maintained the 51% benchmark across both Months 1 to 2 and Months 3 to 4 time periods (i.e., Maintained Group) or did not (failed to achieve the benchmark at either time period). Only three teachers (6.25%) maintained the benchmark toward target student RVs at both time points and 12 teachers (25%) maintained the benchmark toward other student RVs at both time points. On average, those maintaining the benchmark toward target students responded appropriately to 89% ($SD = 18\%$) of target student RVs during Months 3 to 4, whereas those not

Table 2. Rates of RVs for Target Students and Other Student by Time and Teacher Benchmark Groups.

Student variables	Teacher variables		
	Teachers below 51% benchmark over Months 1–2	Teachers meeting 51% benchmark over Months 1–2	Between group effect size (<i>g</i>)
Target student RVs			
Months 1–2	6.50 (5.74)	3.87 (3.04)	.41
Months 3–4	4.81 (6.39)	2.06 (1.44)	.46
Target student DRCRVs			
Months 1–2	2.82 (3.69)	0.43 (0.65)**	.71
Months 3–4	2.41 (5.21)	0.40 (0.57)	.42
Other Student RVs			
Months 1–2	47.86 (19.52)	35.93 (20.36) †	.60
Months 3–4	42.68 (33.72)	21.84 (15.32)*	.80

Student variables	Teacher variables		
	Teachers not maintaining a 51% benchmark over Months 1–4	Teachers maintaining 51% benchmark over Months 1–4	Between group effect size (<i>g</i>)
Target student RVs			
Months 3–4	5.32 (6.28)	1.11 (1.11)	.70
Target student DRCRVs			
Months 3–4	2.54 (5.29)	0.40 (0.70)	.42
Other student RVs			
Months 3–4	44.57 (31.30)	19.55 (13.19)*	1.00

Note. The 51% benchmark represents an average of 51% appropriate response for the given type of RV (i.e., target student RV or other student RV). RV = rule violation.

† $p < .09$. * $p < .05$. ** $p < .01$.

maintaining the benchmark responded to 42% ($SD = 34\%$) of target student RVs during Months 3 to 4, $t(23) = 2.32$, $p = .03$. Similarly, on average, those maintaining the benchmark toward other students responded appropriately to 84% ($SD = 18\%$) of other student RVs during Months 3 to 4, whereas those not maintaining the benchmark responded to 47% ($SD = 25\%$) of other student RVs during Months 3 to 4, $t(21) = 3.93$, $p = .001$.

Independent-samples t tests were conducted on the average rate of target student, DRC violations, and other student rule violations among teachers who did and did not achieve each maintenance benchmark (see bottom half of Table 2). Teachers who maintained the benchmark experienced about half of the RVs than teacher who did not maintain this benchmark. Hedges's g effect sizes are moderate to large for these comparisons (the p value for the significant effect, other student rule violations, was .011).

Student outcomes as a function of teacher integrity and growth. In an attempt to replicate the pattern of data provided in Owens, Holdaway et al., (2017), teachers were sorted based on their percent ARR_V to target student (see Figure 2a) and other student RVs (see Figure 2b), and rates of student RVs were examined across the groups. Because there was minimal variability in DR_{CR}RVs, we did not assess this outcome

further. A few patterns are noteworthy. First, the pattern across both figures indicates that student rule violations are highest when teacher percent ARR_V is lowest. Second, the patterns suggest that earlier in the year (i.e., Months 1–2), there may be slight incremental benefit with higher levels of percent ARR_V beyond the 51% benchmark for both types of RVs. For example, at Months 1 to 2, target students of teachers with a 51% ARR_V violated an average of four rules per hour, whereas target students of teachers with 90% or higher ARR_V violated less than one rule per hour (see Figure 2a). Similarly, at Months 1 to 2, the rate of other student RVs per hour among teachers with a 51% ARR_V was an average of 37, and the rate per hour among teachers with 90% or higher ARR_V violated less than 23 (see Figure 2b). However, at Months 3 to 4, this incremental pattern was not observed.

Finally, to try to better understand this pattern, average rates of other student RVs were examined considering both various levels of growth and whether or not teachers achieved the 51% minimum integrity benchmark. The average rates of other student RVs for each of these groups are depicted in Figure 3. As revealed in Table 2, achieving the benchmark of 51% is related to lower rates of rule violations. However, the pattern depicted in Figure 3 reveals that growth within each benchmark group is also related to rule violations. For example, the average rates of rule violations among teachers who

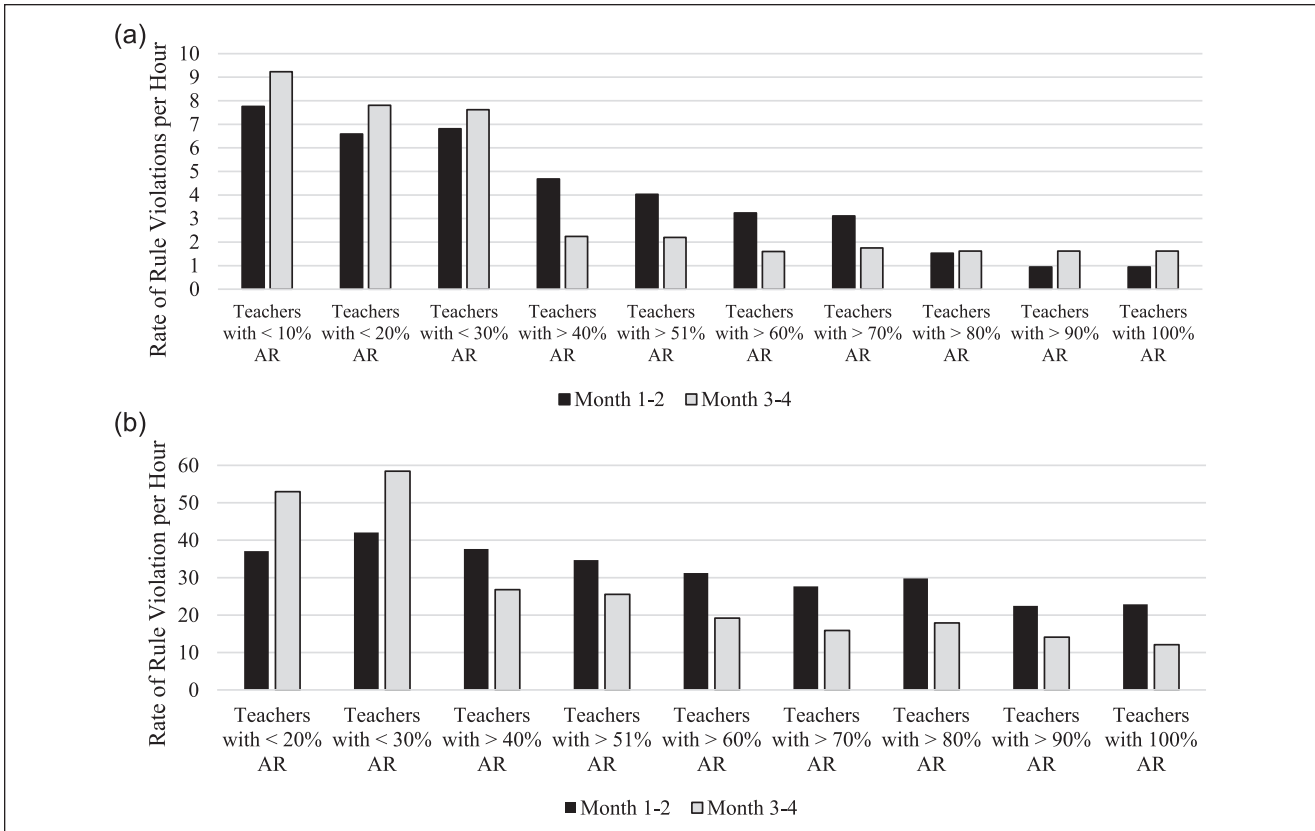


Figure 2. Rate of (a) target student rule violations/hour and (b) other student rule violations/hour as a function of teacher appropriate response (AR) group during Months 1 to 2 and Months 3 to 4. Note. AR = appropriate response.

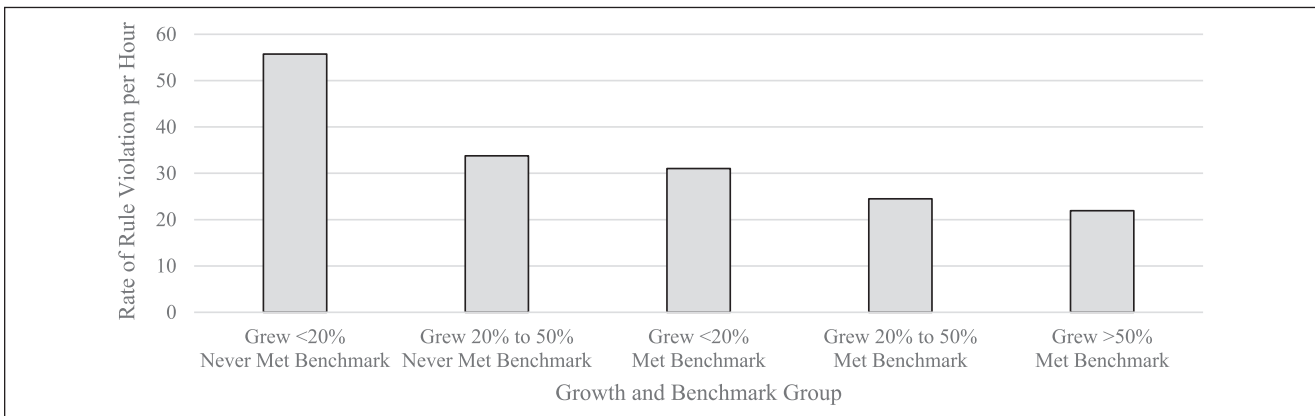


Figure 3. Rates of other student rule violations at Months 3 to 4 based on teacher's growth and attainment of 51% benchmark of appropriate response to rule violations.

improved less than 20% and never crossed the 51% benchmark ($n = 7$) were 55.75, whereas the average rate among teachers who improved 20% to 50% and never cross the 51% benchmark ($n = 5$) was 33.77. Furthermore, among teachers who achieved the 51% benchmark and improved <20%

($n = 4$) had on average 31.04 rule violations per hour. Those who achieved the 51% benchmark and improved >50% ($n = 4$) had 21.92 rule violations per hour. Thus, reductions in rule violations appear related to both growth in ARR and exceeding the minimum benchmark of 51% ARR.

Discussion

This study examined the association between teacher response to rule violations and student behavior in the context of up to 4 months of teacher consultation. This relationship was analyzed in a sample of teachers identified as needing consultation supports, and as a function of integrity benchmarks and teacher growth in skills. The results advance the science of consultation research by documenting levels of student outcomes at various levels of teacher integrity and growth in skills and by revealing a possible minimum benchmark for integrity in teacher practices needed to produce desired change in student behavior. Such data stimulate a host of testable hypotheses that will guide our understanding of how change in teacher practices mediates the relationship between consultation and student outcomes.

First, our study reveals the relationship between teacher behavior and student behavior prior to consultation and at two time points within the consultation process (over Months 1–2 and Months 3–4). As hypothesized, there is a modest negative relationship between teacher behavior and student behavior at baseline ($r_s = -.22$ to $-.34$); however, as teachers enhance their use of effective practices, this relationship becomes stronger ($r_s > .50$ at Months 3–4; see Table 1). These correlations are consistent with previous findings (DiGennaro et al., 2007) and suggest that student learning over time how teachers respond to misbehavior and modify their behavior accordingly. However, this pattern was not observed for target student DRCRVs. This may have occurred because behaviors that are targeted on the DRC likely represent the behaviors that are most impairing to the student and most difficult to change (otherwise a targeted intervention would likely not have been used). Given that there are many other influences on these behaviors beyond that of the teacher (e.g., home environment, neurocognitive development), this specific strategy (i.e., response to rule violation) may have a proportionally smaller impact on these behaviors relative to other target student behaviors and/or to the behavior of other students. Alternatively, this pattern may have been found because DRCRVs were observed at the lowest rate of the three student variables and, as a result, had a more restricted range.

Second, based on patterns detected in previous studies, we hypothesized that achieving a minimal benchmark of 51% integrity could have a meaningful impact on student outcomes. Indeed, the data in Table 2 show that teachers who achieved the benchmark by Months 1 to 2 and/or who maintained the benchmark across both time periods witnessed about half the rule violations (among target students and other students) as compared with teachers who did not achieve or maintain this benchmark (effect sizes were medium to large). Given that behavioral infractions detract from classroom instruction time (Robb et al.,

2011), are stressful for teachers (Greene et al., 2002), and are a top contributor to teacher job dissatisfaction and attrition from the profession (Ingersoll, 2001), reducing violations by half would likely have a substantial impact on learning time for all students, student–teacher relations, and teacher job satisfaction. Furthermore, this finding offers provocative new ideas for future research on teacher training, consultation, and assessing a student’s response to intervention (see “Conclusions and Implications” section below).

Third, the patterns in Figure 2a and 2b document the wide variability in teacher skills throughout consultation. This variability suggests that the field needs to shift from a one-size-fits-all approach to an individually tailored approach to professional development and consultation (Gage, MacSuga-Gage, & Crews, 2017; Owens, Coles et al., 2017), as teachers with different profiles may need different types of support to achieve adequate growth and benchmarks. The patterns offer support to the hypothesis that a benchmark of 51% integrity may be a meaningful and reasonable *minimum* benchmark to achieve; however, the data also suggest that there are benefits of incremental integrity. Rule violations in target students and other students are observed at lower rates when teachers achieve 80% or 90% ARRIV than when teachers achieve 40% or 50% ARRIV, particularly in the earlier phases of consultation. In addition, the pattern in Figure 3 documents the impact that growth (in the context of benchmarks) may have on student outcomes. Collectively, these findings highlight the need to further develop tools for assessing teacher’s willingness to engage in consultation (Owens, Schwartz et al., 2017), as well as methods for tailoring consultation to the variety of teacher needs and response patterns (e.g., Gage et al., 2017; Owens, Coles et al., 2017; Reinke et al., 2008; Sanetti et al., 2014).

Finally, it is notable that nearly 60% of teachers achieved the 51% benchmark for other student ARRIV during Months 1 to 2, yet less than 20% met the 51% benchmark for target student ARRIV during Months 1 to 2. This supports the notion that it is more challenging for teachers to change their behavior in relation to target students than other students and that greater attention may need to be devoted to teacher behaviors directed toward target students.

Limitations

First, the small sample size prevented further statistical analyses and the examination of multiple benchmarks for growth and integrity. Second, we only analyzed one teacher behavior. Clearly, there are multiple important teacher practices that affect student behavior and should be assessed before results can fully inform teacher evaluations systems. Third, observers recorded a standardized set of rule violations common to elementary classrooms that may not have corresponded to the rules posted and enforced in each

individual classroom. Although this decision increases the precision of observations, it may affect generalizability of results across classrooms.

Conclusions and Implications

Because DRC target behaviors represent priority behaviors that are critical to improving a student's functioning, yet were only minimally related to the teacher practice studied here, additional work is needed to understand how to help teacher behaviors differentially impact these impairing behaviors. For example, DRC target behaviors may be more difficult to change than behaviors exhibited by students who are not at-risk for ADHD and may require a high degree of coordinated implementation (i.e., high consistency of teacher implementation and high degree of home-based implementation) to shift student behavior. In addition, that these behaviors were observed at the lowest rate of the three student variables highlights the challenges associated with capturing, via observation, student behaviors associated with targeted interventions and the need for the development of adequate systems for studying such behaviors.

The findings also stimulate new ideas related to teacher training, teacher evaluation systems, and assessing a student's response to intervention. For example, with additional research in teacher training programs and/or evaluation systems, we may be able to identify benchmarks for minimal competencies and for mastery of a given skill, as well as goals and processes for individualized professional development plans. Investigation is also needed to better understand how benchmarks might vary across different teacher skills (e.g., response to violations vs. praise) and based on intensity of student needs. Furthermore, as mentioned previously, SMHPs could begin to assess and use benchmarks to determine where limited consultation resources should be directed. Such benchmarks could be systematically considered prior to intensifying an intervention for a student who seems to be insufficiently responsive. Namely, school teams could systematically assess both benchmarks and student response before making intervention decisions. Finally, these data suggest that further study of benchmarks of integrity defined in various increments is warranted with larger samples to enhance confidence in the findings and further demark minimum benchmarks and meaningful increments to achieve.

This is the one of the first studies to examine possible benchmarks for integrity and/or growth in skills among high needs teacher receiving best practices consultation. The findings provide a foundation for a variety of hypotheses that can be tested related to response to rule violations, as well as other critical teacher behaviors.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant (R324A120272). The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Becker, K. D., Bradshaw, C. P., Domitrovich, C., & Ialongo, N. S. (2013). Coaching teachers to improve implementation of the good behavior game. *Administration and Policy in Mental Health and Mental Health Services Research, 40*, 482–493.
- Becker, K. D., Darney, D., Domitrovich, C., Keperling, J. P., & Ialongo, N. S. (2013). Supporting universal prevention programs: A two-phased coaching model. *Clinical Child and Family Psychology Review, 16*, 213–228.
- Bradshaw, C. P., Pas, E. T., Goldweber, A., Rosenberg, M. S., & Leaf, P. J. (2012). Integrating school-wide positive behavioral interventions and supports with tier 2 coaching to student support teams: The PBISplus model. *Advances in School Mental Health Promotion, 5*, 177–193.
- Cappella, E., Hamre, B. K., Kim, H. Y., Henry, D. B., Frazier, S. L., Atkins, M. S., & Schoenwald, S. K. (2012). Teacher consultation and coaching within mental health practice: Classroom and child effects in urban elementary schools. *Journal of Consulting and Clinical Psychology, 80*, 597–610.
- Carr, E. G., Taylor, J. C., & Robinson, S. (1991). The effects of severe behavior problems in children on the teaching behavior of adults. *Journal of Applied Behavior Analysis, 24*, 523–535.
- Coles, E. K., Owens, J. S., Serrano, V., Slavec, J., & Evans, S. W. (2015). From consultation to student outcomes: The role of teacher knowledge, skills, and beliefs in increasing integrity in classroom behavior management. *School Mental Health, 7*, 34–48. doi: 10.1007/s12310-015-9143-2
- 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.
- Conroy, M. A., Sutherland, K. S., Vo, A. K., Carr, S., & Ogston, P. L. (2014). Early childhood teachers' use of effective instructional practices and the collateral effects on young children's behavior. *Journal of Positive Behavior Interventions, 16*, 81–92.
- DiGennaro, F. D., Martens, B. K., & Kleinmann, A. E. (2007). A comparison of performance feedback procedures on teachers' treatment implementation integrity and students' inappropriate

- behavior in special education classrooms. *Journal of Applied Behavior Analysis*, 40, 447–461.
- Domitrovich, C. E., Pas, E. T., Bradshaw, C. P., Becker, K. D., Keperling, J. P., Embry, D. D., & Jalongo, N. (2015). Individual and school organizational factors that influence implementation of the PAX good behavior game intervention. *Prevention Science*, 16, 1064–1074.
- Fabiano, G. A., Pelham, W. E., Jr., Majumdar, A., Evans, S. W., Manos, M. J., Caserta, D., . . . Carter, R. L. (2013). Elementary and middle school teacher perceptions of attention-deficit/hyperactivity disorder prevalence. *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., . . . Burrows-Maclean, L. (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 & Adolescent Psychology*, 35, 369–385.
- Fabiano, G. A., Vujnovic, R., Pelham, W. E., Waschbusch, D. A., Massetti, G. M., Yu, J., . . . Volker, M. (2010). Enhancing the effectiveness of special education programming for children with ADHD using a daily report card. *School Psychology Review*, 39, 219–239.
- 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.
- Gage, N. A., MacSuga-Gage, A. S., & Crews, E. (2017). Increasing teachers' use of behavior-specific praise using a multitiered system for professional development. *Journal of Positive Behavior Interventions*, 19, 239–251.
- Greene, R., Beszterczey, S. K., Katzenstein, T., Park, K., & 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.
- Ingersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal*, 38, 499–534.
- Jones, H. A., & Chronis-Tuscano, A. (2008). Efficacy of teacher in-service training for attention-deficit/hyperactivity disorder. *Psychology in the Schools*, 45, 918–929.
- Kratochwill, T. R., Elliott, S. N., & Busse, R. T. (1995). Behavior consultation: A five-year evaluation of consultant and client outcomes. *School Psychology Quarterly*, 10, 87–117.
- Motoca, L. M., Farmer, T. W., Hamm, J. V., Byun, S. Y., Lee, D. L., Brooks, D. S., . . . Moohr, M. M. (2014). Directed consultation, the SEALS model, and teachers' classroom management. *Journal of Emotional and Behavioral Disorders*, 22, 119–129.
- Noell, G. H., & Gansle, K. A. (2014). Research examining the relationships between consultation procedures, treatment integrity, and outcomes. In W. P. Erchul & S. M. Sheridan (Eds.), *Handbook of research in school consultation* (2nd ed., pp. 386–408). New York, NY: Routledge.
- 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.
- Owens, J. S., Coles, E. K., & Evans, S. (2014, November). Developing a multi-component consultation package to enhance classroom intervention integrity and student outcomes. In G. DuPaul (Chair) Treatment development for children and adolescents with ADHD and behavior disorders: Promoting engagement, integrity and successful outcomes. Symposium presented at the Annual Convention for the Association for Behavioral and Cognitive Therapies, Philadelphia, PA.
- Owens, J. S., Coles, E. K., Evans, S. W., Himawan, L. K., Girio-Herrera, E., Holdaway, A. S., Zoromski, A., Schamberg, T., & Schulte, A. (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. doi: 0.1007/s12310-017-9217-4
- Owens, J. S., Holdaway, A. S., Smith, J., Evans, S. W., Himawan, L. K., Coles, E. K., Girio-Herrera, E., Mixon, C. Egan, T. E., & Dawson, A. E. (2017). Rates of common classroom behavior management strategies and their associations with challenging student behavior in elementary school. *Journal of Emotional and Behavioral Disorders*, 26, 156–169. doi: 10.1177/106342661771250
- Owens, J. S., Holdaway, A. S., Zoromski, A. K., Evans, S. W., Himawan, L. K., Girio-Herrera, E., & Murphy, C. E. (2012). Incremental benefits of a daily report card intervention over time for youth with disruptive behavior. *Behavior Therapy*, 43, 848–861.
- Owens, J. S., Schwartz, M., Erchul, W. P., Himawan, L. K., Evans, S. E., Coles, E. K., & Schulte, A. (2017). Teacher Perceptions of School Consultant Social Influence Strategies: Replication and Expansion. *Journal of Educational and Psychological Consultation*, 27, 411–436. doi: 10.1080/10474412.2016.1275649
- Pas, E. T., Bradshaw, C. P., & Cash, A. H. (2014). Coaching classroom-based preventive interventions. In M. S. Weist, N. A. Lever, C. P. Bradshaw, & J. S. Owens (Eds.), *Handbook of school mental health: Research, training, practice, and policy* (2nd ed., pp. 255–267). New York, NY: Springer.
- 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 & Adolescent Psychiatry*, 31, 210–218.
- Pelham, W. E., Greiner, A. R., & Gnagy, E. M. (2008). *Student behavior teacher response observation code manual* (Unpublished manual).
- Reddy, L. A., Fabiano, G., Dudek, C. M., & Hsu, L. (2013). Development and construct validity of the Classroom Strategies Scale-Observer Form. *School Psychology Quarterly*, 28, 317–341.
- Reinke, W. M., Lewis-Palmer, T., & Merrell, K. (2008). The classroom check-up: A classwide teacher consultation model for increasing praise and decreasing disruptive behavior. *School Psychology Review*, 37, 315–332.

- Reinke, W. M., Stormont, M., Herman, K. C., Wang, Z., Newcomer, L., & King, K. (2014). Use of coaching and behavior support planning for students with disruptive behavior within a universal classroom management program. *Journal of Emotional and Behavioral Disorders, 22*, 74–82.
- Robb, J. A., Sibley, M. H., Pelham, W. E., Jr., Foster, E. M., Molina, B. S., Gnagy, E. M., & Kuriyan, A. B. (2011). The estimated annual cost of ADHD to the US education system. *School Mental Health, 3*, 169–177.
- Rose, J. S., & Medway, F. J. (1981). Measurement of teachers' beliefs in their control over student outcome. *The Journal of Educational Research, 74*, 185–190.
- Sanetti, L. M. H., Collier-Meek, M. A., Long, A. C. J., Kim, J., & Kratochwill, T. R. (2014). Using implementation planning to increase teachers' adherence and quality to behavior support plans. *Psychology in the Schools, 51*, 879–895.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420–428.
- Solomon, B. G., Klein, S. A., & Politylo, B. C. (2012). The effect of performance feedback on teachers' treatment integrity: A meta-analysis of the single-case literature. *School Psychology Review, 41*, 160–175.
- Stormont, M., Reinke, W. M., Newcomer, L., Marchese, D., & Lewis, C. (2015). Coaching teachers' use of social behavior interventions to improve children's outcomes: A review of the literature. *Journal of Positive Behavior Interventions, 17*, 69–82.
- 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*, 654–672.
- Visser, S. N., Danielson, M. L., Bitsko, R. H., Holbrook, J. R., Kogan, M. D., Ghandour, R. M., . . . Blumberg, S. J. (2014). Trends in the parent-report of health care provider-diagnosed and medicated attention-deficit/hyperactivity disorder: United States, 2003–2011. *Journal of the American Academy of Child & Adolescent Psychiatry, 53*, 34–46.
- Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence* (2nd ed.). Bloomington, MN: Pearson.
- Wilkinson, L. A. (2007). Assessing treatment integrity in behavioral consultation. *International Journal of Behavioral Consultation and Therapy, 3*, 420–432.