Running Head: MULTI-COMPONENT CONSULTATION

Using Multi-Component Consultation to Increase the Integrity with which Teachers Implement Behavioral Classroom Interventions: A Pilot Study

Julie Sarno Owens[1] , Erika K. Coles[2], Steven W. Evans[1], Lina K. Himawan[1], Erin Girio-Herrera[3], Alex S. Holdaway[1], Allison K. Zoromski[1] , Terah Schamberg, and Ann Schulte[4]
[1]Ohio University
[2]Florida International University
[3]Towson University
[4]Arizona State University

Corresponding author: Julie Sarno Owens, Ph.D., Ohio University, Porter Hall 200, Athens, OH, 45701 USA; owensj@ohio.edu. This author declares she has no conflict of interest.

Erika K. Coles, Ph.D., [2]Florida International University, 11200 SW 8th St., Miami, FL 33199, USA, ecoles@fiu.edu. This author declares she has no conflict of interest.

Steven W. Evans, Ph.D., Ohio University, Porter Hall 200, Athens, OH, 45701 USA, evanss3@ohio.edu. This author is the Editor of *School Mental Health.* Thus, the handling of this article will be completed by an Associate Editor. He declares he has no other conflict of interest.

Lina K. Himawan, MS, Ohio University, Porter Hall 200, Athens, OH, 45701 USA, himawan@ohio.edu. This author declares she has no conflict of interest.

Erin Girio-Herrera, Ph.D., Townson University, 8000 York Road, Towson, MD 21252, USA, egirioherrera@towson.edu. This author declares she has no conflict of interest.

Alex S. Holdaway, MS, Ohio University, Porter Hall 200, Athens, OH, 45701 USA, ah218010@ohio.edu. This author declares she has no conflict of interest.

Allison K. Zoromski, MS, Ohio University, Porter Hall 200, Athens, OH, 45701 USA allisonzoromski@gmail.com. This author declares she has no conflict of interest.

Terah Schamberg, MS, Ohio University, Porter Hall 200, Athens, OH, 45701 USA tmschamberg@gmail.com. This author declares she has no conflict of interest.

Ann C. Schulte, Ph.D., Arizona State University, Ann.Schulte@asu.edu
This author declares she has no conflict of interest.

Published online: June 14, 2017 in *School Mental Health*, *9*, 218-234 doi: 0.1007/s12310-017-9217-4

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.

Abstract

The goal of this pilot study was to evaluate the effectiveness of a multi-component consultation package in improving teachers' classroom management skills, particularly among teachers with lower baseline levels of knowledge, skills, and intervention-supportive beliefs. Participants were 58 elementary school teachers (93% female; 50% Non-Hispanic White) who received up to eight biweekly consultation sessions focused on general classroom management strategies and implementation of a daily report card (DRC) intervention with one target student with or at-risk for ADHD. Teachers were randomly assigned to either a comparison consultation condition designed to mirror current best practices (Frank & Kratochwill, 2014; Noell & Gansle, 2014) or a multi-component condition designed to simultaneously address teacher knowledge, skills, and beliefs as possible barriers to implementation of classroom interventions. Teachers in both conditions showed significant improvements in labeled praise, appropriate response to student rule violations, and general competence in classroom management. In support of the hypotheses, teachers with lower baseline levels of knowledge, skills, and intervention-supportive beliefs demonstrated more improvement in key outcomes in response to multi-component consultation, as compared to the comparison consultation (Cohen's *d* ranged from .33 to 1.12). Implications for research and practice in school consultation are discussed.

Keywords: consultation; integrity; classroom management; teacher; beliefs; performance feedback; coaching

Using Multi-Component Consultation to Increase the Integrity with which Teachers Implement Behavioral Classroom Interventions: A Pilot Study

Recent estimates suggest that 10% to 20% of elementary students demonstrate behaviors associated with attention-deficit/hyperactivity disorder (ADHD) (Fabiano et al., 2013; Visner et al., 2014). These behaviors are stressful for teachers (Greene, Beszterczey, Katzzenstein, Park, & Goring, 2002), detract from instruction time (Robb et al., 2011), and are a common reason that teachers feel the need for support from other school professionals (e.g., Shernoff, Mehta, Atkins, Torf, & Spencer, 2011). General classroom management strategies (e.g., use of rules, routines, praise) and targeted interventions (e.g., daily report card: DRC) are effective in improving academic and behavioral functioning in children with inattentive and disruptive behaviors (Epstein, Atkins, Culinan, Kutash, & Weaver, 2008; Evans, Owens, & Bunford, 2014; Vannest, Davis, Davis, Mason, & Burke, 2010). However, few teachers report feeling well trained to implement these strategies (Parsad, Lewis, & Farris, 2001) and teacher adoption of targeted interventions is relatively limited (Martinussen, Tannock, & Chaban, 2011).

One widely used means to support teachers' use of general and targeted interventions is consultation with a mental health professional skilled in evidence-based classroom interventions (Frank & Kratochwill, 2014). Yet, even when teachers receive consultant support, the integrity with which they implement classroom interventions is variable (e.g., Owens, Murphy, Richerson, Girio, & Himawan 2008), which can severely compromise positive student outcomes (e.g., Wilder, Atwell, & Wine, 2006). Given the serious personal and societal costs associated with poorly implemented interventions, it is important to understand barriers to teachers' implementation of classroom interventions for disruptive behavior and to develop consultation programs that target these barriers.

Three malleable teacher characteristics that are possible facilitators or barriers to quality implementation of classroom interventions are (a) knowledge about best practices, (b) skills to implement these practices, and (c) beliefs about the acceptability or feasibility of interventions, or one's own skills in implementing interventions (Sanetti & Kratochwill, 2009). Although some researchers have examined the effects of embedding strategies within consultation to target one of these malleable characteristics, such as increasing teacher skills (e.g., Sanetti, Collier-Meek, Long, Kim, & Kratochwill, 2014) or changing beliefs (e.g., Cook, Lyons, Kubergovic, Wright & Zhang, 2015), theories of adult behavior change (Stuart, Tondora, & Hoge, 2004) suggest that all three may need to be addressed to maximize the outcomes of consultation.

The current study was designed to examine whether a multi-component consultation package, that simultaneously addresses knowledge, skills, and beliefs as possible barriers to intervention integrity, can produce greater change in elementary school teachers' classroom management skills than a comparison consultation package that represents current best practices.

Best Practices in School Consultation

School consultation is an indirect service delivery model in which a triadic relationship is developed between the consultant and teacher who then provides direct services to a student. A five-step problem solving approach has been identified as a consultation best practice (Frank & Kratochwill, 2014). The five steps include: forming a collaborative relationship, identifying and analyzing the problem, selecting an intervention, implementing an intervention, and evaluating outcomes. Although some teachers may be hesitant to receive consultation, current legislation related to inclusive education and use of research-based interventions makes the need for consultation a reality for most teachers (Feldman & Kratochwill, 2003; Fuchs & Fuch, 1996).

There is a large body of literature that shows that behavioral consultation, particularly

when combined with observation and performance feedback procedures, is effective in improving teacher integrity and student outcomes (Noell & Gansle, 2014; Solomon, Klein, & Politylo, 2012), and acceptable to teachers (Sheridan, 1992). However, even in the context of these best practices, there is variability in teachers' integrity (Noell, Witt, Gilbertson, Ranier, & Freeland, 1997; Owens et al., 2008). Although there are requirements that interventions be tailored to the needs of individual students (see DuPaul, Power, Evans, Mautone, & Owens, 2016 for discussion), less attention has been given to how to tailor consultation to the needs of a given teacher. The variability in teachers' intervention integrity prior to and after best practice consultation (e.g., Noell et al., 1997) suggests that some teachers need more intensive support and others do not, and some teachers need different types of support. Our goal was to develop and evaluate a multi-component consultation package that builds upon best practices and addresses knowledge, skills, and beliefs that posed limits to intervention integrity.

Development of a Multi-Component Consultation Package

Many factors can affect integrity (e.g., practical barriers such as time and resources, and contextual factors such as administrative support for an intervention; see Aarons, Hurlburt, & Horowitz, 2011). In this study we focus on malleable teacher-level factors (i.e., baseline knowledge, skills, and beliefs) known to be related to integrity. Indeed, by enhancing these factors, teachers' may be better equipped to overcome some of the practical barriers (Han & Weiss, 2005). Theories of adult learning (e.g., Miller, 1990; Stuart et al., 2004) suggest that the path to adult behavior change is an iterative, individually-tailored process that systematically and distinctively addresses these three malleable factors.

Knowledge. Extensive data shows that professional development training increases participant knowledge (see Beidas & Kendall 2010 for review; Jones & Chronis-Tuscano, 2008).

When developing this consultation package, we viewed teachers' knowledge of ADHD and behavior theory as foundational to implementing general classroom management strategies and a targeted intervention (i.e., a DRC) for a specific student. Thus, in addition to an initial 3-hour workshop, we included a knowledge enhancement component in each multi-component consultation session. It was hypothesized that these activities would offer individualized information to enhance knowledge, and thus, facilitate higher quality implementation, particularly among teachers with lower baseline levels of knowledge. We also recognized that this component, while necessary, is likely insufficient for enhancing integrity on its own (Beidas & Kendall, 2010). Thus, our multi-component package also included tools for modifying skills and beliefs relevant to intervention implementation.

Skills. There is extensive evidence that problem-solving consultation, coupled with observation and performance feedback, increases teachers' implementation skills, resulting in improvements in students' outcomes (see Noell & Gansle, 2014 for review). Thus, the skills component that we developed and evaluated included opportunities to (a) observe the consultant or video model, (b) practice skills under simulated conditions, and (c) receive feedback in a graphic form about student progress and implementation procedures. It was hypothesized that this component would be an important mechanism for growth in classroom management skills. However, we recognized that variability in integrity may remain even with best practice consultation (Noell et al., 1997), suggesting that it may be necessary to target knowledge, skills, *and* intervention-related beliefs to achieve high integrity (Cook et al., 2015; Stuart et al., 2004).

Beliefs. Teacher beliefs relevant to intervention integrity include perceptions about acceptability of the intervention (Allinder & Oats, 1997), self-efficacy in delivering the intervention (Tschannen-Moran, Hoy, & Hoy, 1998), and agency (internal locus of control) and

motivation to implement the strategies (Kealey, Peterson, Gaul, & Dinh, 2000). A recent study provides evidence that targeting teacher beliefs in consultation can produce change in beliefs and that change in beliefs is associated with higher quality implementation scores (Cook et al., 2015). These results support the inclusion of a component targeting beliefs in consultation. However, the intervention evaluated in the Cook et al. (2015) study was schoolwide and scores from the beliefs measure and implementation measure reflected an aggregate across teachers within a school; thus, it is unclear how individual beliefs about specific practices affect implementation.

The multi-component consultation package that we developed and evaluated was focused on the beliefs of individual teachers and included techniques informed by the motivational interviewing (Miller & Rollnick, 2013; Reinke, Herman & Sprick, 2011) and cognitive behavioral therapy literatures. Consultants (a) assessed each teacher's values and intervention-related beliefs, (b) elicited change talk (i.e., statements that evoke the teacher's desire to change, reasons or benefits of implementation, and beliefs in his/her ability to implement the intervention), as change talk has been found to predict change in adult behavior in other contexts (Amrhein, Miller, Yahne, Palmer, & Fulcher, 2003), and (c) used decisional balance (Miller & Rollnick, 2013) and Socratic questioning techniques to consider alternatives to beliefs that were possible barriers to implementation (teachers were encouraged to see the situation through "an alternative lens"). We hypothesized that modifying beliefs that are obstacles to implementation, while also applying knowledge or skill enhancements, would enhance implementation integrity (i.e., use of labeled praise and appropriate response to student rule violations).

Current Study

The goal of this pilot study was to evaluate the effects of a multi-component consultation package on teacher behaviors that are key indicators of effective classroom interventions (i.e.,

labeled praise, appropriate response to student rule violations). While we expected teachers in both conditions to demonstrate improvements in skills, we hypothesized that the multi-component condition may produce greater improvement in teacher skills than a comparison condition (that represents current best practices in consultation; Aim 1). Further, we expected that the difference between the two consultation conditions would be largest in the subsample of teachers with barriers to integrity (i.e., low baseline levels of knowledge, skills, and intervention-supportive beliefs (Aim 2). Namely, there is evidence that current best practices are sufficient to improve the skills of many teachers; yet, an individualized approach may be more effective when barriers to implementation are present.

Method

Participants

General education teachers were recruited from eight participating schools across two sites (Ohio and Florida). Participating teachers (N = 58; 28 from Ohio, 30 from Florida) were Non-Hispanic White (50%) and Hispanic (any race; 44.8%) women (93.1%), with an average of 14.23 years (*SD* = 8.57) of teaching experience and 8.32 years (*SD* = 6.63) teaching at their current school. Most (62%) had obtained a master's degree. The five Ohio 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. The three Florida schools had an average of 1,024 students and 50 general education teachers, with 4% - 11% receiving special education services and 76 - 95% receiving free or reduced lunch services. The average class size across sites ranged from 19 to 25 and the teacher was the sole educator in the room. Consultants (referred to as facilitators) were post-doctoral fellows (*n*=2), master's level clinicians (*n*=2) or graduate students in a master's or doctoral program in psychology (*n* = 5). Six

identified as Caucasian, one identified as African American, and two identified as Hispanic.

Target student participants were 58 elementary school students (77.6% male; 53.4% Latino). Most (93.1%) met criteria for ADHD (67.2% combined presentation; 22.4% inattentive presentation; 3.4% hyperactive/impulsive presentation) and 6.9% were at risk for ADHD (elevated symptoms plus impairment). The sample had an average IQ estimate of 97.77 (*SD* = 12.65), as assessed by the Wechsler Abbreviated Scales of Intelligence, Second Edition (WASI-II; Wechsler, 2011). The socio-economic status of their families was low to middle class (17.2% had a household income under $15,000, 55.2% had an income between $15,000-49,999; 18.9% were above $50,000; 8.6% did not report income). Per parent report at intake, 8.6% had been diagnosed with a learning disability and 24.1% had a prescription for a psychiatric medication.

Study Procedures

All general education teachers in each elementary school were invited to a 3-hour workshop conducted by the investigators that focused on best practices in general classroom management strategies and the DRC. Following the workshop, teachers provided consent and completed a battery of questionnaires (that included those described below), in a computer lab using the Research electronic data capture (REDCap) system (Harris et al., 2009). Teachers received $60 for the workshop and $25 for completing the questionnaires.

At the workshop, teachers were recruited for the consultation project. To participate, teachers were required to identify one student with or at risk for ADHD. School staff obtained permission from parents to be contacted by program staff. Program staff contacted the parents, described the DRC intervention and consultation process and scheduled an appointment to obtain parent consent, child assent, and determine student eligibility. Inclusion criteria for child participants were the following: (a) children were enrolled in a general education classroom (K-

5) for at least 50% of the day, (b) had an IQ estimate that fell in the 90% confidence interval for a score of 80, and (c) met diagnostic criteria for DSM-IV ADHD or were at-risk for ADHD. ADHD was defined as the presences 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 et al., 1998) or the parent- or teacher-version of the *Disruptive Behavior Disorders Rating Scale* (Pelham et al., 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 in the context of the P-CHiPs interview helped to rule out other disorders as sources of ADHD symptoms and to assess the chronicity and dual setting presence of symptoms. At risk status was defined as four or more symptoms and impairment in the school setting. Children were excluded from participation if they had a previous diagnosis of an Autism Spectrum Disorder, Bipolar Disorder, or Cognitive or Developmental Disability per parent report.

Once a target child was identified and teacher consent was obtained, at least two classroom observations (see description below) were conducted to obtain a baseline assessment of each teacher's competence in classroom management. Using the global classroom management competence ratings, averaged across all baseline observations, teachers' skills were categorized as low (5 or below), medium (6 or 7), or high (8 or above) on a 10-point scale. Within each school, teachers were block randomized on competence to one of two consultation conditions to ensure similar numbers of teachers at each competence level in each condition within schools. This resulted in 31 teachers in the multi-component condition and 27 in the comparison condition; teachers in each condition did not differ in baseline competence ratings (M = 6.36; SD = 1.51 for multi-component; M = 6.56, SD = .1.21 for comparison, $t(58) = .570$, $p = .571$).

Consultation Procedures

Consultation in both conditions focused on general classroom management strategies (i.e., labeled praise, use of rules, effective instructions, and appropriate response to rule violations) and the use of a DRC intervention. In both conditions, teachers participated in (a) an interview about the teacher's approach to classroom management, (b) a target behavior interview to assess the target child's strengths and weaknesses and identify DRC target behaviors, (c) baseline tracking of target behaviors for five school days, and (d) a DRC development meeting to review baseline data and finalize target behaviors and goal criteria (see Table 1).

Once the DRC was launched, teachers in both conditions were observed weekly and received biweekly (every other week) performance feedback. Procedures included (a) review of the teacher's use of specific classroom management strategies as observed over the last two observations, (b) review of graphs depicting the target child's progress on DRC target behaviors, (c) praise for correct implementation of classroom management strategies, and (d) corrective feedback about integrity dimensions as dictated by the data ("areas for growth"). Sessions ranged from 30 minutes to 1 hour, and were conducted during, before, or after school. Teachers were not incentivized for participation in any of the consultation activities (e.g., attendance at consultation meetings; classroom observations, implementation behaviors).

Comparison condition. Once the DRC was launched, consultation in the comparison condition followed a general problem-solving process including identifying and analyzing the problem (e.g., child not responding to the DRC), brainstorming possible solutions, selecting a strategy to address the problem, and evaluating possible consequences/outcomes of the strategy selected. Teachers received brief performance feedback, including graphs of the classroom observation data and DRC progress, with teacher strengths and areas for growth highlighted.

Feedback was designed to mirror (in duration, content, and process) best practice procedures reported in previous research (Gilbertson, Witt, Singletary, & VanDerHeyden, 2007; Noell et al., 1997) and refrain from incorporating the active ingredients in the multi-component condition described below. The guiding principles were that performance feedback should be limited to 5 to 10 minutes and unless the teacher *initiated* discussion of other content, the problem solving remained child-focused. Discussion of teacher values and beliefs, and attempts to facilitate change talk were contra-indicated.

Multi-component condition. Facilitators 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 described below (see Table 1).

Knowledge component. In addition to the initial 3-hour in-service, teachers received education in each consultation session via fact sheets titled *News You Can Use (and Things You Might Remember).* They included checklists, reminders, facts from the literature, self-assessments related to the strategies discussed in consultation, and a vignette to highlight "real-world" applicability. Factsheets were given to the teacher one week prior to each session.

Skills component. Teachers received enhanced performance feedback to address areas for growth in skills. Namely, after reviewing the data from the observations and child's DRC graphs, facilitators highlighted connections between teacher integrity behaviors and child outcomes and used this conversation as a catalyst for (a) skills practice or (b) discussion of related beliefs (see below). Techniques informed by motivational interviewing were used to highlight the discrepancy between the teacher's desired goal (e.g., improved child behavior) and his/her current practices (e.g., limited appropriate response to rule violations). Facilitators attempted to elicit change talk from the teacher (e.g., possible benefits of making improvement

in a given skill) and to facilitate the teacher's willingness to engage in a skills practice and use the identified skill in the upcoming weeks. Skills practice involved selecting a skill that was identified as an area for growth and participating in an activity that involved application of that skill (e.g., role playing with or without a script; generating a list of specific praise statements for the upcoming week; observing and discussing the strengths and weaknesses of a video model).

Beliefs component. The multi-component condition included a Values Interview (Frey et al., 2013) that invited teachers to identify their top three values/priorities (e.g., student achievement, self-esteem, community) related to teaching. Facilitators were trained and supervised to (a) incorporate the teacher's values into the problem-solving process (e.g., *if we decide to do X, how will that facilitate your value of Y?*) and (b) elicit change talk from teachers, as motivational interviewing theory suggests that voicing statements aloud to others about intentions to change has a greater impact on actual behavior change than internal statements alone (Miller & Rollnick, 2013). Facilitators also listened for sustain talk (i.e., reasons for not implementing the selected strategies) and beliefs that may have served as barriers to implementation (e.g., *If I give consequences, it will lead to poor student self-esteem*; *I don't have time to implement the DRC*). When such beliefs were identified, the facilitator and teacher discussed the belief, its potential impact on teaching and student-teacher interactions, and an alternative belief was generated (e.g., *Giving consequences is a way to show I care about the child's development*). Then, the facilitator used a decisional balance technique (Miller & Rollnick, 2013) to explore the pros and cons of each belief, or a Socratic questioning technique to explore the short and long-term outcomes of each belief. The facilitator and teacher then reflected on the extent to which the alternative belief may facilitate better implementation and improved student behavior, and align with the teacher's values.

Consultation Integrity. To ensure integrity to each consultation condition, facilitators attended a 3-day training. Prior to working in the schools, facilitators were required to role play each session until they achieved mastery in demonstrating and differentiating techniques the in each condition. For the duration of the trial, consultation sessions were audio recorded and sections were reviewed before each weekly supervision session, where facilitators were given feedback about their adherence to each condition. Approximately 10% of all sessions were coded to examine facilitator adherence to the steps, competence in implementing the steps (e.g., with attention to individual differences), and differentiation of conditions. Percent adherence for each session was calculated by dividing the total number of steps achieved by the total prescribed for the session. Across all sessions, the average percent achieved was 92.93% (*sd* = 6.65).

We examined several behaviors that were intended to occur only in the multi-component condition. For the use of open-ended questions and/or reflections that facilitated change talk, these behaviors occurred, on average, 12.36 times per session (*sd* = 6.41) in the multi-component sessions and 0.43 time per session (*sd* = .30) in the comparison condition. Similarly, there were several contra-indicated behaviors in the comparison condition, including talking about teacher values or beliefs or engaging in role play activities. These behaviors occurred only once across all coded comparison sessions (one facilitator mentioned teacher values). Further, each consultant's competence in consultation was coded on a 5-point scale ranging from 1 (*weak/below expectations: engaged in several behaviors that warrant improvement in this domain* to 5 (e*xceeded expectations in all behaviors in this domain),* with ratings of *3* indicating *met expectations.* For competencies that applied to both conditions (e.g., *Maintains a professional, yet warm demeanor*) facilitators were consistently (100% of the time) at or above expectations. For competencies that involved the use of MI-informed skills (e.g., *uses strategies to increase*

motivation for change or commitment), ratings for multi-component condition met expectations or exceeded 66% of the time, and ratings for the comparison condition were weak/below expectations for these skills 100% of the time.

Equivalence of dose across condition. The intent was for teachers in both conditions to have eight biweekly sessions and two observations between each session. Nonetheless, due to unforeseen circumstances (e.g., weather-related school closings, teacher or student absences), sessions and observations were occasionally canceled. The conditions did not differ on average number of biweekly sessions completed (multi-component: $M = 6.23$; $SD = 2.25$; comparison: $M = 6.19$; $SD = 2.32$; $t(56) = -.068$, $p = .946$), average number of total observations (multi-component: $M = 18.58$; $SD = 5.77$; comparison $M = 16.78$; $SD = 5.40$; $t(56) = -1.22$, $p=.227$)[1], or observations per week between any biweekly consultation. On average, multi-component sessions lasted 38 minutes and comparison sessions lasted 23 minutes.

Measures

Teacher Knowledge of Behavioral Principles as Applied to Children. This is a 16-item multiple-choice measure that assesses teacher knowledge of behavioral principles. Items were developed for this study based on a review of the literature and the Behavior Modification Test (Kratochwill, Elliott, & Busse, 1995). A total percent correct was calculated. 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).

Teacher Knowledge of ADHD. This is a 24-item True/False/Don't Know measure, inspired by Jones and Chronis-Tuscano (2008) that assesses teacher knowledge of ADHD (prevalence, etiology, treatment). Responses were coded as correct or incorrect. A total percent

[1]If a bi-weekly meeting was rescheduled for a subsequent week, we continued to obtain one observation per week, which resulted in more observations than expected.

correct was calculated. The measure has demonstrated sensitivity to change as a function of participating in a workshop focused on ADHD and classroom management (Owens et al., 2014).

Teacher Locus of Control. This is a 25-item measure that assesses teacher perceptions of personal control and responsibility for student academic and behavioral outcomes (Rose & Medway, 1981). For each item, there are two response options, one aligned with greater internal sense of responsibility and one aligned with greater external sense of responsibility. Two subscales include student success (11 items) and student failure (14 items). Higher scores indicate greater internal responsibility for student success and failure, respectively. Scores are predictive of teacher's use of techniques learned during an in-service (Rose & Medway, 1981). We viewed higher scores (i.e., internal locus of control) to be associated with intervention-supportive beliefs. Acceptable internal consistency has been found in previous studies; Kuder-Richardson formula 20 (KR20) reliability scores of .81 for failure and .71 for success (Rose & Medway, 1981). The KR20 scores with this sample were .71 for failure and .46 for success. Given the low score for success (and that removing items from this scale did not improve the reliability scores), this subscale was not used in the analyses.

Teacher skills in classroom management. Under the umbrella of integrity, we conceptualized adherence as the enactment of specific teacher behaviors. These dependent variables were obtained via a modified version of the Student Behavior-Teacher Response Observation Rating System (SBTR; Pelham, Greiner, & Gnagy, 2008); adequate inter-rater reliability, convergent validity, and sensitivity to change has been found in preschool and elementary classrooms (Fabiano et al., 2010; Vujnovic et al., 2014). Using this system, observers obtained (a) frequency counts of rule violations by the target student separately and by all other students in the classroom collectively, (b) frequency counts of how the teacher responded to each

rule violation, and (c) frequency counts of the teacher's use of praise. 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), for coding how the teacher responded to each rule violation (i.e., appropriately[2], inappropriately, or no response), and for coding labeled and unlabeled praise. During the first consultation meeting, all teachers were informed of the rules to be coded. Although consultants asked which rules do not apply in each teacher's classroom, all seven rules were coded for all teachers to maintain consistency in the data.

We conceptualized competence as global indicator of the extent to which all of the classroom management strategies employed by the teacher were consistent with best practices in classroom management. Namely, following each observation, observers rated the teacher's global competence in classroom management on a 10-point scale (ranging from *inconsistent with (1)* to *entirely consistent with (10) best practices in classroom management*).

Because appropriate response to rule violations (of the target child and all other children) and use of labeled praise are best practices in classroom management and DRC implementation, and because these were primary skills taught throughout consultation, these variables, and the global competence rating were the primary outcome variables. In addition, in a recently-published study (Owens et al., in press), we found that higher percentages of appropriate teacher response to rule violations were significantly associated with lower rates of student rule violations, such that class wide rule violations dropped to less than 1 per minute once teachers

[2] 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, 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.). All definitions are available upon request from the first author.

reached a threshold of 51% appropriate response, with little incremental benefit at higher levels.

Observation durations ranged from 15 to 45 minutes; thus frequency counts for rule violations (target student and other students) and labeled praise counts were transformed into rates per hour for each observation. These rates were averaged across all observations for the case. The total number of appropriate teacher responses to target student violations for a given observation was divided by the total number of violations by the target student for that observation period. This produced the percent of appropriate response to violations per observation. These percentages were averaged across all observations. This was repeated for appropriate response to violations by all other students.

Facilitators and research assistants (unaware of teacher condition) were trained to reliability on the SBTR. They attended an initial training, were required to pass (100% accurate) a written definitions test, were required to pass (100% accuracy) coding of 2-5 minute video clips, and were required to achieve at least 80% reliability across all coded behaviors in a classroom with a master observer. Maintenance of reliability was checked throughout the year. Inter-observer assessments were conducted for 24% of all observations. We computed Intraclass Correlations (ICC) of type 1 for average of k raters (ICC(1,k) as outlined in Shrout and Fleiss (1979) or ICC(k) for Case 1 as outlined in McGraw and Wong (1996); henceforth, we use the notation of ICC(1,k)) because we did not have a set of consistent raters across all ratees and we would like to observe the reliability across a set of raters rather than one rater only. Across all frequency count variables, the ICC(1,k) ranged from 0.78 to 0.98 with an average of 0.90. The ICC(1,k) for the global classroom management competence rating was .68.

Teacher satisfaction with consultation. At the end of the program, teachers responded to several satisfaction items about the program. Two items asked about the teacher's perception

of the helpfulness of the bi-weekly consultations and the consultant. Some items asked about the benefits for the teacher and other items asked about the benefits observed in student behavior and academic functioning. Lastly, teachers were asked if they would recommend the program to other teachers.

Results

Aim 1: Whole Sample Analyses

To examine possible growth in teacher behavior as a function of consultation condition, the data were analyzed using hierarchical linear modeling (HLM). The primary dependent variables were (1) percent appropriate response to target child (2) percent appropriate response to other child rule violations, (3) rate of labeled praise/hour, and (4) the global competence rating, each averaged across observations in a given biweekly time period[3]. Time (i.e., biweekly period) was treated as continuous and was centered at the last data point. Several growth curves (i.e., linear, log and quadratic) were fitted and Akaike information criteria (AIC) and Bayesian information criteria (BIC) were used to determine the best fit. A log growth curve fit two variables: percent appropriate response to target child rule violations and percent appropriate response to other child rule violation. A quadratic growth curve fit the two other variables: rate of labeled praise, and global competence in class management. The models fitted were as follows

Log growth curve

Level-1: $y_{ij} = \pi_{0j} + \pi_{1j}\,(\text{Log of Time})_{ij} + e_{ij}$

Level-2: $\pi_{0j} = \gamma_{00} + \gamma_{01}\,(\text{Consultation Condition})_j + r_{0j}$

Level 2: $\pi_{1j} = \gamma_{10} + \gamma_{11}\,(\text{Consultation Condition})_j$

Quadratic growth curve

Level-1: $y_{ij} = \pi_{0j} + \pi_{1j}\,(\text{Time})_{ij} + \pi_{2j}\,(\text{Time}^2)_{ij} + e_{ij}$

Level-2: $\pi_{0j} = \gamma_{00} + \gamma_{01}\,(\text{Consultation Condition})_j + r_{0j}$

Level 2: $\pi_{1j} = \gamma_{10} + \gamma_{11}\,(\text{Consultation Condition})_j$

[3]Although there were some differences across sites, the variability across schools within a given site was two to ten times greater than the differences between schools when aggregated by site. Thus, data from the total sample, combined across sites, are presented and analyzed.

Level 2: $\pi_{2j} = \gamma_{20} + \gamma_{21}$ (Consultation Condition)$_j$

In all models, the effect of time was significant (all *p*s <.01), indicating significant improvement in all teacher behaviors. However, the effect of consultation condition was not significant, indicating that, on average, teachers in the multi-component condition did not demonstrate significantly more improvement in these targeted behaviors than teachers in the comparison condition.

Aim 2: Subsample Analyses

Identification of teachers with knowledge and belief barriers. First, we conducted a latent class analysis to statistically identify a class of teachers with lower knowledge, and intervention-supportive beliefs (i.e., internal responsibility for student outcomes). The latent class analysis was conducted using three global variables: knowledge of ADHD, knowledge of behavior modification, and locus of control for students' failure (locus of control for students' success was dropped due to reliability issues mentioned above). Two-, three- and four-class models were fit to the data and compared on several statistics: Bayesian Information Criteria (BIC), entropy and several Likelihood Ratio Test (LRT). For BIC a lower number represent a better fit. Entropy is a measure of classification uncertainty; an entropy of > 0.8 is preferable since it indicates high certainty. There are several variations of the LRT, three were compared here: Lo-Mendell-Rubin LRT (LMR LRT), Vuong-Lo-Mendel-Rubin LRT (VLMR LRT) and bootstrap LRT (BLRT). As can be seen from Table 2, the two-class model has the lowest BIC and the LRT indicating that the two-class model should be chosen over the one-class and three-class solutions. Although the entropy of the two-class solution is the lowest among the three, the three-class and four-class models resulted in classes that were quite small (e.g., n=3). Thus, the two-class model was chosen over the other models.

Table 3 provides the means, standard deviations, and effect sizes for teachers' knowledge

and beliefs by latent class. Relative to teachers in Latent Class 1, teachers in Latent Class 2 were less knowledgeable and took less credit for their students' failure (*d* effect sizes range from 1.55 to 2.89). Thus, Latent Class 2 represents teachers whom we hypothesized would most benefit from the multi-component consultation.

Identification of teachers with skill barriers. Next, we identified teachers within each latent class who were lower in baseline skills (i.e., in each of the dependent variables); those who were higher in baseline had minimal room to improve. Thus, for each dependent variable and each latent class, we (a) sorted teachers by individual baseline scores for that dependent variable and (b) divided the teachers into two groups using a median spilt (high in baseline skill and low baseline skill). Descriptive statistics for slope, baseline value, and endpoint value (i.e., bi-weekly period before the last consultation) for each dependent variable for high and low baseline teachers are presented in Tables 4 and 5. We expected those with the low baseline scores in Latent Class 2 (knowledge and belief barriers) to be most in need of and to benefit most from the multi-component condition relative to the comparison condition.

We recognize that a median split is arbitrary and that by using a median split on each skill, a given teacher may fall in the low baseline group for one dependent variable and in the high baseline group for another dependent variable; nonetheless this best allowed us to test whether the multi-component condition produced better outcomes than the comparison condition for teachers in Latent Class 2 with low levels of a given skill.

Growth by consultation condition, latent class and baseline groups. The log growth curve for two dependent variables (i.e., percent appropriate response to target child rule violations and other child rule violations) showed that change happened in the same direction over all time points. Thus, a linear regression was fit to the baseline and the means for each of

the possible eight biweekly periods of each teacher to obtain an estimate of each teacher slope for these variables. For the two other variables (i.e., rate of labeled praise and global competence in class management), one slope could not represent the patterns of change over time because the quadratic growth curve showed that there was improvement over the first few time points and the trajectory changed direction over the latter time points. Because we were most interested in initial response to the consultation, we used the HLM model to obtain the optimum biweekly observation point (that is before the curve changed its direction) and fitted a linear regression line to the baseline and the means of the biweekly observations before the optimum point was reached by each teacher. Effect sizes[4] in average improvement (slope) in teacher skills as a function of Consultation Condition, Latent Class, and High/Low Baseline Groups are provided in Table 6.

There were several noteworthy patterns across the dependent variables (comments below are based on effect sizes not statistical tests). First, in support of the hypothesis, teachers who had lower knowledge and intervention-supportive beliefs (Latent Class 2), who started below the median on a given skill (Low Baseline) demonstrated greater change in the multi-component condition relative to the comparison condition (all four effect sizes in boldface type in Table 6 are positive). Second, for one variable (i.e., response to other child rule violations) the Latent Class 2 Low Baseline teachers achieved a similar level of skill by the end of consultation as teachers in the Latent Class 2 High Baseline group (see Figure 1). However, this was not achieved for all variables (e.g., response to target child rule violation; see Figure 1 and Tables 4 and 5). Lastly, Latent Class 1 High Baseline teachers (i.e., teachers with no barriers) benefited

[4] Effect sizes were derived from the difference between the slopes in each condition divided by the pooled standard deviation. Positive values indicate an advantage for the multi-component condition; negative values indicate an advantage for the comparison condition.

similarly from both conditions, with an average effect size (across all variables) between the conditions of .06 (see Table 6).

Satisfaction data are presented in Table 7. All satisfaction ratings are high. In most cases, the satisfaction ratings for teachers in the multi-component condition (regardless of latent classes and baseline group) are higher than that of their counterparts in the comparison condition, yet these differences are not statistically significant.

Discussion

The goal of this study was to evaluate the effects of a multi-component consultation package on effective teacher classroom management practices. To our knowledge, this is the first study to evaluate a consultation package that simultaneously addresses individual teacher knowledge, skills, and beliefs. Our results provide support for current best practices in consultation (i.e., problem solving with observation and performance feedback), and offer evidence that an individually-tailored, multi-component consultation package may be particularly effective among teachers whose profile of skills, knowledge, implementation-related beliefs may serve as barriers to intervention implementation.

First, our findings are consistent with prior studies documenting the powerful impact of problem solving consultation with performance feedback for all teachers and contribute to the collection of data supporting this as an evidence-based practice (Noell & Gansle, 2014). Namely, teachers in both conditions demonstrated significant improvement in many skills and were satisfied with the consultation they received. With the total sample, there were no differences in skill growth between the multi-component and comparison conditions. This finding is likely for two reasons. Our comparison condition represents best practices in consultation, and thus, is a difficult threshold to improve upon. Further, as hypothesized in Aim 2, the value of the multi-

component condition may be strongest (or only observed) with teachers who demonstrate knowledge, skills, and beliefs that impact intervention implementation. Findings from Aim 1, suggest that among teachers who do not have barriers to integrity, the enhancement strategies within the multi-component condition would not be needed.

Second, the data in Table 2 document the wide variability in teachers' baseline knowledge and beliefs. Further, the results of our latent class analyses suggest that the variability converges around two teacher profiles, with striking differences between the two groups (effect sizes >1.5). These differences suggest that, with additional study, researchers and school personnel may be able to, a priori, consider which intensity of consultative resources (e.g., from minimally intensive to most intensive) would be most beneficial for an individual teacher. Further, this variability provided the opportunity to examine the differential impact of the multi-component consultation condition among teachers whose knowledge, skills, and beliefs may limit intervention implementation.

Third, the relative growth of teachers in the different groups provides support for the theory of change (Coles et al., 2015). The results of Aim 2 analyses, indicate that teachers who were low in general knowledge and intervention-supportive beliefs, and low in baseline skills showed more growth in response to the multi-component package, compared to the comparison consultation package (see Table 6). In schools where the availability of consultants is limited, identifying those teachers who are most in need of comprehensive consultation could make the use of consultants more efficient. Taken together, the outcomes are consistent with theories of adult learning (Stuart et al., 2004) that suggest that a consultation package equipped with strategies to simultaneously address limitations in knowledge, skills, and beliefs may maximize teacher outcomes.

Lastly, several patterns warrant additional consideration. For example, it is interesting that the effect sizes and the significant findings of the HLM models were greater for the competence rating than the behavioral frequencies (within both Latent Class 2 groups; see Table 6). Competence ratings are intended to reflect a global indicator of effectiveness, and thus encompass multiple teacher behaviors. It is possible that change in competence ratings reflect a cumulative benefit across multiple behaviors (those coded and not coded in the SBTR system), thus producing a larger effect size than any single micro-behavior. Further, it is noteworthy to recognize that the competence ratings have a cap (i.e., a score of 10) whereas the behavioral frequencies do not. Thus, the effect sizes for the competence ratings among high baseline teachers may be smaller (than the effect sizes for the behavior frequencies) as a function of statistical compression rather than a limited intervention effect (see Tables 4 and 5).

In addition, it is worth acknowledging the pattern of response across the four groups of teachers. Namely, teachers in High Baseline Latent Class 1 group benefited from both conditions nearly equivalently, with an average effect size (across all variables) between the conditions of .06 (see Table 6). These teachers did not have barriers targeted by this intervention. Thus, the focus on enhancing knowledge, skills and beliefs was not a value to them and their response to the two conditions were essentially equivalent. In contrast, Latent Class 1 Low Baseline teachers did not have knowledge or beliefs barriers, but did have lower skills; and the Latent Class 2 High Baseline teacher had knowledge or beliefs barriers but not skill deficits. Thus, the opportunity to focus on the existing barriers may have produced the slight advantage for the multicomponent condition. Lastly, those with all three barriers (Latent Class 2 Low Baseline teachers) benefitted the most (and the most consistently across all variables) from the multi-component condition.

Lastly, of particular concern are the low levels of all teachers' baseline levels of effective

classroom management strategies. Alarmingly, data in Table 4 show that, on average, teachers are responding appropriately to less than 30% of student rule violations. This low rate is consistent with the national surveys that document teachers' perceptions that they are ill-prepared to manage disruptive classroom behavior (Parsad et al., 2001). The low rate is also consistent with a recent study that found high rates of rule violations in classrooms in which teacher respond to less than 50% of rule violations (Owens et al., in press). Given the strong relationship between teachers' use of specific instructional and classroom management strategies, and student academic and behavioral outcomes (e.g., Gable, Hester, Rock, & Hughes, 2009; Owens et al., in press; Wang, Haertel, & Walhberg, 1993), these data reaffirm the need for better training (pre-service and in-service) and ongoing support on classroom management.

Implications

We believe the above outcomes advance the science of consultation and have several implications for future research, practice, and policy. First, these data offer sufficient evidence that the theory of change that guided this study (Coles et al., 2015) is worthy of study in a large clinical trial to address many remaining questions. Second, the variability in teacher knowledge and beliefs suggests there is utility in additional research that establishes reliable and valid tools for measuring teacher characteristics that are most related to high quality classroom management skills. Such research could facilitate the identification of teacher characteristics that may moderate response to consultation and/or prioritize groups of teachers for specific types of support. Such specificity could help school personnel expend consultative resources efficiently. Third, in addition to identifying moderators of response, it will be important for researchers to develop systems for identifying the most critical teacher behaviors for effectively managing the broad range of student behavior (e.g., Owens et al., in press). As noted previously, there may be

important teacher behaviors that we did not observe that changed in response to consultation and accounted for the large improvements in competence ratings.

Lastly, our data highlight the critical need for greater attention to workforce issues, both for educators and school mental health professionals who may serve as consultants. The data reaffirm the call to action to transform teacher preparation programs (NCATE, 2010), including enhancement of the practical training in classroom behavior management practices. Because stress related to disruptive student behavior is one of the top two reasons why early career teachers leave the field within the first five years (Ingersol, 2001), innovative training approaches could have a significant impact on both student and teacher outcomes. Similarly, given the ineffectiveness of one-time didactic in-service training for teachers (Blank et al., 2008), policies are needed to ensure that professional development for teachers takes a multi-component approach that aligns with theories for adult learning and behavior change (Stuart et al., 2004).

Relatedly, the effectiveness of problem-solving consultation, coupled with the likely enhancements that may be available in the future, underscore the need to prepare school mental health providers (i.e., school psychologists, school counselors, school social workers) to be competent to provide such consultative services. Unfortunately, data from recent surveys suggest that this is an underutilized practice by many school-based practitioners (Mixon, Owens & Holdaway, 2015).

Our study's contributions must be considered in the context of limitations. First, because this was a development project, we entered into the endeavor with an underpowered design (IES, 2016). Thus, support for our theory of change relies on interpretation of trends rather than statistical significance. Clearly, a large clinical trial is needed before strong conclusions about the effectiveness of the multi-component condition can be made. Second, although we selected

measures of knowledge, skills, and beliefs that have acceptable psychometric properties, additional research is needed to establish the psychometric strength of these constructs. Indeed, the reliability of the locus of control measure with this sample was lower than expected given the findings of prior studies.

Third, our primary outcome variables are based on observation, the inter-rater reliability of which was stronger for frequency behaviors than for ratings of global competence (which is likely more susceptible to observer bias). Likewise, to maintain consistency across teachers, we chose to evaluate all classroom rules for all teachers, even if they acknowledged that some rules did not apply in their classroom. In addition, the classroom activity taking place during the observation was not coded. Future research could examine the impacts of coding classroom activity and rules individualized to each teacher on outcomes.

Fourth, some readers may be concerned that the benefits derived by the Latent Class 2, Low Baseline group may reflect regression to mean. However, there was a differential response to condition for the two low baseline groups (suggesting an effect of our condition beyond regression to the mean) and the high baseline groups of Latent Class 2 also showed benefit and did not reach a ceiling effect. Fifth, the quadratic trends showing initial improvement followed by a decline while still receiving the consultation is puzzling and other variables (e.g., teacher stress relative to state testing time) need to be considered to fully understand this phenomenon. This study was focused on teacher-level factors; additional measurement of practical barriers and contextual supports would likely provide a more comprehensive picture of impacts over time. Indeed, it is possible that until school-based accountability systems for high quality implementation of classroom management are present, teacher implementation may be highly susceptible to the impact of external factors. Sixth, it is a limitation that inter-rater reliability has

not been established for our consultation integrity coding manual. Lastly, given the focus of this study (i.e., change in teacher behaviors), we did not include the impact of change in teacher behavior on change in student behavior. Although there is evidence to expect this relationship within best practice consultation (e.g., see Noell & Gansle, 2014 for review), this relationship should be tested in future studies.

Overall, this study generated findings that refine our hypotheses about methods for tailoring consultation approaches to the characteristics of individual teachers. For teachers with implementation barriers related to knowledge, skills, and beliefs, it may be best to include strategies to address these obstacles in addition to problem solving and performance feedback techniques. Continuing to identify techniques for modifying these and other obstacles to implementation of best practices for classroom management has the potential to make dramatic differences in the success experienced by many students with and without behavior problems.

Compliance with Ethical Standards: We have no conflicts of interest to disclose. 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 was obtained from all individual participants included in the study.

Funding: 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

Aarons, G. A., Hurlburt, M., & Horwitz, S. M. (2011). Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Administration and Policy in Mental Health and Mental Health Services Research, 38*(1), 4-23.

Allinder, R. M., & Oats, R. G. (1997). Effects of acceptability on teachers' implementation of curriculum-based measurement and student achievement in mathematics computation. *Remedial and Special Education, 18*, 113-120.

Amrhein, P.C., Miller, W.R., Yahne, C.E., Palmer, M., & Fulcher, L. (2003). Client commitment language during motivational interviewing predicts drug use outcomes. *Journal of Consulting and Clinical Psychology, 71,* 862–878.

Beidas, R. S., & Kendall, P. C. (2010). Training therapists in evidence-based practice: A critical review of studies from a systems-contextual perspective. *Clinical Psychology: Science and Practice, 17*, 1-30. doi: 10.1111/j.1468-2850.2009.01187.x

Blank, L., Peters, J., Pickvance, S., Wilford, J., & Macdonald, E. (2008). A systematic review of the factors which predict return to work for people suffering episodes of poor mental health. *Journal of Occupational Rehabilitation, 18,* 27–34.

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

Cook, C. R., Lyon, A. R., Kubergovic, D., Wright, D. B., & Zhang, Y. (2015). A supportive beliefs intervention to facilitate the implementation of evidence-based practices within a multi-tiered system of supports. *School Mental Health, 7,* 49-60.

DuPaul, G., D., Power, T. J., Evans, S. W., Mautone, J. A, & Owens, J. S. (2016). Students with ADHD and Section 504 Regulations: Challenges, Obligations, and Opportunities for School Psychologists. *Communiqué, 45(3)* 1, 26-67.

Epstein, M., Atkins, M., Culinan, D., Kutash, K., Weaver, R. (2008). *Reducing behavior problems in the elementary school classroom.* IES Practice Guide. (NCEE 2008-012) U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Evans, S. W., Owens, J. S., & Bunford, N. (2014). Evidence-based psychosocial treatments for children and adolescents with attention-deficit/hyperactivity disorder. *Journal of Clinical Child and Adolescent Psychology, 43,* 527-551.

Fabiano, G.A., Pelham, W.E., Majumdar, A., Evans, S.W., Manos, M.J., Caserta, D., Girio, E.L., Pisecco, S., Hannah, J.N. & Carter, R.L. (2013). Elementary and middle school teacher perceptions of attention-deficit/hyperactivity disorder incidence. *Child and Youth Care Forum, 42,* 87-99.

Fabiano, G. A., Pelham, Jr, W. E., 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 and Adolescent Psychology*, *35*(3), 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.

Fabiano, G. A., Vujnovic, R. K., Waschbusch, D. A., Yu, J. Mashtare, T. Pariseau, M. E. . . . Smalls, K. J. (2013). A comparison of workshop training versus intensive, experiential training for improving behavior support skills in early educators, *Early Childhood Research Quarterly*, *28*(2), 450-460.

Feldman, E. S., & Kratochwill, T. R. (2003). Problem solving consultation in schools: Past, present, and future directions. *The Behavior Analyst Today, 4*(3), 318.

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 Edition (pp. 18-39).* Routledge, New York.

Frey, A. J., Lee, J., Small, J. W., Seeley, J. R., Walker, H. M., & Feil, E. G. (2013). Transporting motivational interviewing to school settings to improve the engagement and fidelity of tier 2 interventions. *Journal of Applied School Psychology*, *29*(2), 183-202.

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*(4), 221-226.

Fuchs, D. & Fuchs, L. S. (1996). Consultation as a technology and the politics of school reform:

reaction to the issue. *remedial and special education*, *17*, 386–392.

Gable, R. A., Hester, P. H., Rock, M. L., & Hughes, K. G. (2009). Back to basics: Rules, praise, ignoring and reprimands revisited. *Intervention in School and Clinic, 44,* 195–205.

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.

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.

Han, S., & Weiss, B. (2005). Sustainability of teacher implementation of school-based mental health programs. *Journal of Abnormal Child Psychology, 33,* 665-679.

Harris, P.A., Taylor, R., Thielke, R., Payne, J, Gonzalez, N. & Conde, J.G. (2009) Research electronic data capture (REDCap) – a metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Information, 42,* 377-381.

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.

Ingersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal*, *38*(3), 499-534.

Institute of Educational Sciences (2016). *Request for applications.* Retrieved February, 2016, from http://ies.ed.gov/funding/pdf/2016_84324A.pdf.

Kealey, K. A., Peterson, A. V., Gaul, M. A., & Dinh, K. T. (2000). Teacher training as a behavior change process: principles and results from a longitudinal study. *Health*

Education & Behavior, *27*(1), 64-81.

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.

Martinussen, R., Tannock, R., & Chaban, P. (2011). Teachers' reported use of instructional and behavior management practices for students with behavior problems: Relationship to role and level of training in ADHD. *Child & Youth Care Forum*, *40*(3), 193-210.

McGraw, K.O. & Wong, S.P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, *1*(1), 30-46.

Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine, 65,* S63–S67.

Miller, W. R., & Rollnick, S. (2013). Motivational interviewing: helping people change. New York. *NY: Guilford*.

Mixon, C. S., Owens, J. S., Holdaway, A. S. (2015). *Exploring the job-related activities and services provided by school-based mental health professionals*. Poster presented at the annual Conference on Advancing School Mental Health, New Orleans, LA.

National Council for Accreditation of Teacher Education (NCATE). (2010). *Transforming teacher education through clinical practice: A national strategy to prepare effective teachers.* Retrieved on November, 30 2016 from http://www.ncate.org/LinkClick.aspx?fileticket=zzeiB1OoqPk%3D&tabid=715

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 Edition (pp. 386-408).* Routledge, New York.

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., 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. (in press). Rates of common classroom behavior management strategies and their associations with challenging student behavior in elementary school. *Journal of Emotional and Behavioral Disorders.*

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. doi: 10.1080/15374410801955912

Parsad, B., Lewis, L., & Farris, E. (2001). *Teacher preparation and professional development: 2000* (NCES 2001-088). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

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

Reinke, W. M., Herman, K. C., & Sprick, R. (2011). *Motivational interviewing for effective classroom management: The classroom check-up*. Guilford Press.

Robb, J. A., Sibley, M. H., Pelham Jr, W. E., 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*(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*(3), 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 behavior support plans. *Psychology in the Schools, 51,* 879–895. doi: 10.1002/pits.21787

Sanetti, L. M. H., & Kratochwill, T. R. (2009). Treatment integrity assessment in the schools: An evaluation of the Treatment Integrity Planning Protocol. *School Psychology Quarterly*, *24*, 24-35.

Sheridan, S. M. (1992). Consultant and client outcomes of competency-based behavioral consultation training. *School Psychology Quarterly*, *7*(4), 245.

Shernoff, E. S., Mehta, T. G., Atkins, M. S., Torf, R., & Spencer, J. (2011). A qualitative study of the sources and impact of stress among urban teachers. *School mental health*, *3*(2), 59-69.

Shrout, P.E. & Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 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*(2), 160-175.

Stuart, G., Tondora, J., & Hoge, M. (2004). Evidence-based teaching practice: Implications for behavioral health. *Administration and Policy in Mental Health, 32,* 107–130.

Tschannen-Moran, M., Hoy, A. W., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research*, *68*(2), 202-248.

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.

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*(1), 34-46.

Vujnovic, R. K., Fabiano, G. A., Waschbusch, D. A., Pelham, W. E., Greiner, A., Gera, S., ... & Buck, M. (2014). Preliminary psychometric properties of an observation system to assess teachers' use of effective behavior support strategies in preschool classrooms. *Education and Treatment of Children*, *37*(2), 323-346.

Wang, M. C., Haertel, G. D., & Walhberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research, 63,* 249–294.

Wechsler, D. (2011). *Wechsler abbreviated scale of intelligence (2nd ed.).* Bloomington, MN. Pearson.

Wilder, D. A., Atwell, J., & Wine, B. (2006). The effects of varying levels of treatment integrity on child compliance during treatment with a three-step prompting procedure. *Journal of Applied Behavior Analysis*, *39*(3), 369-373. doi: 10.1901/jaba.2006.144-05

Table 1

Components for the Comparison and Multi-component Consultation Conditions

	Comparison Consultation	Multi-component Consultation
Initial Workshop (ADHD, GCM, DRC)	X	X
Classroom Management Interview	X	X (+ Values Interview)
Target Behavior Interview	X	X
Baseline Tracking	X	X
DRC Development Meeting	X	X
DRC Launch Meeting	X	X
Biweekly Problem Solving	X	X
Weekly Integrity Observations	X	X
Biweekly Performance Feedback	Brief	Enhanced
Teacher Values Interview		X
MI-Informed Strategies		X
Cognitive Behavioral Strategies		X
Knowledge Component		X
Beliefs Component		X
Skills Component		X

Note. GCM = general classroom management; Knowledge Component involved *News You Can Use* fact sheets; Beliefs Component involved identification and modification of teacher beliefs; Skills Component involved role plays and skills practice. Brief = limited to 5 to 10 minutes and unless the teacher *initiated* discussion of other content, the discussion and problem solving remained child-focused; Enhanced = comprehensive review of child and teacher behaviors from the observations, highlighting connections between teacher integrity behaviors and child outcomes to facilitate either skills practice or a discussion of related beliefs. MI = motivational interviewing.

Table 2

Fit Statistics Associated with the Two-, Three-, and Four-Latent Class Models

Model	N	Class Probability	BIC[a]	Entropy[b]	LMR LRT p-value[c]	VLMR LRT p-value[c]	BLRT p-value[c]
One-Class	56		1091.89				
Two-Class			1079.61	0.76	0.05	0.04	< 0.01
Class 1	27	0.96					
Class 2	29	0.92					
Three-Class			1084.31	0.83	0.30	0.28	0.14
Class 1	4	0.99					
Class 2	27	0.95					
Class 3	25	0.90					
Four-Class			1092.52	0.79	0.33	0.31	0.60
Class 1	8	0.85					
Class 2	26	0.92					
Class 3	5	0.97					
Class 4	17	0.78					

Note. BIC = Bayesian Information Criteria; LMR LRT = Lo-Mendell-Rubin Likelihood Ratio Test; VLMR LRT = Vuong-Lo-Mendel-Rubin Likelihood Ratio Test; BLRT = bootstrap Likelihood Ratio Test. [a]In interpreting fit, a lower number represent a better fit for the BIC; [b]Entropy of > 0.8 is preferable; [c]A significant p-value in LRT tests signifies that the (k-1) class model is rejected in favor of the k-class model.

Table 3

Descriptive Data on Knowledge, Skills, and Beliefs for Teachers in Latent Class 1 and 2.

Variable	Class 1[a] *M (SD)*	Class 2[b] (Target Population) *M (SD)*	Effect Size[c]
Knowledge of ADHD	90.64 (7.20)	74.77 (10.08)	1.81
Knowledge of behavior modification	67.13 (8.24)	50.86 (12.36)	1.55
Locus of control for student failure	11.41 (2.00)	5.54 (2.06)	2.89

Note. [a]N = 27; [b]N = 29; [c]Positive Cohen's d effect sizes represent higher scores for Class 1 relative to Class 2. The Target Population is the group of teachers hypothesized to demonstrate greater benefits from the multi-component condition than the comparison condition.

Table 4

Baseline, Slope and End Point Values for Dependent Variables by Condition and Latent Class for Low Baseline Teachers

Variable	MCC[a] Latent Class 1		CC[b] Latent Class 1		MCC[c] Latent Class 2		CC[d] Latent Class 2	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Responded Appropriately - Target Child								
Baseline	0.12	0.09	0.06	0.07	0.14	0.05	0.12	0.04
Slope	0.03	0.08	0.10	0.26	0.09	0.13	0.03	0.06
End Point	0.21	0.23	0.17	0.25	0.40	0.29	0.17	0.25
Responded Appropriately - Other Child								
Baseline	0.22	0.10	0.21	0.09	0.28	0.10	0.22	0.07
Slope	0.02	0.07	0.00	0.02	0.14	0.17	0.04	0.05
End Point	0.35	0.26	0.27	0.14	0.68	0.25	0.39	0.15
Rate Labeled Praise								
Baseline	2.64	2.50	4.03	2.43	4.81	2.22	4.47	3.34
Slope	4.49	8.73	3.95	3.62	4.26	6.19	2.11	6.83
End Point	6.92	5.23	16.07	10.02	13.14	8.20	10.57	22.20
Global Competence - Class Management								
Baseline	5.32	0.98	5.88	0.52	5.01	1.03	5.36	0.84
Slope	0.45	0.31	0.23	0.20	1.04	0.98	0.23	0.28
End Point	6.64	0.94	6.90	0.77	7.31	1.99	6.24	1.12

Note. MCC = Multi-component Consultation Condition; CC = Comparison Condition. Slopes indicate that for each biweekly session, teachers changed by one unit of the variable (e.g., percent of appropriate responses, labeled praises per hour, score on the global competence subscale).

[a]*N*s range from 5 to 10; [b]*N*s range from 3 to 7; [c]*N*s range from 7 to 9; [d]*N*s range from 6 to 11.

Table 5

Baseline, Slope and End Point for Dependent Variables by Condition and Latent Class for High Baseline Teachers

Variable	MCC[a] Latent Class 1		CC[b] Latent Class 1		MCC[c] Latent Class 2		CC[d] Latent Class 2	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Responded Appropriately - Target Child								
Baseline	0.33	0.05	0.41	0.16	0.44	0.15	0.38	0.12
Slope	0.02	0.05	0.03	0.06	0.04	0.05	0.04	0.12
End Point	0.51	0.41	0.55	0.45	0.58	0.36	0.55	0.47
Responded Appropriately - Other Child								
Baseline	0.47	0.09	0.47	0.12	0.62	0.22	0.57	0.12
Slope	0.04	0.07	0.05	0.10	0.06	0.03	0.06	0.07
End Point	0.63	0.22	0.67	0.29	0.79	0.24	0.72	0.32
Rate Labeled Praise								
Baseline	19.06	9.10	20.23	9.81	30.55	19.92	17.99	10.52
Slope	5.41	6.93	2.34	0.88	10.34	23.47	3.32	18.69
End Point	24.44	9.10	22.11	6.56	40.59	34.01	9.63	8.28
Global Competence - Class Management								
Baseline	7.63	0.35	7.44	0.58	7.66	0.68	7.60	0.79
Slope	0.23	0.18	0.32	0.32	0.07	0.19	-0.12	0.53
End Point	8.39	1.25	8.10	1.23	7.92	1.24	7.75	1.98

Note. MCC = Multi-component Consultation Condition; CC = Comparison Condition.

[a]*N*s range from 6 to 11; [b]*N*s range from 4 to 8; [c]*N*s range from 5 to 7; [d]*N*s range from 4 to 9.

Table 6

Effect Sizes Representing Benefits of the Multi-Component Condition Over the Comparison Condition By Latent Class and Baseline Groups

Variable	LC 1 Low Baseline	LC 1 High Baseline	**LC 2 Low Baseline**	LC 2 High Baseline
Responded Appropriately - Target Child	-0.39	-0.23	**0.51**	-0.03
Responded Appropriately - Other Child	0.29	-0.17	**0.82**	-0.03
Rate of Labeled Praise	0.08	0.53	**0.33**	0.32
Global Competence - Class Management	0.82	-0.35	**1.12**	0.46
Average Across All Variables	0.20	-0.06	**0.69**	0.18

Note. Boldface indicates the group hypothesized to benefit most from multi-component condition, relative to the comparison condition. Within each latent class, median splits were performed on baseline levels of each dependent variable. Therefore, participants may change baseline groups (e.g., low or high baseline) depending on dependent variable selected. LC = Latent Class; Effect sizes were derived from the difference between the slope in multi-component condition and the slope in comparison condition divided by the pooled standard deviation. Positive values indicate an advantage for the multi-component condition; negative values indicate an advantage for the comparison condition.

Table 7. *Teacher Satisfaction Items by Treatment Condition and Latent Class and Baseline Status**

	Multi-Component Condition				Comparison Condition			
Satisfaction Item	LC 1 Low Baseline	LC 1 High Baseline	LC 2 Low Baseline	LC 2 High Baseline	LC 1 Low Baseline	LC 1 High Baseline	LC 2 Low Baseline	LC 2 High Baseline
The bi-weekly consultations were...[1]	5.00 (0.0)	4.72 (.49)	4.57 (.54)	5.00 (0.0)	5.00 (0.0)	4.60 (.55)	4.33 (.58)	4.57 (.54)
Regarding my facilitator, I found him/her...[1]	5.00 (0.0)	4.86 (.39)	5.00 (0.0)	5.00 (0.0)	5.00 (0.0)	5.00 (0.0)	4.67 (.58)	5.00 (0.0)
Regarding discipline techniques, I feel I have learned...[2]	4.56 (.53)	4.14 (.69)	4.00 (1.16)	4.50 (.58)	4.25 (.96)	3.00 (1.00)	3.33 (1.53)	3.71 (.76)
Regarding the progress my student has made in his/her behavior, I am...[3]	5.00 (.00)	4.29 (.49)	4.14 (1.22)	3.75 (1.89)	4.75 (.50)	4.20 (.84)	4.67 (.58)	3.43 (1.72)
Regarding the progress my student has made in his/her academics, I am …[3]	4.67 (.50)	3.57 (1.40)	3.29 (1.60)	3.50 (1.73)	4.00 (.82)	4.00 (1.23)	3.33 (1.16)	2.71 (1.60)
Thinking about the overall program, how much did your student benefit?[4]	1.11 (.33)	1.57 (.79)	2.43 (1.13)	1.5 (1.00)	1.00 (0.0)	2.00 (.71)	2.00 (1.00)	2.43 (1.13)
Thinking about the overall program, how much did you benefit?[4]	1.22 (.44)	1.86 (.69)	2.43 (1.13)	1.50 (1.00)	1.00 (.00)	2.40 (.89)	1.33 (.58)	2.29 (.76)
Would you use this intervention with other students if you had the chance?[5]	4.67 (.50)	4.57 (.54)	3.83 (.98)	4.75 (.50)	5.00 (0.0)	4.00 (1.00)	4.67 (.58)	4.00 (.58)
Would you recommend the program to other teachers?[5]	4.89 (.33)	4.57 (.54)	3.83 (.98)	4.75 (.50)	5.00 (.00)	3.80 (1.30)	4.00 (1.00)	3.71 (.76)

Note: *Baseline status is based on the median split for Percent Appropriate Response to Target Child Rule Violations. [1]Response options included: *Not helpful at all (1), Somewhat unhelpful (2), Neither helpful or unhelpful (3), Somewhat helpful (4), Very helpful (5), Prefer not to answer*. [2]Response options included: *Nothing useful (1), Very little (2), A few new techniques (3), Several useful techniques (4), and very many techniques (5).* [3]Response options included*: very dissatisfied (1), somewhat dissatisfied (2), neutral (3), somewhat satisfied (4), and very satisfied (5).* [4]Response options included*: very much (1), much (2) somewhat (3) and (4) not at all;* [5]Response options included*: definitely not (1), probably not (2), not sure (3), probably yes (4), definitely yes (5).*

Figure 1. *Improvements in the Appropriate Reponse to Rule Violations of Target Child and Other Children Among Latent Class 2, High and Low Baseline Teacher Recievig Multi-Component Consultation*