An Empirical Study of Multidimensional Fidelity of COMPASS Consultation

Venus Wong,
Department of Educational, School, and Counseling Psychology, University of Kentucky

Lisa A. Ruble,
Department of Educational, School, and Counseling Psychology, University of Kentucky

John H. McGrew, and
Department of Psychology, Indiana University-Purdue University Indianapolis

Yue Yu
Department of Psychology, Indiana University-Purdue University Indianapolis

Abstract

Consultation is essential to the daily practice of school psychologists (NASP, 2010). Successful consultation requires fidelity at both the consultant (implementation) and consultee (intervention) levels. We applied a multidimensional, multilevel conception of fidelity (Dunst et al., 2013) to a consultative intervention called the Collaborative Model for Promoting Competence and Success (COMPASS) for students with autism. The study provided three main findings. First, multidimensional, multilevel fidelity is a stable construct and increases over time with consultation support. Second, mediation analyses revealed that implementation-level fidelity components had distant, indirect effects on student IEP outcomes. Third, three fidelity components correlated with IEP outcomes: teacher coaching responsiveness at the implementation level, and teacher quality of delivery and student responsiveness at the intervention levels. Implications and future directions are discussed.

School consultation is an evidence-based approach for improving teaching quality and student learning outcomes (Kampwirth & Powers, 2012; Sanetti & Kratochwill, 2007) and is a required core competency in accredited school psychology programs (NASP, 2010). Despite evidence of effectiveness, consultation practices vary widely and little is known about applications in everyday practice (Gresham & Lopez, 1996). One critical factor that influences success for any evidence-based intervention is fidelity - the degree to which the intervention is delivered as originally intended (Power et al., 2005). Consultation fidelity addresses both what the consultant does during the consultation and what the teacher or consultee does as a result of the consultative process (Miller & Rollnick, 2014; Noell, 2008; Sanetti & Kratochwill, 2007; Sanetti & Kratochwill, 2009). The implementation science model outlined by Dunst, Trivette, and Raab (2013) and the multilevel consultation framework by Noell (2008) provide frameworks for assessing these two levels of fidelity -

implementation fidelity (consultant level) and intervention fidelity (consultee level) (Dunst et al., 2013). Implementation fidelity captures the consultant's behavior or the degree to which the consultation is implemented as designed, whereas intervention fidelity focuses on the consultee's behavior or the degree to which the intervention developed during the consultation is implemented as planned. Assessment of both implementation and intervention fidelity is critical to capture delivery of the necessary components of an effective program that impacts consultee and client outcomes (Durlak & DuPre, 2008; Ruble & McGrew, 2015; Schoenwald, Sheidow, & Letourneau, 2004). Dunst and colleagues (2013) provided some examples with regard to the key features of quality implementation and intervention fidelity in the area of early intervention. For instance, essential implementation fidelity includes active learners' participation in using evidence based practices and quality coaching feedback for performance, whereas intervention fidelity includes reciprocal child-teacher interactions and contingent responsiveness to child's learning. Dunst and colleagues (2013) believed that implementation fidelity should influence intervention fidelity and thus treatment outcomes. Specifically, they posited that intervention fidelity would fully or partially mediate the relationships between implementation fidelity and treatment outcomes.

Despite its importance, fidelity measurement for school consultation is limited (Noell, 2008; O'Donnell, 2008; Sheridan et al., 2009). Unfortunately, the failure to assess fidelity is not isolated to school consultation research and practice. Several studies have documented the relatively small percentage of research and practice-based articles reporting intervention fidelity (Schulte, Easton, & Parker, 2009; Weisz, Jensen-Doss, & Hawley, 2006). For instance, half of the intervention studies published in four school psychology journals between 1995 and 2008 failed to report intervention fidelity data (Sanetti & Fallon, 2011). Another recent study reported that fewer than 50% of the intervention studies published in general and special education journals provided information about the fidelity of implementation (Swanson, Wanzek, Haring, Ciullo, & McCulley, 2013). Even lower percentages have been reported within particular special education eligibility groups. For example, in randomized controlled treatment studies of autism, treatment fidelity data appear in less than 18% of studies (see Perepletchikova, Treat, & Kazdin, 2007; Wheeler, Baggett, Fox, & Blevins, 2006).

One of the challenges in fidelity assessment is inconsistent and unclear use of definitions. Fidelity has been defined differently, and its meaning has evolved across time from a unidimensional to a multidimensional concept (Power et al., 2005). The traditional unidimensional definition of fidelity focused solely on the extent to which an intervention was delivered as planned (Gresham, 1989). More recently, researchers have suggested that fidelity should be studied in terms of quantity (i.e., adherence and dosage) and quality (i.e., quality of delivery and participant responsiveness) (Dusenbury, Brannigan, Falco, & Hansen, 2003; O'Donnell, 2008). For example, adherence has been defined as the extent to which program components are implemented; and dosage or exposure as the amount of the intervention delivered to participants, such as length and frequency of sessions (Dane & Schneider, 1998). Quality of delivery, in turn, measures interventionists' skills in delivering the program (e.g., treatment alliance, positive regard), and participant responsiveness measures participants' engagement during intervention sessions, such as attention and

participation. Thus, over time, the construct of fidelity has expanded from a narrow focus on adherence to a multidimensional concept of quantity and quality.

A multilevel, multidimensional measurement approach is rarely applied in school-based consultation (Sheridan et al. 2009) in part because measuring multidimensional fidelity is time consuming and complex, requiring psychometrically sound tools (Noell, 2008; Schulte el al., 2009). However, a few researchers have used multidimensional fidelity. Reinke and colleagues (2013) evaluated the multidimensional fidelity of a teacher's classroom management intervention and found that ratings of participant responsiveness and coaching dosage correlated with important teacher- and student-level treatment outcomes. Also, Odom and colleagues (2010) and Hamre and colleagues (2010) studied the dosage and quality of delivery of early education programs and found that both were related to student outcomes. Crawford, Carpenter, Wilson, Schmeister, and McDonald (2012) found that teacher adherence, student engagement, and total time in a math intervention predicted students' math outcomes. Finally, Wanless and colleagues (2014) used mediational analyses and found that teachers' engagement in training in a social-emotional intervention mediated the relationship between observational ratings of classroom climate and teacher adherence.

These studies have advanced our understanding of the interactions among fidelity components (Hoagwood, Atkins, & Ialongo, 2013); however, other than these few studies, school-based consultation studies have failed to utilize multilevel, multidimensional fidelity to monitor the consultative process and outcomes.

COMPASS and the Multilevel, Multidimensional Fidelity Model

The current study builds on Ruble and colleagues (2012)'s work on the Collaborative Model for Promoting Competence and Success (COMPASS), a manualized teacher-parent consultation intervention for students with autism spectrum disorder (ASD) (Ruble, Dalrymple, & McGrew, 2012). COMPASS was tested in two RCTs (Ruble, Dalrymple, & McGrew, 2010; Ruble, McGrew, Toland, Dalrymple, & Jung, 2013) that provided opportunities to understand multilevel, multidimensional fidelity as both implementation fidelity (consultant implementation of COMPASS) and intervention fidelity (teacher implementation of COMPASS intervention plans). COMPASS consists of five sessions: an initial 3-hr consultation that includes the teacher and parent, and four, 1-hr, subsequent coaching sessions. In the initial 3-hr consultation, the consultant, teacher, and parent used the COMPASS decision-making approach to engage in an open discussion of parent and teacher concerns and observations related to the child's strengths, preferences, fears, frustrations, adaptive skills, problem behaviors, social skills, communication skills, and learning skills to identify IEP goals related to social skills, communication, and independence/learning skills within an ecological framework. After personalized goals were identified, the team developed intervention plans using evidence based strategies that were selected, and when necessary adapted, based on the child's learning strengths/preferences and environmental resources / supports and the classroom context using an Evidence Based Practice in Psychology Framework (McGrew, Ruble, & Smith, 2016). Following the consultation, teachers were asked to update the students' IEPs with the new goals. The consultant then met with the teacher for four subsequent (1-hr) coaching sessions spread

evenly throughout the remaining school year (occurring every 4–6 weeks). All consultations for the study were led by one of two consultants, the co-developers of COMPASS (Ruble & Dalrymple, 2002), who split the case load. The consultants were highly experienced with more than 25 years of experience. They were both female and Caucasian. The total time spent consulting with the teacher was less than 10 hrs over the school year. A large effect size was obtained for the first RCT ($d = 1.5$; Ruble, Dalrymple, & McGrew, 2010). This was replicated in a second RCT that included both face-to-face and a web-based teacher coaching condition ($d = 1.4$; 1.1 respectively; Ruble, McGrew, Toland, Dalrymple, & Jung, 2013.

Although parents are an integral part of COMPASS, they are not the main consultees. That is, only teachers are required to attend the coaching sessions and execute the plans developed after consultation. Based on the implementation frameworks by Dunst and colleagues (2013) and Dane and Schneider (1998), we assessed four fidelity components at both the implementation and intervention levels: adherence, quality of delivery, dosage, and participant responsiveness (Figure 1).

Treatment fidelity is essential to the success of COMPASS. At the implementation level, high fidelity implies that COMPASS-specific components and generally good consultation practices are implemented by the consultant and received by the consultee or teacher. At the intervention level, this means that evidence-based practices recommended by the National Professional Development Center on Autism Spectrum Disorder (see Wong, 2015) individualized for students with ASD (McGrew, Ruble, & Smith, 2016) were implemented by the teacher and received by the student. Thus, fidelity at the intervention level directly links to fidelity at the implementation level and positively influences students' learning outcomes (National Professional Development Center on Autism Spectrum Disorder, n.d.; Wheeler, Baggett, Fox, & Blevins, 2006).

Research Questions

Because we lack knowledge of multidimensional fidelity reliability, construct validity, and predictive validity, the study aimed to explore the nature of and relationships among multidimensional fidelity components. We asked four basic research questions: (1) Are measures of implementation- and intervention-level fidelity consistent across time?; (2) Do measures of implementation- and intervention-level fidelity correlate with each other?; (3) Are measures of implementation- and intervention-level fidelity associated with student outcomes?; and (4) Do measures of intervention-level fidelity mediate the relationship between implementation-level fidelity and student outcomes? The idea of intervention-level variables as mediators is commonly accepted and found in other conceptual models (see Noell, 2008; Sheridan, Rispoli, & Holmes, 2013). Based on previous empirical work (Odom et al., 2010; Reinke et al., 2013) and the general assumption that measures of similar constructs are likely to be positively related (Campbell & Fiske, 1959), we expected that the four fidelity components would be positively correlated with each other at both the implementation and intervention levels, with correlations strongest within, as opposed to, across levels. We also expected that implementation-level fidelity would have a weaker,

indirect association with student outcomes mediated through intervention-level fidelity, which would have a direct and stronger association with student outcomes.

Methods

Participants

The data come from a secondary analysis of two RCTs of COMPASS (Ruble et al., 2010; Ruble et al., 2013). The same inclusion eligibility, recruitment procedures, and group assignment procedures were used in both studies. Across the two studies, a total of 79 special education teachers and one student with autism selected randomly from the teacher's caseload were recruited. That is, if a teacher had more than one student with ASD, random selection was used to select one child with ASD from her caseload as the student participant. Forty-seven dyads (i.e., one teacher and one student with ASD) were assigned to the COMPASS experimental groups (18 for the first study and 29 from the second study). See Ruble et al., (2010) and Ruble et al., (2013) for a more detailed description of the recruitment and randomization process. The Autism Diagnostic Observation Schedule (Lord et al., 2000) was used to verify the ASD diagnosis. See Table 1 for demographic information.

Implementation Level Measures

Consultation and coaching adherence—Two scales were used. A 25-item Consultant Consultation Adherence Form (CCAF) measured the degree to which critical elements of the initial consultation were delivered. Teachers rated items using a dichotomous scale (Yes/No; Kuder–Richardson Formula 20 [KR20] >.80). Sample items included "goals include those suggested from home and family" and "treatment goals that came from the COMPASS consultation are described in clear behavioral terms." The 14-item Consultant Coaching Adherence Form (CCoAF) assessed the critical elements of a quality coaching session. Teachers rated items on a 4-point Likert-type scale (1=not at all; 4=very much; a = .70). Sample items included "we reviewed the consultation/coaching written summary report and answered questions" and "we discussed at least one idea (what teaching methods to keep in place or what teaching methods to consider changing) for each objective."

Consultation and coaching quality of delivery—Two scales were used. The 8-item Consultation Quality of Delivery Questionnaire (CQDQ) captured quality of delivery of the initial consultation, e.g., consultant enthusiasm, attitude, communication skills and professionalism. Teachers rated items using a 4-point Likert scale (1=not at all; 4=very much). Sample items included "the consultant's communication skills were effective" and "the consultant listened to what I had to say." The CQDQ demonstrated excellent consistency (a = .99). The 4-item Coaching Quality of Delivery Questionnaire (CoQDQ) was similar to the CQDQ, assessing similar features focused on coaching sessions, and was teacher-rated using a 4-point Likert scale (1=not at all; 4=very much). The CoQDQ had good internal consistency across the coaching sessions (a =.75 –.99).

Teacher responsiveness—Consultants used a 5-point Likert scale (1=not very much; 4=very much) to rate teacher involvement at initial consultation and across coaching

sessions using the 9-item Consultation/Coaching Impression Scale (CIS). The CIS demonstrated good internal consistency across initial consultation and coaching sessions ($a= .74–.91$). Sample items included "was the teacher involved in discussion?" and "was the teacher cooperative / collaborative?"

Intervention Level Measures

Teacher adherence—Teacher adherence assessed whether the teaching plans from the COMPASS consultation and coaching sessions were delivered as planned (Fogarty et al., 2014). Immediately following each coaching session, the two consultants completed a simple one-item, 5-point ($1 = 0 – 19\%$; $2 = 20 – 39\%$; $3 = 40 – 59\%$; $4 = 60 – 79\%$; $5 = 80 – 100\%$) scale rating the degree (percentage) to which teachers had implemented the teaching plans. The two consultants independently rated 45% of the coaching sessions for reliability. Interrater reliability using intraclass correlation was good (ICC; $r = .90$, $p < .05$).

Teacher quality of delivery—The quality of implementation of the intervention was assessed using the Social Interaction Rating Scale (SIRS; Ruble & McGrew, 2013; Ruble, McDuffle, King, & Lorenz, 2008) which measures teaching quality of delivery across six domains of teacher behavior, e.g., teacher's level of affect. Consultants rated items using a 5-point Likert scale. The SIRS demonstrated good internal consistency ($a = .89$). Two consultants independently rated 20% of the SIRS across sessions, demonstrating good interrater reliability using ICC ($r = .94$, $p < .01$).

Student responsiveness—The level of student engagement and reaction to the intervention was assessed using the Autism Engagement Scale (AES; Ruble & McGrew, 2013). Student engagement was rated along six domains of behavior: (a) cooperation; (b) functional use of objects; (c) productivity; (d) independence; (e) consistency of the child's and the teacher's goals; and (f) attention to the activity. Items were rated on a 5-point Likert scale. The total score was used. The AES demonstrated good overall internal consistency ($a = .86$). In addition, the two consultants independently rated 20% of the AES across sessions, obtaining good interrater reliability (ICC = 0.88, $p < .01$).

Intervention Dosage—Teacher report of weekly frequency of teaching the targeted skills and goals (i.e., "how many times a week is the skill worked on") was used as the measure of intervention dosage.

Student Outcome Measure

Student goal attainment—Psychometric Equivalence Tested Goal Attainment Scale (PET-GAS) was used to measure student learning outcomes (i.e., IEP outcomes; Ruble, McGrew, & Toland, 2012). PET-GAS was developed following the consultation and guided performance feedback for the coaching sessions. Goal attainment scaling is considered standard practice for school consultation studies (Sheridan et al., 2006; Sladeczek et al.2001) and it is especially useful when student outcomes are individualized, such as IEP goals. The use of PET-GAS addressed the limitations of some standardized measures, such as sensitivity to changes and relevance (McConachie et al., 2015).

For each student, three IEP goals representing a social, communication, and learning skill were evaluated. To enhance comparability of PET-GAS across different participants, we applied a protocol for goal writing to ensure equivalence in goal difficulty, measurability, and equidistance; Ruble et al., 2012). In addition, two raters (i.e., two doctoral-level graduate students) independently rated all of the PET-GAS goals for difficulty, measurability, and equidistance, and confirmed similarity across goals in the three areas. For study one, the ICC for agreement was .96 for measurability, .96 for equidistance, and .59 for difficulty (difficulty was controlled in the study one analyses). For study two, the ICC for agreement was 1.0 for measurability, .96 for equidistance, and .96 for difficulty, meaning that the PET-GAS goals were similar with regard to difficulty, measurability, and equidistance across groups.

The PET-GAS was made up of a 5-point scale for each goal (-2 = present levels of performance, -1 = progress, 0 = expected level of outcome, $+1$ = somewhat more than expected, $+2$ = much more than expected). The aggregated PET-GAS score across the three skill domains was used to represent the overall learning outcome. Two independent raters (i.e., graduate research assistants blind to group assignment) evaluated 39% of PET-GAS goals and achieved good interrater reliability ($r > .90$, $p < .01$).

Data Structure and Data Analysis

Data were collected at five different time points corresponding to the five sessions in COMPASS. At the implementation level, during the initial consultation, consultation adherence, teacher responsiveness, and consultant quality of delivery were collected for both studies (time 0). In addition, teacher responsiveness was obtained for both study one and two at each coaching session. However, consultant coaching adherence and consultant coaching quality of delivery for coaching sessions (times 1 to 4) were collected only from study two. The procedures and measures for both studies were identical, except two measures were added for study two only (i.e., consultant coaching adherence and consultant coaching quality of delivery for coaching sessions) (see Figure 2; Ruble & McGrew, 2015). Moreover, to lower participants' burden, only one-fifth of teacher participants (i.e., 6) were randomly selected to fill out the fidelity measures for each coaching session, with each participant completing at least one set of coaching fidelity measures. Data at this level were analyzed in aggregate only (see Figure 2).

To ensure the reliability of the measures, interrater reliability data were collected on all the consultant-rated measures at the intervention level (i.e., teacher adherence, teacher quality of delivery, and student responsiveness). First, to promote good agreement on scoring, the two consultants selected 10% of the measures and worked on initial reliability, coming to consensus on any disagreements. Second, based on the consensus scoring protocols developed in the first step, the two consultants scored at least 20% to 45% of the measures independently; the results were compared to assess interrater reliability (ICC was used for reliability).

At the intervention level, for both study one and study two, four fidelity components were consistently assessed across coaching sessions —teacher adherence, quality of delivery,

intervention dosage, and student responsiveness. Thus, data at this level were analyzed both concurrently (i.e., at each time point) and aggregately (i.e., overall mean score).

To answer the first two research questions, correlational analyses were used to understand the relationships among the variables. Because study cohort (i.e., whether teachers participated in the first or second study) may have influenced study variables, cohort was used as a control variable, and partial correlation coefficients were used in the correlational analyses. Repeated measure analyses were conducted using a linear mixed model to evaluate change in the levels of different fidelity components across the coaching sessions. A linear mixed model can handle within- and between-subject analyses simultaneously and thus is appropriate for the current data set because it is structured between cohorts (i.e., study one and two) and within individuals (i.e., four data points within an individual). To answer the third research question, regression analyses were used to model predictiveness of different fidelity components on student IEP outcomes.

To answer the fourth research question, six serial mediation analyses were performed using the PROCESS procedure within SPSS to assess whether intervention fidelity components mediated relationships between outcomes and implementation fidelity (Hayes, 2012). Serial mediation is "a causal chain linking the mediators, with a specified direction of causal flow" (Hayes, 2012, p. 14). Bootstrapping with 5000 repetitions was used to perform the mediation analyses. Given our sample size limitations, the bootstrapping PROCESS program allowed us to test our hypotheses appropriately and to do so within the power recommendations. PROCESS examines all possible variable combinations for a particular sequence of mediating variables specified by the user. In the current study, we specified the mediating variables with the following sequence: teacher adherence, teacher quality of delivery, student engagement, and dosage. This order was based on a logical and temporal sequence; teacher adherence to the COMPASS teaching plans should increase their teaching quality, which in turn should improve student engagement in learning. Student responsiveness to the intervention should then provide information enabling teachers to adjust their dosage or frequency of instruction. Group assignment was also controlled in the mediation analyses.

Results

Consistency of Fidelity across Time

In general, the fidelity measures showed evidence of consistency over time as well as a general increase over time. The results for each measure are reported below.

Teacher adherence—Teacher adherence at sessions two, three, and four were significantly correlated, $r(42) = .37–.47$, $p < .05$. A linear mixed model analysis showed that time predicted adherence ($F_{3,138} = 14.80$, $p < .01$), indicating an increase in adherence across time.

Teacher quality of delivery—Four of the six correlations of teacher quality of delivery at sessions one, two, three, and four were significantly correlated, $r(42) = .44–.60$, $p < .05$. A linear mixed model showed that time predicted quality of delivery ($F_{3,111} = 11.43$, $p < .01$), indicating an increase in quality of delivery across time.

Student responsiveness—Except for one pairwise comparison, student responsiveness at times one, two, three, and four were significantly correlated, $r(42) = .39–.60$, $p < .05$. A linear mixed model showed that time predicted participant responsiveness ($F_{3,125} = 9.10$, $p < .01$), indicating an increase in participant responsiveness across time.

Intervention dosage—Four of the six correlations of dosage at times one, two, three, and four were significantly correlated, $r(42) = .28–.76$, $p < .05$. A linear mixed model showed that time predicted dosage ($F_{3,117} = 2.97$, $p < .05$), indicating increased instruction across time.

Relationship among Fidelity Components

At the implementation level, fidelity components collected during the consultation and coaching sessions were largely uncorrelated, except teacher responsiveness during consultation was positively correlated with mean teacher responsiveness over the four coaching sessions ($r = .34$, $p < .05$). At the intervention level, teacher adherence was marginally correlated with teacher quality of delivery ($r = .29$, $p = .05$). Student responsiveness was positively correlated with teacher quality of delivery ($r = .47$, $p < .01$), but negatively correlated with dosage ($r = −.45$, $p < .05$). With respect to cross level associations, consultation quality of delivery was negatively correlated with intervention dosage ($r = −.42$, $p < .01$), teacher consultation responsiveness was positively correlated with mean teacher adherence ($r = .31$, $p < .05$), and teacher coaching responsiveness was correlated with teacher adherence ($r = .54$, $p < .01$), teacher quality of delivery ($r = .63$, $p < .01$), and student responsiveness ($r = .37$, $p < .01$) (See Table 2).

Student PET-GAS Outcome and Fidelity

At the implementation level, teacher coaching responsiveness was positively correlated with PET-GAS ($r = .39$, $p < .01$). At the intervention level, PET-GAS correlated with mean student responsiveness ($r = .52$, $p < .01$) and mean teacher quality of delivery ($r = .44$, $p < .01$).

Mediation Analyses

Both consultant consultation adherence (indirect effect = .07, $SE = .04$, 95% $CI = .02, .20$; completely standardized effect = .12) and teacher consultation responsiveness (indirect effect = .10, $SE = .11$, 95% $CI = .005, .58$; completely standardized effect = .03) had significant indirect effects on student IEP outcomes via teacher adherence, teacher quality of delivery, and student responsiveness (see Figures 3 and 4). Increased consultant adherence during consultation and increased teacher consultation responsiveness during the consultation were related to greater teacher adherence, and subsequently to teacher quality of delivery, which in turn was related to better student responsiveness in learning and then to better IEP outcomes.

Both consultant coaching adherence (indirect effect = −.81, $SE = .70$, 95% $CI = −3.51, −.08$; completely standardized effect = −.10) and teacher responsiveness during coaching sessions (indirect effect = .36, $SE = .28$, 95% $CI = .04, 1.24$; completely standardized effect = .08) had significant indirect effects on student outcomes via teacher quality of delivery and

student responsiveness. Increased consultant adherence was related to lower teacher quality of delivery, which in turn was related to poorer student responsiveness in learning, and then to poorer IEP outcomes (Figure 4). Increased teacher responsiveness during coaching sessions was related to higher teacher quality of delivery, which in turn was related to better student responsiveness, and then to better IEP outcomes (Figure 6).

Finally, there was a significant indirect effect of consultant consultation quality of delivery on student outcomes via intervention dosage at the intervention level (indirect effect = .14, SE = .11, 95% CI = .0002, .53; completely standardized effect = .08). Increased consultant quality of delivery was related to lower intervention dosage, which in turn was related to better IEP outcomes (Figure 7).

Discussion

Summary of Findings

Relatively little is known about consultation multilevel, multidimensional fidelity measurement, including the interrelationships and predictive associations with outcomes (Harn, Parisi, & Stoolmiller, 2013). The summary of findings is divided into four areas we deemed of particular importance: (1) stability of fidelity, (2) relationship among fidelity components within and across levels, (3) fidelity and outcomes, and (4) the mediating effects of intervention-level fidelity.

Stability of fidelity—The current results are somewhat inconsistent with other reports. Although Spear (2014) noted that the two common ways to capture multidimensional fidelity are a single representative score that measures one particular assessment time point or an average score of all assessment time points throughout the study, our findings show that implementation fidelity is a dynamic rather than static process. Similarly, in contrast to Zvoch (2009), who reported that implementation fidelity decreases over time, or Spear (2014) and Odom et al. (2010), who reported that implementation fidelity stays stable across time, our results suggest that multidimensional fidelity may be both relatively stable and improve across time. One possible explanation of these differences is that Zvoch (2009), Spear (2014), and Odom et al. (2010) passively monitored fidelity and did not directly intervene to alter fidelity over time. In contrast, COMPASS consultants actively monitor and intervene to improve teachers' progress through coaching sessions at different time points. The results also indicate that consultants can expect both that consultees who start out strong in fidelity to remain strong later in time and that consultees will likely improve over time and reinforce the potential advantages of consultation that is ongoing and not a one-time activity. Given the finding of improved multidimensional fidelity over time, the results also imply that analyzing data concurrently (scores at time points) or aggregately (averaged scores) may produce slightly different results in consultation research. However, existing knowledge about the stability of multidimensional fidelity is still limited and both approaches for measurement of multidimensional fidelity are suggested.

Relationships among fidelity components within and across levels—Another important contribution was the examination of relationships among fidelity components at the implementation and intervention levels. Currently, there is only a very preliminary

understanding of how fidelity components correlate with each other. For instance, Knoche, Sheridan, Edwards, and Osborn (2010) reported that adherence was significantly correlated with participant responsiveness. In contrast, our findings varied depending on the level examined. Fidelity components were largely uncorrelated at the implementation level. We suspect that these findings might be due to two reasons. First, fidelity components at the implementation level were high overall and demonstrated low variance. Second, the consultation and coaching sessions were conducted by two well-trained consultants, which might produce a lack of variance in implementation quality.

Fidelity components were more likely to be correlated at the intervention level, however. For example, teacher quality of delivery was marginally positively related to intervention adherence, echoing Knoche and colleagues (2010) results. Teachers who implemented teaching plans with adherence were rated as having better teaching quality. Thus, teachers faithful to the unique elements of COMPASS also were faithful to the common elements of good teaching practice, in general. This may indicate that teachers good in one area of teaching are likely to be good in other areas. Also at the intervention level, teacher quality of delivery was positively correlated with student responsiveness. Students tended to be more engaged and active when teaching was rated high quality, similar to some previous findings (Hamre, Pianta, Hatfield, & Jamil, 2014; Klem & Connell, 2004). Taken together these results identify potential links in a chain between consultation activities, resulting changes in teaching behavior, and impact on student responsiveness. Moreover, these results are consistent with the notion that different aspects of fidelity should be positively correlated and provide support for the potential usefulness of multiple measures of fidelity.

However, somewhat surprisingly, intervention dosage and student responsiveness were negatively correlated with each other. One possible explanation for the negative relationship is that teachers may work more intensively with less engaged students due to their lack of responsiveness to the interventions. That is, dosage, unlike the other fidelity measures may not indicate better or higher quality of implementation per se, but may instead be a proxy for student need or indicator of the teacher's sensitivity of the optimal amount of instruction. Another possible explanation is questionable validity of the measures. We employed a one-item, self-report measure to capture dosage. Although self-report is commonly used as a source of implementation information and single item measures have been used successfully previously (see Fogarty et al., 2014), there is always the danger of inadequate construct coverage when using single item measures. Thus, better measures of dosage and a more comprehensive longitudinal data set will be needed to more clearly understand and untangle the causal relationships between these factors.

Fidelity and outcomes—Consultation outcomes (student IEP progress) were related to three different fidelity components of quality (teacher coaching responsiveness at the implementation level, teacher quality of delivery and student responsiveness at the intervention level). These findings underscore the usefulness of multidimensional, multilevel measurement of fidelity and the potential importance of qualitative aspects of fidelity. Thus, in addition to adherence, consultants and researchers might wish to closely monitor quality of delivery (i.e., teacher engagement) and participant responsiveness (i.e., teacher responsiveness during consultation and student engagement during instruction), when

conducting consultation. Moreover, adherence, the traditional measure of fidelity, and dosage did not correlate with IEP outcomes. These non-significant relationships between treatment quantity and outcomes are consistent with some previous findings (e.g., Rowe et al., 2013), and illustrate the problem with simple approaches to the measurement of fidelity. A key additional question is what is meant by fidelity. Simply ensuring implementation of specific treatment elements may not be sufficient (i.e., quantity), instead it may be more important to measure how well elements are implemented (i.e., quality). Also, the findings support the idea that directly interacting /intervening (e.g., consultant impact on teacher) will be more impactful than indirectly intervening (e.g., consultant impact on student via the teacher). Consultants did not directly interact with students, whereas teachers did, thus it should not be surprising that fidelity focused on teacher behavior, as opposed to consultant behavior, were more closely related to student behavior and outcomes. However, additional studies are needed to obtain a more conclusive understanding of fidelity when measured quantitatively or qualitatively or when focusing on direct vs. indirect interventions.

The mediating effects of intervention-level fidelity—Consistent with these speculations about direct and indirect influences, we conducted mediation analyses to understand the complex relationships among the multilevel and multidimensional fidelity components and student outcomes. The analyses revealed that implementation-level variables, including consultant consultation adherence, consultant coaching adherence, teacher consultation responsiveness, teacher coaching responsiveness, and consultant consultation quality of delivery, were related to student outcomes via indirect pathways mediated by intervention-level fidelity components. Moreover, none had direct effects on student outcomes. The implications are twofold. First, even though implementation-level fidelity components either did not predict or were weaker predictors of student outcomes, they did have an influence on student outcomes through intervention-level fidelity components which had more direct influences on student outcomes. This finding is consistent with the general feature of consultation as an indirect intervention that supports the student via the teacher consultee. However, the current findings require future replications to confirm these mediating effects.

Second, the results provide insight into the differential effects of fidelity. Implementation-level fidelity indirectly influenced student outcomes via somewhat different paths. Consultant consultation adherence and teacher consultation responsiveness shared the same path (i.e., consultant consultation adherence/teacher consultation responsiveness -> teacher adherence -> teacher quality of delivery -> student responsiveness -> IEP outcomes), whereas consultant coaching adherence and teacher coaching responsiveness shared another path (consultant coaching adherence/teacher coaching responsiveness -> teacher quality of delivery -> student responsiveness -> IEP outcomes) and consultant consultation quality of delivery had a unique mediating path through intervention dosage (consultant consultation quality of delivery -> intervention dosage -> IEP outcomes). These results illustrate the complexity of the influence of multilevel, multidimensional fidelity on consultation outcomes. We encourage future studies to further examine these relationships.

Not surprisingly, the mediation analyses revealed that fidelity's influence on student outcomes during the initial consultation (i.e., consultant adherence and teacher

responsiveness), proceeded through a longer indirect chain compared to coaching sessions. The initial consultation laid out the consulting framework, set goals, and identified the ecological influences impacting the student and teaching plans, whereas the coaching sessions helped teachers to carry out the teaching plans through self-reflection, performance-based feedback, and outcomes-based monitoring. These results highlight the relative importance of coaching. Coaching is more closely tied to the actual teaching situation, and thus may be more predictive of actual teaching behavior, and in turn, of student outcomes. However, the significant indirect pathways also are consistent with the importance of having a comprehensive, clear consultation session before coaching to reinforce teacher's adherence to the intervention plans later.

Two mediation paths were unexpected. First, increased consultant adherence to COMPASS during coaching was related to lower teacher quality of delivery, which was related to poorer student responsiveness in learning, and then to poorer IEP outcomes. A similar finding was reported by Domitrovich and colleagues (2010). We believe that low consultant adherence represented teacher ability to adapt instructions based on the children's learning. This flexibility might lead to better learning outcomes because consultation also is a dynamic process. Quality coaching requires flexibility to adjust the support process based on the consultee's needs. We also speculate that low coaching adherence was related to consultants' ability to flexibly prioritize and select important implementation components case-by-case, which then led to better teaching behaviors. The current findings also emphasize the difference between consultation adherence and coaching adherence. To obtain positive student outcomes, consultants need to follow the protocol during the initial consultation to build a solid foundation for the following consultative process. After the consultation, consultants may need to exercise clinical judgment and individualize the manualized programs to fit consultees' needs. However, future studies will be needed to confirm these findings.

Second, the mediation path between consultant consultation quality of delivery and IEP outcomes was surprising. Increased consultant consultation quality of delivery was related to lower intervention dosage, which in turn was related to better IEP outcomes. As mentioned earlier, intervention dosage may not necessarily indicate better or higher quality implementation. It also is plausible that quality and quantity of an intervention can operate reciprocally. That is, there may be less need for intervention (dosage) if the intervention is high quality. Based on this logic, the mediation result might mean that better consultation quality of delivery is more effective, allowing teachers to adjust training dosage to a somewhat lower or optimal amount, in turn leading to better student outcomes. However, more exploratory, longitudinal research is needed.

Limitations

Lastly, we note our limitations. First, the fidelity measurements are COMPASS-specific and have not yet been thoroughly validated (e.g., using videotapes and independent observations). Moreover, the critical elements captured by these measures do not necessarily refer to consultation and coaching models in general but only to COMPASS consultation and coaching. Moreover, some are rather simple; for instance, dosage was represented by the

frequency of teaching per week. Although teaching frequency can serve as a proxy for dosage (Dane & Schneider, 1998), the exact duration spent teaching (e.g., minutes or hours) would likely be a more useful measure of dosage. Similarly, teacher adherence was measured by only one item, which may pose construct coverage issues. Second, although the current sample is relatively large compared to other consultation studies (e.g., Sanetti & Fallon, 2011), a larger sample size would allow more sophisticated statistical analyses. The current sample was also limited by the research design. For instance, only two consultants were used at the implementation level, limiting variability in consultant fidelity. Also, the sample was restricted to students with ASD and their teachers within special education, which might limit generalizability of the results to other populations and to general classroom teaching. Third, data at the implementation level were not complete at each time point, resulting in the use of mean scores to represent fidelity components in the mediation analyses. Thus, mediating effects were snapshots of four interwoven but temporally distinct processes. For instance, teacher quality of delivery at time one might influence coaching adherence at time two. Fourth, although researchers have identified four common sources of fidelity measures (i.e., direct observation, self-report, interviews, and archival records; Durlak & DuPre, 2008; O'Donnell, 2008), the current study relied heavily on direct observations, including from consultants and teachers, which may pose measurement bias. In addition, teachers were asked to rate the performance and fidelity of the consultants. Even though there was variability across the scores, teachers' ratings might be confounded by either the consultative relationships or social desirability. Lastly, it is worth noting that the current study addresses a very limited set of questions within implementation science. Many implementation- and intervention- level variables that potentially impact student outcomes were not included. For instance, factors related to intervention characteristics (e.g., complexity), outer settings (e.g., external incentives for teacher consultees), inner settings (e.g., administrative support, tension for change), characteristics of individuals (e.g., teacher's self-efficacy), and process (e.g., quality of teaching plans) may influence student outcomes, but the current study either did not collect or did not analyze this level of data (Damschroder & Lowery, 2013).

Areas for Future Research

As mentioned in the general discussion section, more study is needed to verify the mediating effects of the intervention-level fidelity. However, based on existing conceptual frameworks (Noell, 2008; Sheridan et al., 2013) and clinical experience, we expect that the results should generalize to other structured consultation models designed for special education teachers. Additionally, more efforts are needed to develop standardized measures of these fidelity components (Sheridan et al., 2009). Such efforts will lead to a clearer understanding of the nature and application of multilevel, multi-dimensional fidelity components. In particular, we need more development of and understanding of the reliability and validity of different sources of fidelity measures (i.e., direct observation, self-report, interviews, and archival records; Durlak & DuPre, 2008; O'Donnell, 2008). Comparative studies are needed to provide a comprehensive picture of the utility and efficiency of each source and potential measure. In addition, as noted above, we need more longitudinal studies to examine how fidelity components change and interact over time, as well as larger studies that include the many additional variables thought to impact outcomes, to examine how they correlate with

and moderate or mediate different aspects of fidelity. Finally, it would be useful to include additional input from parents in terms of fidelity and outcome measurement.

Acknowledgments

References

Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin. 1959; 56(2):81–105. [PubMed: 13634291]

Crawford L, Carpenter DM, Wilson MT, Schmeister M, McDonald M. Testing the relation between fidelity of implementation and student outcomes in math. Assessment for Effective Intervention. 2012; Advance online publication. doi: 10.1177/1534508411436111

Damschroder LJ, Lowery JC. Evaluation of a large-scale weight management program using the consolidated framework for implementation research (CFIR). Implementation Science. 2013; 8(1): 51–68. DOI: 10.1186/1748-5908-8-51 [PubMed: 23663819]

Dane AV, Schneider BH. Program integrity in primary and early secondary prevention: Are implementation effects out of control? Clinical Psychology Review. 1998; 18(1):23–45. [PubMed: 9455622]

Domitrovich CE, Gest SD, Jones D, Gill S, Sanford DeRousie RM. Implementation quality: Lessons learned in the context of the Head Start REDI trial. Early Childhood Research Quarterly. 2010; 25:284–298. DOI: 10.1016/j.ecresq.2010.04.001 [PubMed: 22844183]

Dunst CJ, Trivette CM, Raab M. An implementation science framework for conceptualizing and operationalizing fidelity in early childhood intervention studies. Journal of Early Intervention. 2013; 35:85–101. DOI: 10.1186/1748-5908-2-40

Durlak JA, DuPre EP. Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. American Journal of Community Psychology. 2008; 41:327–350. doi:1007/s10464-008-9165-0. [PubMed: 18322790]

Dusenbury L, Brannigan R, Falco M, Hansen WB. A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. Health Education Research. 2003; 18:237–256. DOI: 10.1093/her/18.2.237 [PubMed: 12729182]

Fogarty M, Oslund E, Simmons D, Davis J, Simmons L, Anderson L, Roberts G. Examining the effectiveness of a multicomponent reading comprehension intervention in middle schools: A focus on treatment fidelity. Educational Psychology Review. 2014; 26:425–449. DOI: 10.1007/s10648-014-9270-6

Gresham FM. Assessment of treatment integrity in school consultation and prereferral intervention. School Psychology Review. 1989; 18:37–50.

Gresham FM, Lopez MF. Social validation: A unifying concept for school-based consultation research and practice. School Psychology Quarterly. 1996; 11:204–227. DOI: 10.1037/h0088930

Hamre BK, Justice LM, Pianta RC, Kilday C, Sweeney B, Downer JT, Leach A. Implementation fidelity of MyTeachingPartner literacy and language activities: Association with preschoolers' language and literacy growth. Early Childhood Research Quarterly. 2010; 25:329–347. DOI: 10.1016/j.ecresq.2009.07.002

Hamre B, Pianta R, Hatfield B, Jamil F. Evidence for general and domain-specific elements of teacher-child interactions: Associations with preschool children's development. Child Development. 2014; 85:1257–1274. DOI: 10.1111/cdev.12184 [PubMed: 24255933]

Harn B, Parisi D, Stoolmiller M. Balancing fidelity with flexibility and fit: What do we really know about fidelity of implementation in schools? Exceptional Children. 2013; 79:181–193. DOI: 10.1177/001440291307900204

Hayes, AF. PROCESS: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling. 2012. Retrieved from http://www.afhayes.com/public/process2012.pdf

Hoagwood K, Atkins M, Ialongo N. Unpacking the black box of implementation: The next generation for policy, research and practice. Administration and Policy in Mental Health and Mental Health Services Research. 2013; 40(6):451–455. DOI: 10.1007/s10488-013-0512-6 [PubMed: 23942647]

Kampwirth, TJ., Powers, KM. Collaborative consultation in the schools: Effective practices for students with learning and behavior problems. Upper Saddle River: New Jersey: Pearson Education; 2002.

Klem AM, Connell JP. Relationships matter: Linking teacher support to student engagement and achievement. Journal of School Health. 2004; 74:262–273. DOI: 10.1111/j. 1746-1561.2004.tb08283.x [PubMed: 15493703]

Knoche LL, Sheridan SM, Edwards CP, Osborn AQ. Implementation of a relationship-based school readiness intervention: A multidimensional approach to fidelity measurement for early childhood. Early Childhood Research Quarterly. 2010; 25:299–313. DOI: 10.1016/j.ecresq.2009.05.003 [PubMed: 20824112]

Lord C, Risi S, Lambrecht L, Cook EH Jr, Leventhal BL, DiLavore PC, Rutter M. The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. Journal of Autism and Developmental Disorders. 2000; 30:205–223. DOI: 10.1023/A:1005592401947 [PubMed: 11055457]

McConachie, H., Parr, JR., Glod, M., Hanratty, J., Livingstone, N., Oono, IP., Garland, D. Systematic review of tools to measure outcomes for young children with autism spectrum disorder. 2015. Retrieved from https://ore.exeter.ac.uk/repository/bitstream/handle/10871/17760/FullReport-hta19410-2.pdf?sequence=1&isAllowed=y

McGrew J, Ruble L, Smith I. Autism Spectrum Disorder and Evidence-Based Practice in Psychology. Clinical Psychology: Science and Practice. 2016; 23(3):239–255.

Miller WR, Rollnick S. The effectiveness and ineffectiveness of complex behavioral interventions: impact of treatment fidelity. Contemporary Clinical Trials. 2014; 37:234–241. DOI: 10.1016/j.cct. 2014.01.005 [PubMed: 24469237]

National Association of School Psychologists. Model for comprehensive and integrated school psychological services. 2010. Retrieved from www.nasponline.org/Documents/.../PracticeModelGuide_Web.pdf

Noell, GH. Studying relationships among consultation process, treatment integrity, and outcomes. In: Erchul, WP., Sheridan, SM., editors. Handbook of research in school consultation: Empirical foundations for the field. Mahwah, NJ: Erlbaum; 2008. p. 323-342.

O'Donnell CL. Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. Review of Educational Research. 2008; 78:33–84. DOI: 10.3102/0034654307313793

Odom SL, Fleming K, Diamond K, Lieber J, Hanson M, Butera G, Marquis J. Examining different forms of implementation and in early childhood curriculum research. Early Childhood Research Quarterly. 2010; 25:314–328. DOI: 10.1016/j.ecresq.2010.03.001 [PubMed: 21874091]

Perepletchikova F, Treat TA, Kazdin AE. Treatment integrity in psychotherapy research: analysis of the studies and examination of the associated factors. Journal of Consulting and Clinical Psychology. 2007; 75:829–841. DOI: 10.1037/0022-006X.75.6.829 [PubMed: 18085901]

Power TJ, Blom-Hoffman J, Clarke AT, Riley-Tillman T, Kelleher C, Manz PH. Reconceptualizing intervention integrity: A partnership-based framework for linking research with practice. Psychology in the Schools. 2005; 42:495–507. DOI: 10.1002/pits.20087

Reinke WM, Herman KC, Stormont M, Newcomer L, David K. Illustrating the multiple facets and levels of fidelity of implementation to a teacher classroom management intervention. Administration and Policy in Mental Health and Mental Health Services Research. 2013; 40:494–506. DOI: 10.1007/s10488-013-0496-2 [PubMed: 23636711]

Rowe C, Rigter H, Henderson C, Gantner A, Mos K, Nielsen P, Phan O. Implementation fidelity of multidimensional family therapy in an international trial. Journal of Substance Abuse Treatment. 2013; 44:391–399. DOI: 10.1016/j.jsat.2012.08.225 [PubMed: 23085040]

Ruble LA, Dalrymple NJ. Compass: A Parent—Teacher Collaborative Model for Students With Autism. Focus on Autism and Other Developmental Disabilities. 2002; 17(2):76–83. DOI: 10.1177/10883576020170020201

Ruble LA, Dalrymple NJ, McGrew JH. The effects of consultation on individualized education program outcomes for young children with autism: The collaborative model for promoting competence and success. Journal of Early Intervention. 2010; 32(4):286–301. DOI: 10.1177/1053815110382973 [PubMed: 21691449]

Ruble, LA., Dalrymple, NJ., McGrew, JH. Collaborative model for promoting competence and success for students with ASD. Springer Science & Business Media; 2012.

Ruble L, McDuffie A, King AS, Lorenz D. Caregiver responsiveness and social interaction behaviors of young children with autism. Topics in Early Childhood Special Education. 2008; 28(3):158–170. DOI: 10.1177/0271121408323009

Ruble L, McGrew JH. Teacher and child predictors of achieving IEP goals of children with autism. Journal of Autism and Developmental Disorders. 2013; 43(12):2748–2763. DOI: 10.1007/s10803-013-1884-x [PubMed: 23838728]

Ruble, LA., McGrew, JH. COMPASS and implementation science: Improving educational outcomes of children with ASD. Springer; 2015.

Ruble L, McGrew JH, Toland MD. Goal attainment scaling as outcome measurement for randomized controlled trials. Journal of Autism and Developmental Disorders. 2012; 42(9):1974–1983. DOI: 10.1007/s10803-012-1446-7 [PubMed: 22271197]

Ruble LA, McGrew JH, Toland MD, Dalrymple NJ, Jung LA. A randomized controlled trial of COMPASS web-based and face-to-face teacher coaching in autism. Journal of Consulting and Clinical Psychology. 2013; 81(3):566–572. DOI: 10.1037/a0032003 [PubMed: 23438314]

Sanetti LMH, Kratochwill TR. Treatment integrity in behavioral consultation: Measurement, promotion, and outcomes. International Journal of Behavioral Consultation and Therapy. 2007; 4:95–114. http://dx.doi.org/10.1037/h0100835.

Sanetti LMH, Kratochwill TR. Toward developing a science of treatment integrity: Introduction to the special series. School Psychology Review. 2009; 38:445–259. DOI: 10.1177/0014402914551738

Sanetti LM, Fallon LM. Treatment integrity assessment: How estimates of adherence, quality, and exposure influence interpretation of implementation. Journal of Educational and Psychological Consultation. 2011; 21:209–232. DOI: 10.1080/10474412.2011.595163

Schoenwald SK, Sheidow AJ, Letourneau EJ. Toward effective quality assurance in evidence-based practice: Links between expert consultation, therapist fidelity, and child outcomes. Journal of Clinical Child and Adolescent Psychology. 2004; 33(1):94–104. DOI: 10.1207/S15374424JCCP3301_10 [PubMed: 15028545]

Schulte AC, Easton JE, Parker J. Advances in treatment integrity research: Multidisciplinary perspectives on the conceptualization, measurement, and enhancement of treatment integrity. School Psychology Review. 2009; 38(4):460–475. DOI: 10.1177/1053815110382973

Sheridan SM, Clarke BL, Knoche LL, Edwards CP. The effects of conjoint behavioral consultation in early childhood settings. Early Education and Development. 2006; 17:593–617. DOI: 10.1207/s15566935eed1704_5

Sheridan, SM., Rispoli, K., Holmes, SR. Treatment integrity in conjoint behavioral consultation: Conceptualizing active ingredients and potential pathways of influence. In: Sanetti, LM., Kratochwill, TR., editors. Treatment integrity: A foundational for evidence-based practice in applied psychology. Washington, DC: American Psychological Association; 2013. p. 107-123.

Sanetti, T., Kratochwill. Treatment integrity: Conceptual, methodological, and applied considerations for practitioners. Washington, DC: American Psychological Association; p. 255-278.

Sheridan SM, Swanger-Gagné M, Welch GW, Kwon K, Garbacz SA. Fidelity measurement in consultation: Psychometric issues and preliminary examination. School Psychology Review. 2009; 38:476–495. DOI: 10.1177/1053815113504625

Sladeczek IE, Elliott SN, Kratochwill TR, Robertson-Mjaanes S, Stoiber KC. Application of goal attainment scaling to a conjoint behavioral consultation case. Journal of Educational and Psychological Consultation. 2001; 12(1):45–58.

Spear, CF. Examining the relationship between implementation and student outcomes: The application of an implementation measurement framework. Doctoral dissertation, University of Oregon. 2014. Retrieved from https://scholarsbank.uoregon.edu/xmlui/handle/1794/18710

Swanson E, Wanzek J, Haring C, Ciullo S, McCulley L. Intervention fidelity in special and general education research journals. The Journal of Special Education. 2013; 47:3–13. DOI: 10.1177/0022466911419516

Wanless SB, Rimm-Kaufman SE, Abry T, Larsen RA, Patton CL. Engagement in training as a mechanism to understanding fidelity of implementation of the Responsive Classroom Approach. Prevention Science. 2015; 16:1107–1116. DOI: 10.1007/s11121-014-0519-6 [PubMed: 25351865]

Weisz JR, Jensen-Doss A, Hawley KM. Evidence-based youth psychotherapies versus usual clinical care: A meta-analysis of direct comparisons. American Psychologist. 2006; 61(7):671–689. [PubMed: 17032068]

Wheeler JJ, Baggett BA, Fox J, Blevins L. Treatment integrity: A review of intervention studies conducted with children with autism. Focus on Autism and Other Developmental Disabilities. 2006; 21(1):45–54. DOI: 10.1177/10883576060210010601

Wong C, Odom SL, Hume KA, Cox CW, Fettig A, Kurcharczyk S, et al. Evidence-based practices for children, youth, and young adults with autism spectrum disorder: A comprehensive review. Journal of Autism and Developmental Disorders. 2015; Advance online publication. doi: 10.1007/s10803-014-2351-z

Zvoch K. Treatment fidelity in multisite evaluation: A multilevel longitudinal examination of provider adherence status and change. American Journal of Evaluation. 2009; 30(1):44–61. DOI: 10.1177/1098214008329523
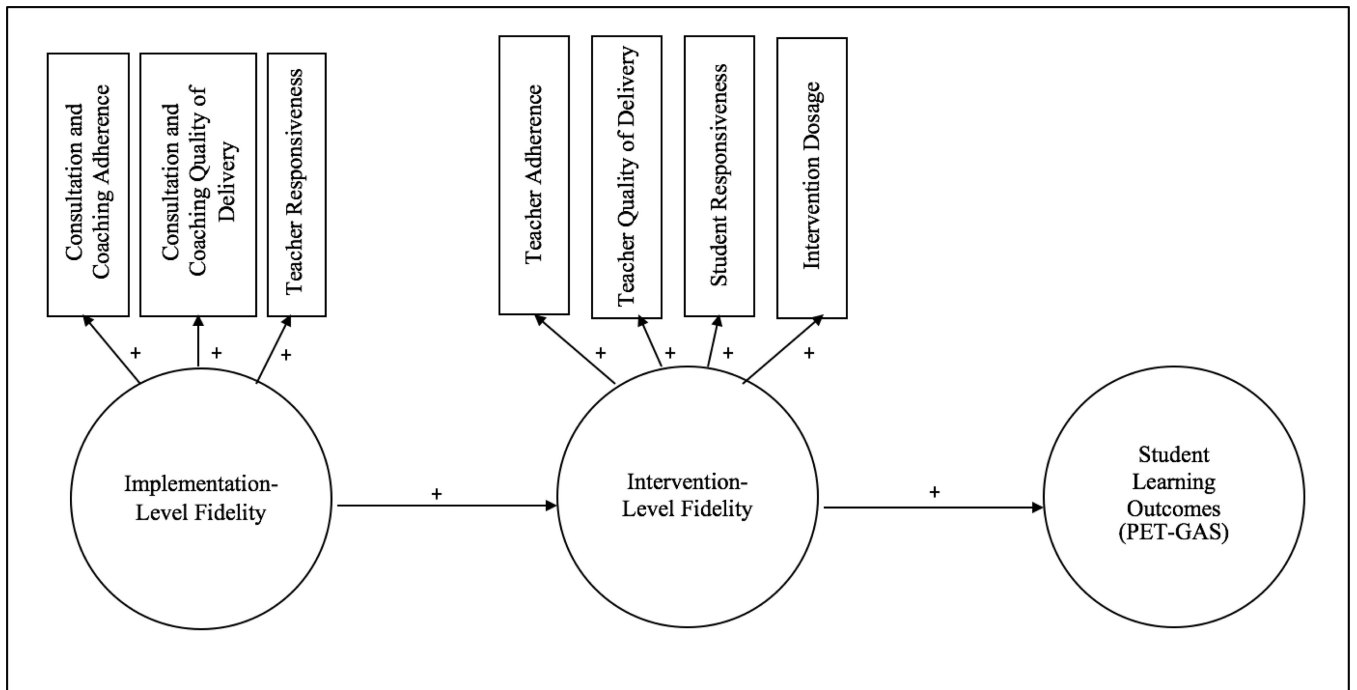
Figure 1.
A working model of multilevel, multidimensional fidelity of COMPASS. Dotted lines represent constructs did not vary across COMPSS sessions. It is worth noting that this figure is akin to a SEM model; however, the current study did not use SEM.

	Constructs	Implementation Level						Intervention Level			
		Adherence		Quality of Delivery		Responsiveness	Dosage	Adherence	Quality of Delivery	Responsiveness	Dosage
	Measures / Time Points	Consultant Consultation Adherence Form	Consultant Coaching Adherence Form	Consultant Quality of Delivery	Coaching Quality of Delivery	Teacher Responsiveness	Consultation Dosage	Teacher Adherence	Teacher Quality of Delivery	Student Responsiveness	Intervention Dosage
Study 1	Consultation	X		X		X	-				
	Coaching 1		X*		X*	X	-	X	X	X	X
	Coaching 2		X*		X*	X	-	X	X	X	X
	Coaching 3		X*		X*	X	-	X	X	X	X
	Coaching 4		X*		X*	X	-	X	X	X	X
Study 2	Consultation	X		X		X	-				
	Coaching 1		X*		X*	X	-	X	X	X	X
	Coaching 2		X*		X*	X	-	X	X	X	X
	Coaching 3		X*		X*	X	-	X	X	X	X
	Coaching 4		X*		X*	X	-	X	X	X	X

Figure 2.

Measurement used at different time points. X=Data collected; - = Constant; Empty = Data not collected. * = Only collected in study 2.
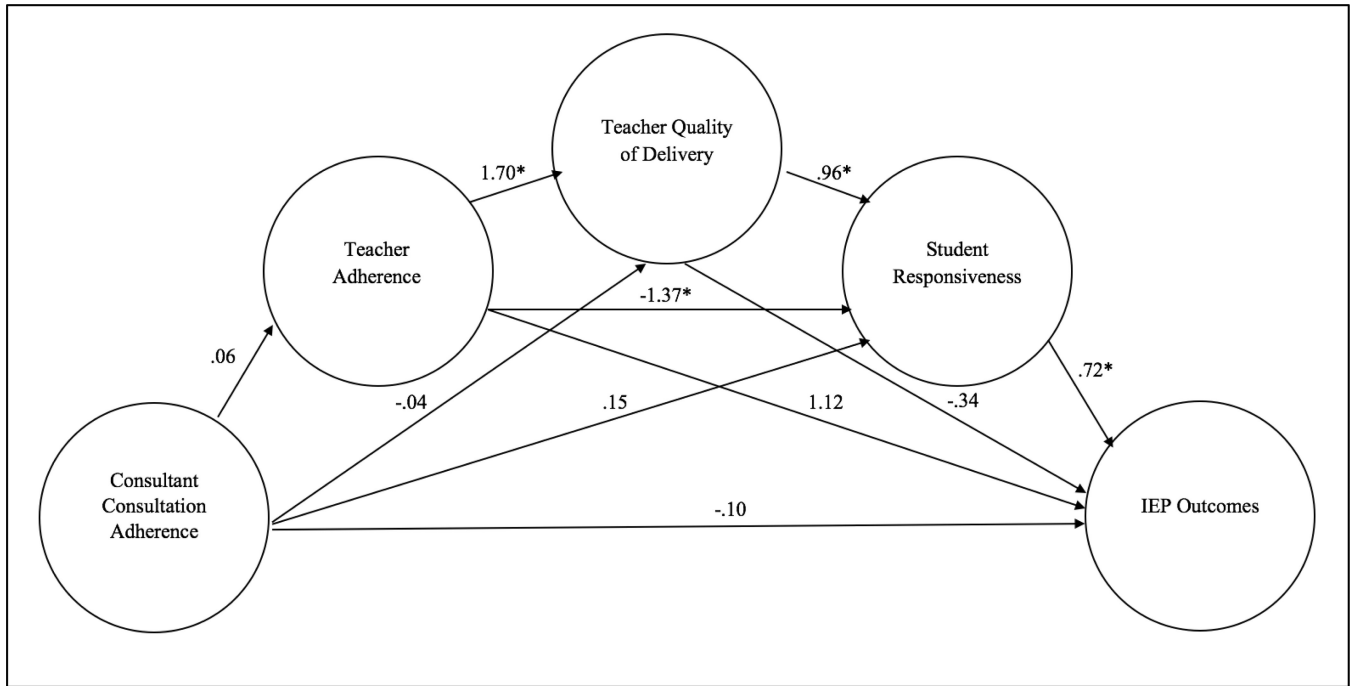
Figure 3.
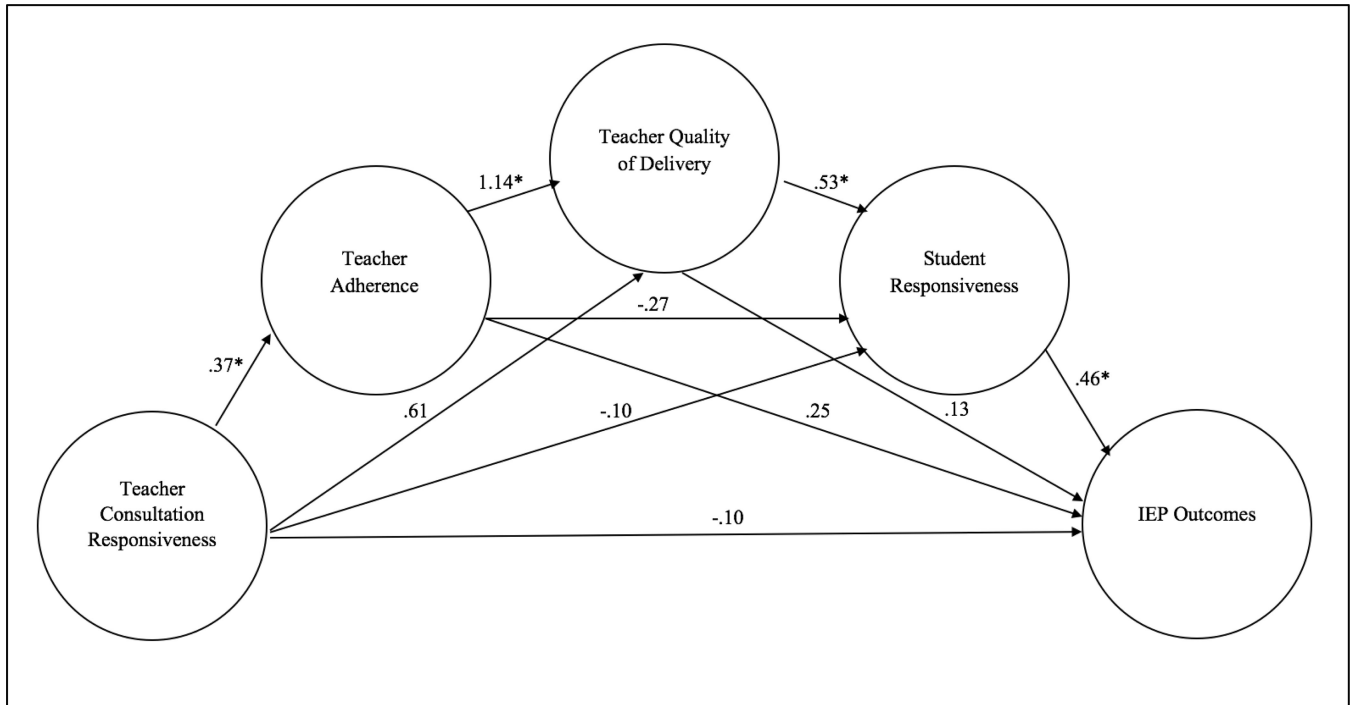The mediating effect between consultant consultation adherence and IEP outcomes. $* = p <$.
05.

Figure 4.
The mediating effect between teacher consultation responsiveness and IEP outcomes. * = *p*<.05.
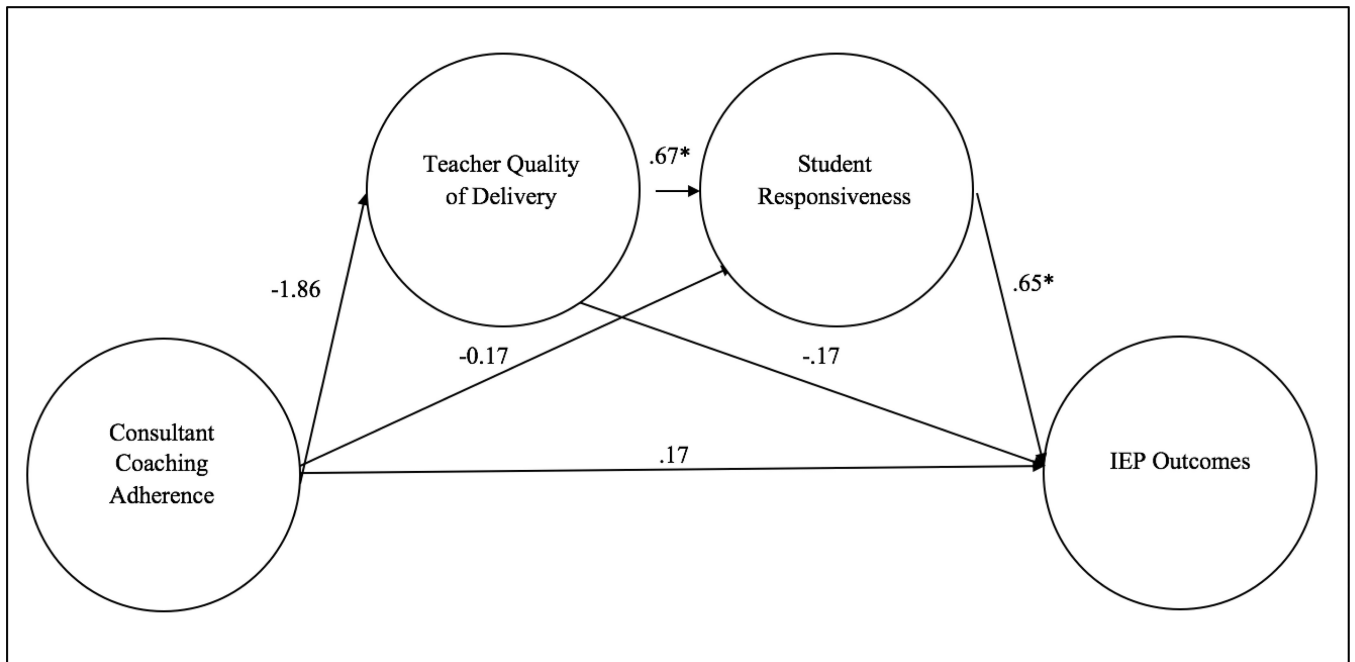
Figure 5.
The mediating effect between consultant coaching adherence and IEP outcomes. * = $p < .05$.
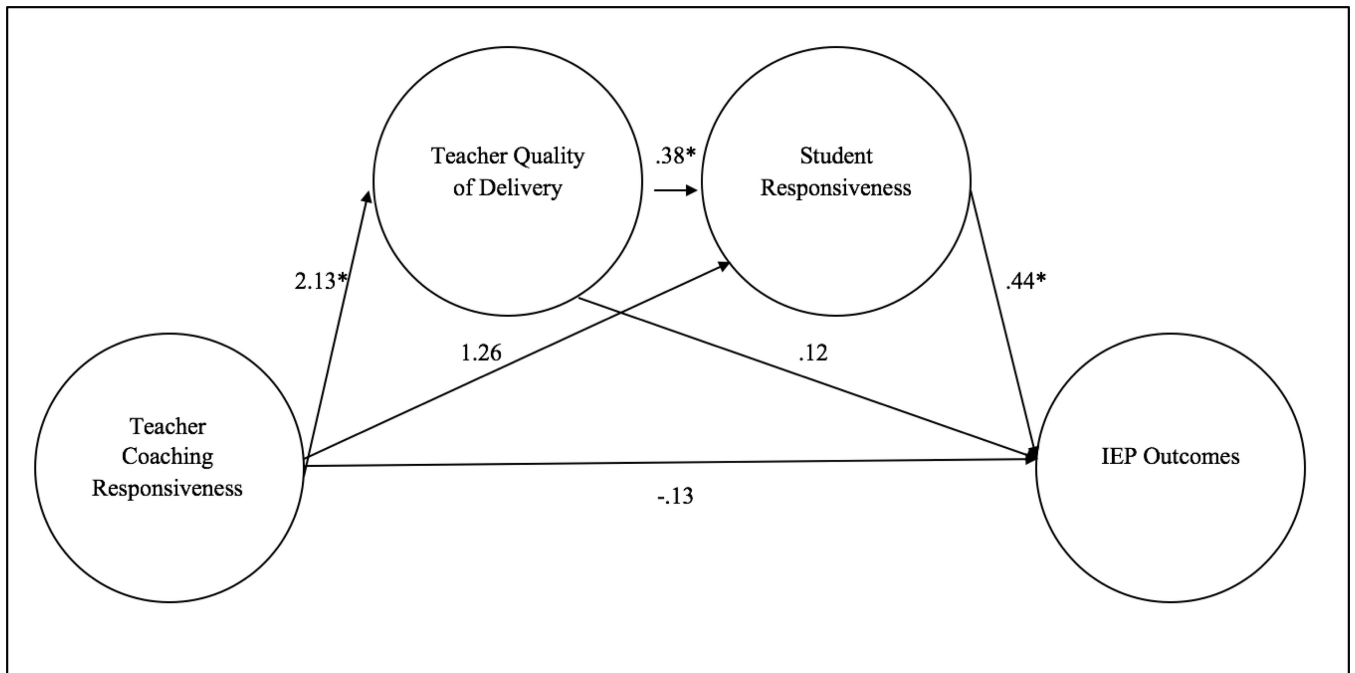
Figure 6.
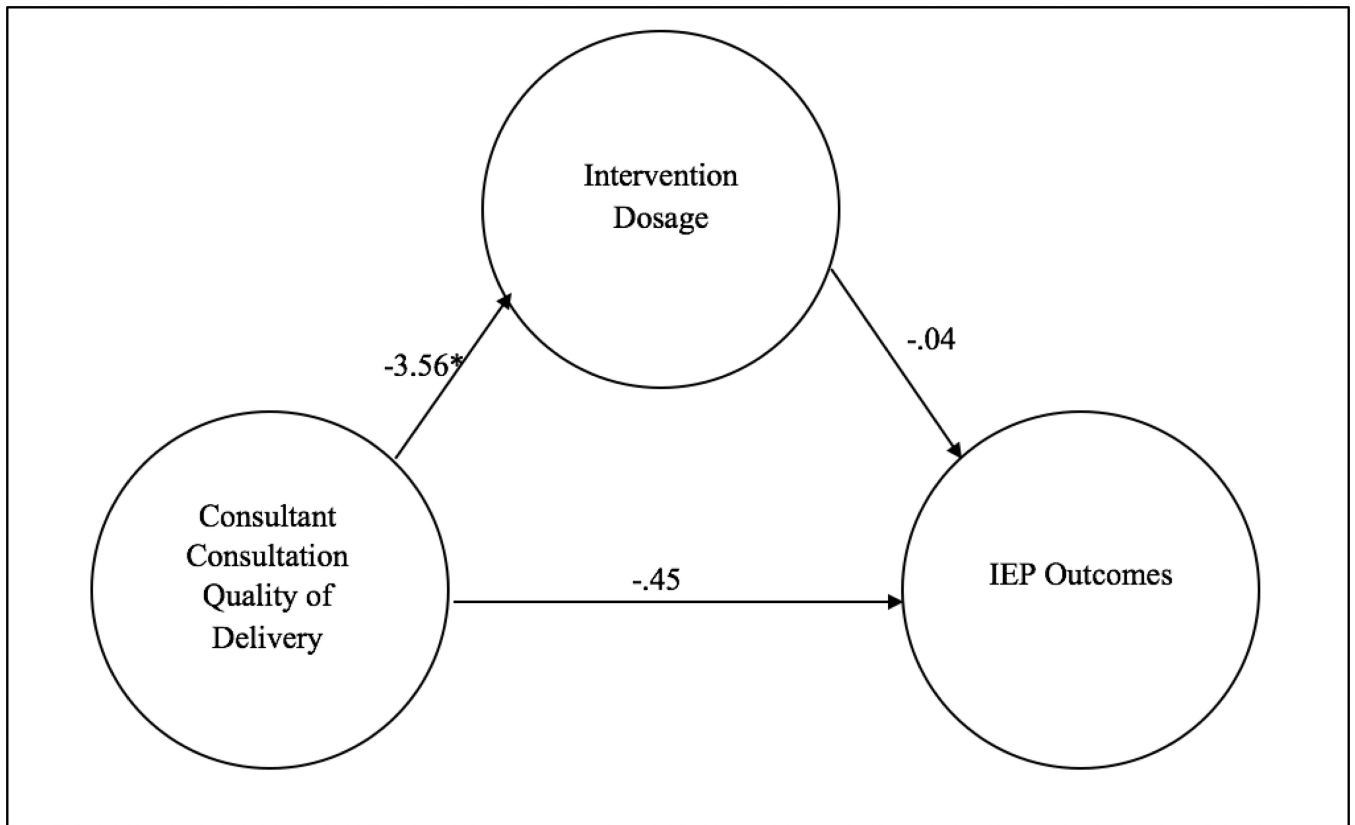The mediating effect between teacher responsiveness and IEP outcomes. * = $p < .05$.

Figure 7.
The mediating effect between consultant consultation quality of delivery responsiveness and
IEP outcomes. * = $p < .05$.

Table 1

Demographic Information

School variables	n	%
Urban/ suburban	42	89.4
Rural (Population less than 5,000)	5	10.6
Teacher variables	*M (SD)*	Range
Teaching experience (years)	11.46 (7.85)	0–32
Caseload (number of current students)	13.22(6.45)	3–34
Education	*n*	*%*
Bachelor	17	39.5
Master	24	55.8
Emergency Certificate	1	2.3
Other	1	2.3
Gender		
Male	2	4.3
Female	45	95.7
Student variables	*M (SD)*	Range
Age (years)	5.95 (1.61)	3–9
Gender	*n*	*%*
Male	39	83.0
Female	8	17.0
Family income		
Less than $10,000	4	11.1
$10,000–24,999	7	19.4
$25,000–49,999	9	25.0
$50,000–100,000	13	36.1
>$100,000	3	8.3
Race		
Caucasian	37	80.4
African American	5	10.9
Asian/Pacific Islander	1	2.2
Other	3	6.5

*
Reported values are based on available data. Sample sizes may vary due to missing data.

Table 2

Summary of Intercorrelations, Means and Standard Deviations for Scores at Implementation and Intervention Levels

Measures	1	2	3	4	5	6	7	8	9	10	M	SD	Range
1. Teacher adherence (Mean)	-										3.65	.98	1.00–5.00
2. Teacher quality of delivery (Mean)	.29†	-									14.14	1.78	10.38–17.13
3. Student responsiveness (Mean)	.12	.47**	-								14.42	1.70	10.50–17.13
4. Intervention dosage (Mean)	.12	−.23	−.45**	-							13.37	11.30	0–60.58
5. Consultant consultation adherence	.15	.06	.14	.11	-						22.28	3.89	9.00–25.00
6. Consultant consultation quality of delivery	.03	−.02	.04	−.42**	.09	-					2.60	1.40	1.00–4.00
7. Teacher consultation responsiveness	.31*	.25	.07	.07	.19	.23	-				3.88	.60	2.43–5.00
8. Consultant coaching adherence (Mean)	.08	−.21	−.20	.22	.09	.10	.33	-			3.80	.23	3.36–4.00
9. Teacher coaching responsiveness (Mean)	.54**	.63**	.37*	−.24	.03	.18	.34*	.30	-		4.27	.45	3.06–4.91
10. Consultant coaching quality of delivery (Mean)	.21	.27	.12	−.03	−.02	−.11	−.05	.15	.16	-	3.76	.38	2.88–4.00
11. PET-GAS (Mean)	.10	.44**	.52**	−.25	.01	−.14	.02	−.03	.38*	.26	−.84	1.81	−.475–5.00

*
= $p < .05$;

**
= $p < .01$;

†= .50

Note: Consultant consultation adherence and consultant coaching quality of delivery were only collected for the second study (n=29). The rest of the variables were collected from the full sample (n=47)