A META-ANALYSIS OF SCHOOL-BASED, BEHAVIORAL CONSULTATIONS FOR EXTERNALIZING BEHAVIORS

A Dissertation

by

KELSEY MARIE RAGAN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Lisa Bowman-Perrott
Co-Chair of Committee,	Krystal Simmons
Committee Members,	Mack Burke
	Robert Heffer
Head of Department,	Shanna Hagan-Burke

August 2019

Major Subject: School Psychology

ABSTRACT

Many teachers feel unprepared to handle the behavior problems that arise during the school day. School-based, behavioral consultation is one method of approaching this issue. While consultation has been demonstrated to be effective at providing teachers with the means to increase appropriate student behavior, there are still several unknown factors with which practitioners and researchers contend. In particular, questions about treatment fidelity and methods for its improvement have been noted in the literature. Of additional concern is the exclusion of single-case and unpublished studies in more recent meta-analytic research.

The purpose of this paper was twofold. The first goal was to re-examine the impact of consultation on externalizing student behavioral outcomes by means of a more recent review of the school-based consultation literature. This analysis was intended to focus on single-case studies and both published and unpublished literature. Additionally, the author sought to investigate a possible moderator of this relationship: performance feedback, a method commonly used to measure and improve treatment integrity in consultations.

To accomplish these tasks, a sample of 26 single-case studies, both published and unpublished, was systematically collected. The data from the sample were evaluated using nonparametric and parametric methods. The results suggested school-based, behavioral consultation can produce improvements in externalizing student behaviors, with a Baseline Corrected Tau of .46 and a statically significant *p*-value according to multilevel modeling analyses. The presence of performance feedback resulted in statistically significant improvements, but relatively small practical changes in behavioral outcomes. Of additional note was the publication bias identified within the sample, with larger effects seen in published literature, suggesting that meta-analyses which do not include unpublished studies may be

biased. Overall, these conclusions support the use of consultation for improving behavioral problems in students and highlight the importance of considering treatment fidelity issues.

DEDICATION

This dissertation is lovingly dedicated to my family, who have been constant supports and have helped me to continually drive myself after my dreams. Also, special thanks to my mother for always being genuinely interested in my dissertation, even when I kept you on the phone for far too long, excitedly talking about graphs and statistical procedures.

ACKNOWLEDGEMENTS

I wish to express my appreciation for Dr. Krystal Simmons, Dr. Lisa Bowman-Perrott, and Dr. Mack Burke who were invaluable during not only the dissertation process, but throughout my time at Texas A&M University. Thank you for your unending support. Your mentorship, kindness, and wisdom have impacted me both professionally and personally. My gratitude also goes out to Dr. Rob Heffer. I have greatly valued your time, advice, and suggestions, which have helped to make this dissertation far better than it would have been without your guidance. Finally, thanks to my friends and colleagues and the department faculty and staff for making my time at Texas A&M University such a joyful and fulfilling experience.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professors Lisa Bowman-Perrott, Krystal Simmons, and Mack Burke of the Department of Educational Psychology and Professor Robert Heffer of the Department of Psychology & Brain Sciences. The coding conducted for inter-coder reliability purposes depicted in Chapter 3 was conducted in part by Bianca Watkins of the Department of Educational Psychology.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by a fellowship from Texas A&M University.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Many students face issues that could put them at risk for developing external behavioral problems (Campbell, Shaw, & Gilliom, 2000). During the 2013-2014 academic year, around 354,000 children were receiving services under the Emotional Disturbance (ED) category of special education in the United States (National Center for Education Statistics, n.d.). Unfortunately, this number likely underestimates the prevalence of students' emotional and behavioral needs, as not all children who exhibit behavioral problems are diagnosed with an emotional disorder and receive special education services. Although it is estimated that about 20% of students are experiencing a serious mental health issue, only around 1% of students are served in schools under the ED category (Walker, Nishioka, Zeller, Severson, & Feil, 2000). Furthermore, these students do not appear to experience the same educational benefits as their peers. Research has suggested that behavioral difficulties are predictive of lower educational achievement across academic subjects and grades (McClelland et al., 2014; Nelson, Benner, Lane, & Smith, 2004). Thus, these students often struggle not only emotionally and behaviorally, but academically as well.

A possible contributing factor to these concerns is the individual teacher's self-efficacy and competency with appropriately addressing behavioral issues. Teacher education programs often are not required to train pre-service teachers on evidenced-based behavioral interventions (Freeman, Simonsen, Briere, & MacSuga-Gage, 2014), and many teachers report feeling unprepared to implement behavioral management strategies within their classroom (Begeny & Martens, 2006). Problematic and disruptive behaviors result in increased teacher burnout and disrupt the teaching process—negatively impacting both the individual student and the classroom

at large (Greene, Beszterczey, Katzenstein, Park, & Goring, 2002). Thus, teachers would appear to benefit from additional training to increase their proficiencies in this area.

Consultation is one possible solution to providing technical support to teachers and increasing the use of evidence-based behavioral interventions. In the field of consultation, several models have been developed which target different areas of intervention, each with their own variations in theoretical oriental and approach to the process, including popular models such as the Behavioral, Mental Health, and Organizational Developmental Models (Kratochwill, Elliott, & Callan-Stoiber, 2002). Despite their differences, all models share an indirect, problem-solving approach when addressing behavior concerns (Kratochwill et al., 2002). In particular, the behavioral consultation approach, developed by Kratochwill and Bergan (1990), is well suited to addressing behavioral concerns. As the name implies, behavioral consultations are typically used to remedy inappropriate behaviors and have been demonstrated to be superior to other models when the goals of a consultation are behavioral in nature (Batts, 1987).

There are three primary roles in a consultation: the *consultant*, who is an expert knowledgeable about the consultation process and appropriate interventions; the *consultee*, who works directly with the client; and the c*lient*, the individual who is intended to experience beneficial outcomes (Erchul & Martens, 2010). In the school-based model specifically, the consultee is a teacher, parent, or other school personnel. They in turn work with a consultant to aid the student-client (Zins, Kratochwill, & Elliott, 1993). Through the school-based behavioral consultation model, teachers can collaborate with expert consultants, a process which is intended to increase the teacher's competency when independently handling behavioral issues.

When evaluating a consultation, researchers are usually interested in client-based changes, such as student behavioral, social, or academic improvements, but studies may also

focus on skill development in the teacher or even impacts on the consultant themselves (e.g., Dart, Cook, Collins, Gresham, & Chenier, 2012). Thus, consultation is conceptualized as a vehicle for both improving client outcomes and increasing the skills of the consultee. Given the theoretical strengths of the consultation process, it holds promise for addressing not only student concerns but teacher competencies as well.

In general, researchers have found school-based consultation effective at meeting the needs of teachers who struggle with behavior management. In fact, several meta-analyses have evaluated consultation with teacher-consultees and found it to have a moderate to large effect on various outcomes for both clients and consultees, including behavioral, academic, attitudinal improvements (e.g., Davis, 2012; Jackson, 1986; Reddy, Barboza-Whitehead, Files, & Rubel, 2000; Sibley, 1986). For example, Reddy and colleagues (2000) found that students improved their attitude, behavior, and academic skills after consultation, with an overall effect size for client outcomes at 1.3 (Glass' Δ).

Although consultation in general has proven to be an evidenced-based vehicle for intervention delivery, questions remain regarding the most effective methods for implementing this process. Concerns surrounding treatment integrity in particular have been prevalent in the research (Batts, 1987; Davis, 2012; Wilkinson, 2006). Treatment integrity (or treatment fidelity) is defined as "the accuracy and consistency with which each component of the treatment or intervention plan is implemented" (Wilkinson, 2006, p. 428). Due to the indirect nature of the consultation process, there have been questions about the fidelity with which teachers are able to implement the consultation. Does a consultation session provide sufficient support and training for teachers who may be unfamiliar with the intervention and who are then expected to implement the plan independently? While previous school-based meta-analyses have noted these

concerns, they have not yet investigated the answers to these questions or what might be done to address treatment integrity issues when they are identified.

Performance feedback may be used as a means of resolving this issue. When using performance feedback, a consultant observes the consultee implementing the intervention and later provides customized feedback about the level of treatment integrity observed, as well as methods for improving it (Sanetti & Kratochwill, 2007). Research has shown performance feedback can result in improved treatment integrity outcomes, and feedback may be provided verbally or in a written format to achieve these results (Kaufman, Codding, Markus, Tryon, & Kyse, 2013). For example, in a study by DiGennaro, Martens, and McIntyre (2005) consultants conducted daily meetings with consultees, until treatment integrity reached 100%. The result was that teachers implemented the intervention with more fidelity and maintained those gains over time. If studies employ this feedback method, the implication is that integrity issues were assessed and dealt with during the consultation process. One might hypothesize, therefore, that studies utilizing performance feedback may result in larger client effect sizes.

Of additional concern is the fact that there has yet to be a meta-analysis looking exclusively at behavioral client-level outcomes. Previous studies have often included behavior as one variable of interest, but have focused on a broader range of consultation goals (e.g., Davis, 2012; Reddy et al., 2000). Thus, it may prove beneficial to evaluate behavior outcomes in isolation, to help determine what issues may be present in this particular literature base, as well as provide recommendations for practice that are specifically tailored to behavioral goals. Furthermore, there have been gaps in more recent school-based meta-analyses, due to the exclusion of single-case studies or unpublished literature (e.g., Batts, 1987; Davis, 2012; Reddy et al., 2000).

In order to investigate these issues, the current dissertation evaluated the impact of school-based, behavioral consultation on externalizing student behaviors. A sample including single-case and both unpublished and published literature was used. Performance feedback was examined as a potential moderator as well, to determine if measuring and addressing treatment fidelity issues using performance feedback methods impacted consultation outcomes. Finally, these findings were used to provide practical recommendations for the implementation of consultation methods in an educational setting.

2. LITERATURE REVIEW

Previous Consultation Meta-Analyses

The purpose of this chapter was to summarize the meta-analytic literature related to school-based consultation. Meta-analysis is a useful approach for summarizing, integrating, and interpreting a body of literature (Lipsey & Wilson, 2001). This method of analysis allows the researcher to calculate an overall, or omnibus, effect size with more statistical power than that of an effect for a single study in isolation (Lipsey & Wilson, 2001). Additionally, meta-analyses can allow researchers to assess possible moderators across studies, to further evaluate the relationship between the dependent and independent variables (Cooper, 2017).

The school-based consultation approach has a long history, and previous meta-analyses have been conducted in this area (e.g., Medway & Updyke, 1985). In order to further consolidate this body of knowledge, a history of consultation-focused meta-analyses was reviewed. More specifically, consultations targeting behavioral outcomes for school-aged clients were of interest. These results were used to identify areas where further investigation was needed and inform the selection of variables for the current dissertation.

Retrieval and Selection of Meta-Analyses

Studies for review were identified through a systematic search procedure. Databases used included *ERIC*, *PsycINFO*, and *Academic Search Ultimate*. The search terms *consultation, school, teacher,* and *meta-analysis* were employed to identify related literature. In order to be included in this literature summary, the studies must have (a) been written in English, (b) employed a systematic review and meta-analytic methods, and (c) evaluated school-based consultations with client behavioral outcomes.

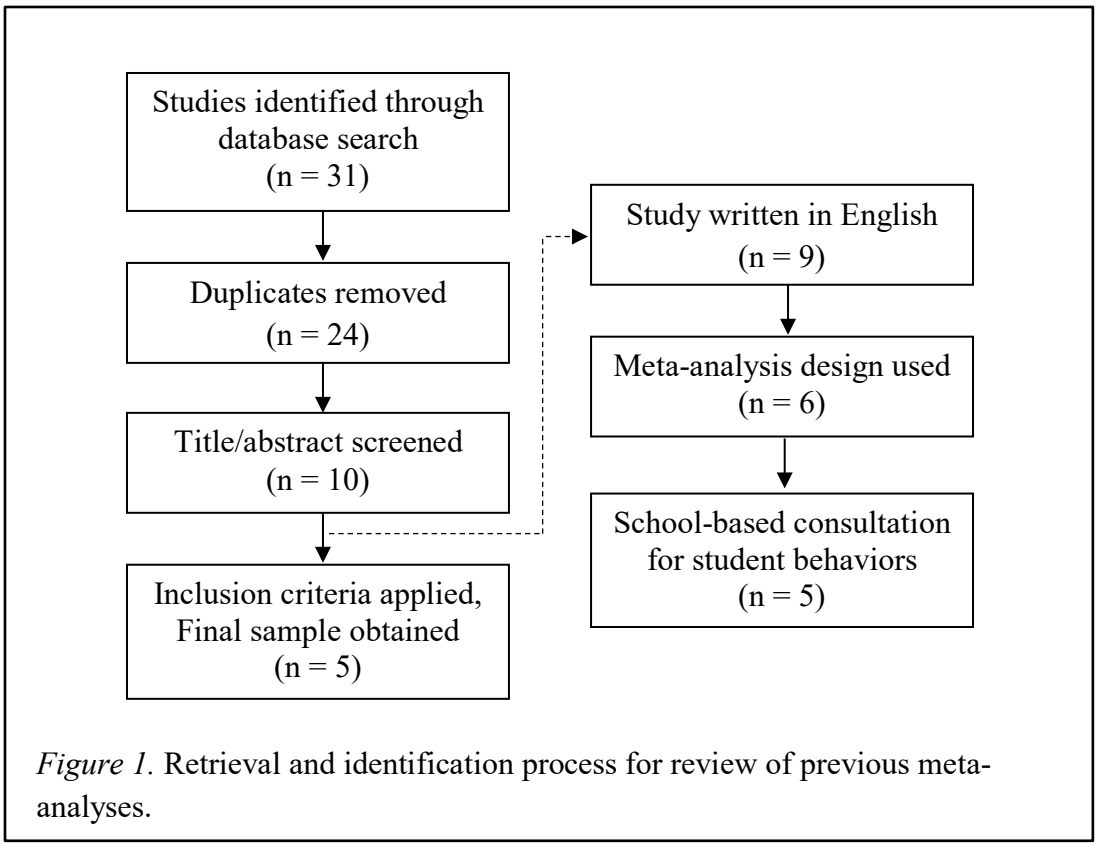

Figure 1. Retrieval and identification process for review of previous meta-analyses.

The initial search returned 31 studies, but only five of which met criteria for inclusion. Figure 1 provides additional information on the study retrieval and selection process. The results spanned a wide range of years, with studies from 1986 to 2012, the majority of which were unpublished dissertations (n = 4). All included meta-analyses evaluated group design studies; however, Jackson (1986) and Reddy and colleagues (2000) incorporated single-case designs as well. Three of the studies included both published studies and unpublished dissertations in their sample (Batts, 1987; Jackson, 1986; Sibley, 1986), and two considered only published literature (Davis, 2012; Reddy et al., 2000). Table 1 provides additional descriptive information about the collected research.

Table 1

Meta-Analyses Investigating School-Based Consultation

Author(s)	P/D	N Studies	Years Covered	Sample Type: Design, P/D	Outcome Source (ES)	Outcome Type (ES)	Overall Effect (95% CI)
Batts (1987)	D	40	Up to 1986	GD, P&D	CE (.77*) CL (.40*)	BEH (NR) ATT (NR) ACH (NR)	0.53* (0.40, 0.66)
Davis (2012)	D	19	1986-2009	GD, P	CL (.42τ)	ACH (.14τ) BEH (.50τ) OTH (.60τ)	0.42τ (0.16, 0.68)
Jackson (1986)	D	21	1974-1984	SCR&GD, P&D	BEH-CE (1.06*) BEH-CL (4.06*) ACH-CE (1.89*) ACH-CL (1.18*)		1.95* (1.13, 2.77)
Reddy et al. (2000)	P	35	1986-1997	SCR&GD, P	SYS (2.25*) CE (1.22*) CL (1.30*)	INT (.45*) BEH (1.49*) SS (0.50*) ACH (0.69*) MED (0.00*) CSA (2.29*) IK (0.58*) ATT (0.51*) SDR (0.86*) RSEP (0.29*) IUS (3.81*)	Not Reported
Sibley (1986)	D	63	Up to 1984	GD, P&D	CO (.90*) CE (.60*) CL (.91*)	BEH (.89*) ATT (.56*) ACH (.93*)	0.75* (0.59, 0.90)

Note. P=published, D=dissertation, N=sample size, GD=group design, SCR=single-case research, ES=effect size, CE=consultee, CL=client, CO=consultant, SYS=system-wide, ACH=achievement, BEH=behavior, ATT=attitude, OTH=other, INT=Internalized, SS=social skills, MED=medical, CSA=consultee skill acquisition, IK=increased knowledge, SDR=system decreased referrals, RSEP=reduced special education placement, IUS=increased use of services, CI=confidence interval

* Glass' Δ or a variant of this equation

τ Hedges' *g*

General Outcomes of Consultation

The identified meta-analyses evaluated progress not only on behavioral outcomes, but on a wide range of additional outcomes including attitude changes, academic and skill gains, social skills, and many others. Overall effect sizes thus represent the impact of consultation on a variety of outcomes, not only those related to behavioral goals. Overall effects ranged from a Hedge's *g* of .42 to a Glass' Δ of 1.95, indicating that the consultation process typically results in positive improvements.

Furthermore, the omnibus effects reported by each meta-analysis suggest consultation can be an impactful method for improving outcomes at various levels of analysis. In particular, researchers considered school-, consultant-, consultee-, and client-level changes and found positive impacts across all levels. These findings imply that not only did student-clients benefit from the consultation process, but so did teachers, consultants, and the school system at large. For example, Jackson (1986) found that not only did students increase their behavioral and academic skills, but teachers improved their ability to helpfully respond to students and demonstrated increased skill knowledge after the consultation concluded.

In contrast, when examining outcomes based on the type of change measured (i.e., the goal of the consultation), the results became more variable. Some skills showed marked increases and others demonstrated minimal or no change after the consultation process. For example, Reddy and colleagues (2000) found notable effects for most of the outcomes examined, but reported that consultations had no impact on client medical issues (e.g., number of doctor visits, health status) and a somewhat limited impact on the rate at which students were placed into special education within the school system. While consultation may improve skills for all

participants in the process, it appears to be better adapted to certain outcomes over others. Of specific interest for this review were the changes in behavioral outcomes.

Student behavioral outcomes.

When looking specifically at the development in behavioral skills across the different meta-analysis, positive changes were observed uniformly. In general, consultations appeared to have a significant effect on client behavioral outcomes, ranging from a Hedge's *g* of .50 to a Glass' Δ of 4.06 (Table 1). These positive improvements in student behavior suggest teachers are capable of effecting change when provided expert support through the consultation process. Furthermore, the range of behaviors featured within these samples indicates various behavioral goals would be appropriate for the consultation process. To illustrate, two meta-analyses reported the behavioral goals included in their sample (Jackson, 1986; Reddy et al., 2000); the most common goals were off-task behavior, aggression, and disruptive conduct or verbalizations.

Moderators Evaluated

In addition to effect sizes, the meta-analyses considered several moderators to determine if consultation outcomes were influenced by other factors. Although these moderators did not evaluate impacts on behavior goals specifically, instead looking at overall outcomes which included a wide range of goals, they were still of interest for informing best practices in general. First, the most frequently used moderator was the type of consultation model employed. The consultation models considered in these meta-analyses included the Mental Health, Behavioral, Organizational Development, Informational/Educational, Instructional, and Process Models. As is implied by the name, the Behavioral Model utilizes theory from behavioral psychology and is often used to consult with teachers serving students with problematic behaviors (Kratochwill & Bergan, 1990). No significant differences were found between the various types of consultations

on overall outcomes (Batts, 1987; Davis, 2012; Reddy et al., 2000; Sibley, 1986), with the exception of Jackson (1986) who found the Mental Health Model outperformed the Behavioral Model. In a more detailed analysis, Batts (1987) evaluated the impact of the various models on different client goals and found clients with behavioral goals typically benefited most from the Behavioral Model and less from the Process and Informational/Educational Models. These results suggest no model is inherently superior, but that the context, purpose, and goals of the consultation should be considered when choosing which model will a good fit for the client.

Second, the duration of the consultation was evaluated as a moderating variable. When considering the length of the consultation, it was questioned whether a longer consultation resulted in more significant outcomes. Batts (1987) was the only author to evaluate duration as a moderator and found it to have a non-significant impact. The author cautioned that these results may be confounded, because longer consultations were typically done as class projects, meaning the consultants were students with less training and the length of the consultation was determined by the class, rather than the client's progress. These results should consequently be replicated before conclusions can be made.

Third, client-level variables were considered, namely age and gender. Three of the authors examined the relationship of client-age to outcomes and the results were inconclusive. Between these three studies, it was determined: (a) there was no difference in consultation results by age (Sibley, 1986), (b) younger children demonstrated higher outcomes (Davis, 2012), and (c) older children benefited at higher rates (Reddy et al., 2000). Based on these disparate findings, it is difficult to hypothesize if the child's age impacts consultation outcomes and in which direction. Again, further research is needed, which would help to determine if these differences are the result of a sampling artifact or if the relationship of age with consultation outcomes is

complicated, requiring it to be placed in the context of other moderators. One such variable that might produce interesting results when evaluated in conjunction with age is gender (Reddy et al., 2000). Within this meta-analytic sample, only one study included client-gender as a variable and found females outperformed males (Reddy et al., 2000). It should be cautioned that the gender moderator was analyzed at the study-level and only two female-majority studies and four male-majority studies were identified. These small group sizes shed some doubt on the reliability of the results.

Fourth, Davis (2012) considered a wide array of other moderating variables (i.e., type of class, consultant type, school type, referral source, referral reason, comparison group, intervention type, design quality, outcome measured, and data type), none of which resulted in significant differences in the dependent variables. However, his evaluation of consultee and consultant features was particularly interesting. These included whether the consultee was a general or special education teacher and if the consultant was still in graduate school training or an experienced practitioner in the field. As mentioned above, these characteristics did not impact student outcomes. More research should be done to verify these findings, but they are a promising indicator that individuals of different backgrounds and various levels of experience can effectively improve student outcomes through the consultation model.

Gaps in Previous Meta-Analyses

There are several concerning gaps in the previous meta-analytic literature. One such gap is the frequent exclusion of single-case designs, including the most recent update to the meta-analytic literature base (Davis, 2012). Although there were strong methodological reasons for this exclusion, the absence of single-case designs is a serious gap in the literature because this method is a recognized approach for identifying evidence-based practices (Horner et al., 2005).

Those previous meta-analyses including single-case studies are somewhat dated, as a span of almost two decades has passed (Jackson, 1986; Reddy et al., 2000). There is a need to update the literature and include more current single-case studies to determine if their results are in accord with the rest of the consultation literature.

Part of the reason for this exclusion may have been due to the relatively recent development of practices for applying statistical methods to single-case designs, and as such, the best practices in this area are still being debated by experts in the field (Horner & Kratochwill, 2012). Meta-analyses in particular are well suited to group designs, but questions have arisen about how to best adapt these procedures for single-case studies. While some researchers argue in favor of a nonparametric method, others support a parametric effect size calculation. These concerns are covered in depth by Maggin, Swaminathan, Rogers, O'Keeffe, Sugai, and Horner (2011) in their article discussing the nuances of the issue.

As Maggin reports, nonparametric effects are beneficial, because they do not rely on the assumption that the data has a normal distribution, something which can rarely be assumed in single-case research. However, these methods are also not appropriately sensitive to trends in the data, outliers, or the magnitude of the effect, possibly resulting in a misleadingly large or small effect size. In comparison, parametric effects are more sensitive to variations in the data, but were created for group designs. Translating parametric effect sizes for use with single-case research often results in various issues, problems that could again result in misleading effect size interpretations (Maggin, Swaminathan, et al., 2011). Thus, authors of meta-analyses featuring single-case designs may consider using a combination of these two methods; however, this has not yet been done in the school-based, behavioral consultation literature.

In addition, none of the meta-analyses which utilized single-case studies implemented What Works Clearinghouse (WWC) quality indicators (Kratochwill et al., 2010). Implementation of these standards can help to evaluate whether a study demonstrated adequate experimental control, thereby allowing the researcher more confidence in their results. When conducting meta-analyses, the WWC standards can be used as a method for screening prospective studies to determine if they meet minimum design standards and should, therefore, be incorporated into the final sample (Bowman-Perrott, Burke, Zhang, & Zaini, 2014). When these standards are not used, there may be more questions or concerns about the validity of the results.

The second major gap is related to the gray literature, which has often been excluded in meta-analyses. Two of the identified meta-analyses evaluated the differences between the published and unpublished literature within their sample (Batts, 1987; Sibley, 1986). Both found evidence of significant publication bias. This is a cause for caution when interpreting the findings of the studies which excluded dissertations, as they may be inflated due to bias. Additionally, dissertations have not been included in the meta-analytic literature since the 1980s, and thus a large body of research has gone unreported in more modern analyses.

The final gap relates to the issue of treatment integrity. Although fidelity issues have been discussed as a concern when conducting consultations, few authors of school-based consultation studies have attempted to evaluate this problem, thus knowledge about treatment fidelity is limited (Noell, 2008). This lack of investigation into treatment fidelity is reflected in the meta-analytic literature. Batts (1987) noted few studies in his sample measured the integrity with which the teacher implemented the intervention, making it is unclear if student outcomes were influenced by low treatment fidelity. Davis (2012) also recommended treatment integrity

be considered as a moderator in future research; however, this variable has not yet been incorporated in the meta-analytic literature. The relationship between treatment integrity and behavioral outcomes appears to be a complicated one, but research has suggested when treatment integrity is low, so are the effects of the intervention (Sanetti & Kratochwill, 2007). If the intervention is not being implemented with fidelity across studies, this raises concerns about the validity of the findings within a meta-analysis.

Performance feedback may be a practical method in the field for measuring and improving low treatment fidelity during consultation. Performance feedback allows a consultant to observe and systematically monitor a consultee's treatment integrity and meet regularly to correct issues as they occur (Sanetti & Kratochwill, 2007). This is a moderator which has not yet been assessed in the consultation meta-analysis literature, and could be considered in future research to better address concerns related to treatment integrity, as well as provide information about best-practice when addressing fidelity concerns in practice.

Purpose and Research Questions

There is a need to synthesize the literature on the use of school-based behavioral consultation models to improve externalizing behaviors of students. Previous research suggests consultation can be a useful intervention method for helping teachers respond to behavioral concerns. However, there remain several questions regarding the best way to provide technical assistance to teachers using consultation models. In particular, evaluating moderators such as performance feedback could inform best practice and help improve the efficacy consultations. In addition, there are several gaps in the current literature, such as the exclusion of dissertations, single-case studies, and WWC design-standards.

The purpose of this study was to fill these gaps and evaluate the effectiveness of using school-based behavioral consultation to improve externalizing problem behavior. Two research questions were considered: (a) What are the overall effects of school-based behavioral consultations on student externalizing behaviors? (b) Do performance feedback procedures moderate the relationship between consultation and externalizing behavioral outcomes?

3. METHOD

Search Process and Study Selection

Articles from 2016 and earlier were incorporated into this dissertation. To identify relevant studies for analysis, several databases were searched, including *ERIC*, *PsycINFO*, *Academic Search Complete*, and *ProQuest Dissertations & Theses Global*. Search terms included: *school*, *teacher*, *student*, and variants from the stems *behav-* and *consult-* (e.g., behavioral, behavior, consult, consultant, consultation). This search returned a total of 8,861 articles and dissertations. Figure 2 provides additional information on the study retrieval and selection process.

An initial screening procedure was conducted, in which the title and abstracts of each article were read and any articles that clearly did not meet inclusion criteria were eliminated. In addition, duplicates were removed at this stage. The inclusion criteria, as described below, were then individually applied to each of the articles and dissertations to determine if they met the requirements of the meta-analysis.

Additionally, an ancestral search was performed, to find studies missed during the search process. This involved examining the references of the articles which had meet inclusion criteria for related literature and identifying authors whose names appeared more than once in the list of included studies so their vita's could be scrutinized for related research. Once complete, this process resulted in a total of 26 studies in the final sample.

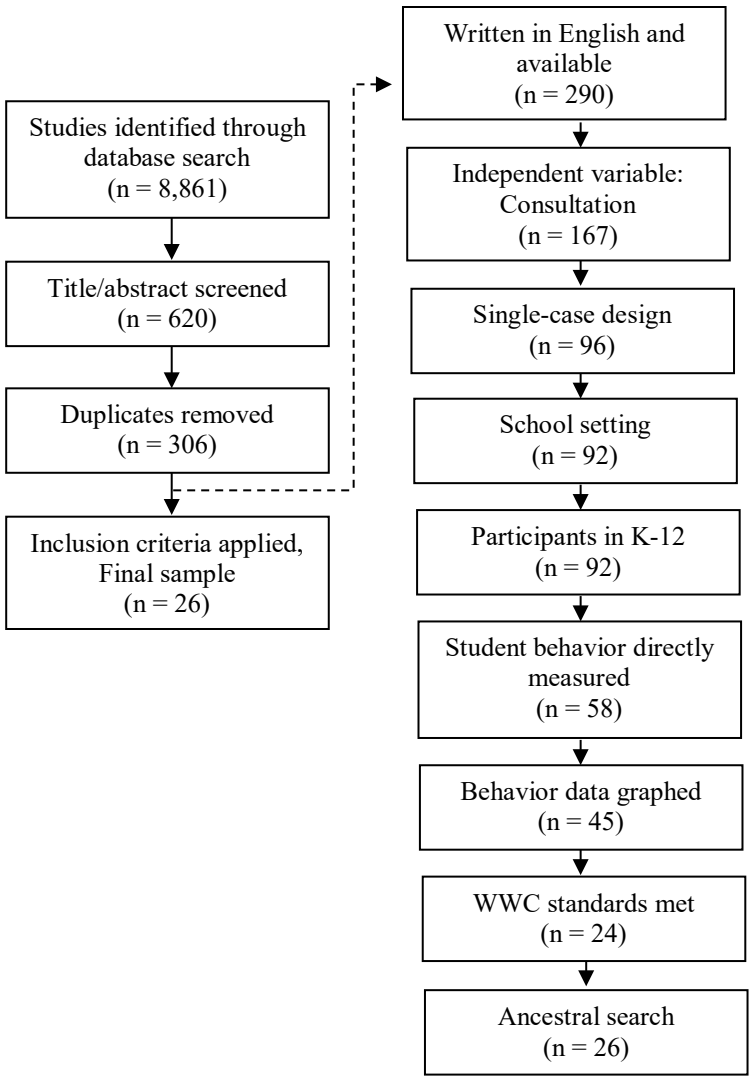

Figure 2. Retrieval and identification process for dissertation sample.

Inclusion Criteria

Studies were required to meet the following criteria to be included. Each study reviewed for inclusion must have: (a) been available in English, (b) included behavioral consultation as an independent variable, (c) used a single-case design, (d) occurred in the school setting, (e) included school-aged clients (grades K-12), (f) directly measured externalizing student behavior

outcomes, (g) featured graphed behavioral data, and (h) met WWC quality indicators. These criteria are discussed in-depth below.

English and availability. When coding for inclusion criteria it was first determined if the article was available in English. English was the primary language spoken by the author of the meta-analysis and the readership of this study. Thus, articles written in another language were excluded from the sample. Additionally, the studies had to be available for the current author to read. In rare cases, a dissertation had been purposefully removed from the public domain or copies of older documents were no longer available.

Consultation as an independent variable. Studies were examined to determine if a behavioral consultation process was implemented as an independent variable. Behavioral consultation was defined as a service featuring a consultee and consultant team using an indirect problem-solving process to improve client behavior. This definition was based on the work of Kratochwill and Bergan (1990). A more specific definition could have included key features of the Behavior Model, such as the four-step process detailed by Kratochwill; however, for purposes of inclusion, a more general definition was chosen, so as not to exclude articles that did not provide high levels of detail regarding their consultation procedures. Additionally, a study would occasionally use the word "consultation" to refer to a teacher training session. For example, the authors might have a specific intervention they wished to train teachers to use and would meet with school staff for the express purpose of training them to use that strategy. In this situation, a problem-solving process was not used; therefore, these studies were excluded.

Single-case design. Only single-case studies were included. Single-case studies are those that demonstrate experimental control using the same individual as both the control and

treatment participant (Kennedy, 2005). With the application of this criterion, group designs, descriptive designs, and qualitative studies were excluded.

School setting and school-aged participants. The setting of the consultation must have included a type of school. The definition of "school" was expanded to include public, private, charter, or alternative schools. Furthermore, the client of the consultation needed to be a grade-school student: an individual attending kindergarten through twelfth grade. The purpose of this criterion was to exclude pre-school or post-secondary settings, which were eliminated in an effort to reduce the potential heterogeneity of the sample. Students were included from both general and special education and no disability was required to meet this criterion.

Directly measured student behavior outcomes. At least one of the dependent variables in the study was required to be client behavioral outcomes. This included any positive or negative behaviors which were directly observed. In an attempt to avoid measurement bias, the direct observation requirement eliminated studies that used retrospective or rating-scale measurements. To be considered "directly observed," a behavior must have been monitored either in person or by means of a video recording, and data collected during this observation period.

Graphed behavioral data. Graphed student behavior was required. To be more specific, the collected data needed to be displayed in a line graph format. This criterion was necessary, as graphed data was essential when conducting the effect size analysis.

What Works Clearinghouse quality indicators. The studies were screened using WWC quality indicators for single-case design (Kratochwill et al., 2010). The WWC quality indicators were used to ensure each study within the sample met basic quality standards. There are three levels at which a study could be classified: (a) *Meets Design Standards,* (b) *Meets*

Design Standards with Reservations, or (c) *Does Not Meet Design Standards.* Studies were classified according to the process detailed by Maggin, Chafouleas, Goddard, and Johnson (2011). For a study to *Meet Design Standards*, it must have: (a) systematically manipulated the independent variable, (b) measured the dependent variable over time and reported interobserver agreement (IOA) of no less than 80% for at least 20% of the sessions, (c) made a minimum of three attempts to demonstrate an intervention effect (e.g., multiple baseline designs with at least three baseline-treatment conditions, ABAB designs), (d) and reported at least five data points for each phase.

In contrast, studies which *Met Design Standards with Reservations* had a minimum of three data points for each phase and while they did report IAO at acceptable levels, they did not report IOA for at least 20% of the sessions. Studies that *Did Not Meet Design Standards* were excluded from the sample. See Table 2 for information about the number of studies that met WWC criteria within this sample. Of the 45 studies that were eligible for the WWC screening procedures (i.e., had passed all other inclusion criteria), 7 met standards, 19 met with reservations, and 19 did not meet standards.

Table 2

Application of WWC Quality Indicators. Number of Studies (Percentage).

WWC Criteria	Meets	Meets With Reservations	Does Not Meet	Total*
Criteria A: IV systematically manipulated	44 (98%)	NA	1 (2%)	44
Criteria B: DV measured over time and IOA criteria met	19 (43%)	14 (32%)	11 (25%)	34
Criteria C: Minimum of three attempts to demonstrate effect	30 (88%)	NA	4 (12%)	30
Criteria D: Required number of data points reported	13 (43%)	13 (43%)	4 (13%)	26

Note. IV=independent variable, DV=dependent variable, NA=not applicable

* Total number of articles that passed the given criteria

Variable Coding

Once a study passed the above inclusion criteria, descriptive variables of interest were identified and methodically coded. Additionally, it was at this stage that each study was evaluated for the presence of the moderating variable, performance feedback.

Descriptive variables. Several study-level descriptives were of interest, such as whether the study was a published article or unpublished dissertation. The sample size of the consultant, consultee, and client participants in each study were coded for as well. Client-level variables including gender, grade, and disability status were noted. Furthermore, details about the specific intervention used and behavioral outcome measured were recorded for descriptive purposes.

Moderator variable. Coding procedures were created for judging the presence or absence of performance feedback, the moderator variable in this dissertation. Performance feedback was defined as a procedure in which the consultant provided repetitive, data-based feedback to the consultee about their level of treatment integrity (Reinke, Lewis-Palmer, & Martin, 2007). This was a dichotomous variable, with studies coded as either including performance feedback or not including it.

Inter-Coder Reliability

The reliability of the study selection and variable coding process was verified through the use of an additional coder. An Excel file was created with a manual for coding procedures, featuring the above inclusion criteria and definitions for the descriptive and moderating variables. The author of this paper then trained an additional coder to correctly use the manual. Both coders reviewed the definitions for each variable and practiced coding until they achieved 100% inter-coder reliability. The second coder independently coded 20% of the studies, which were randomly chosen from the larger sample. Reliability was then calculated using both

percent agreement and Cohen's kappa, as according to the recommendations and procedures outlined by McHugh (2012).

The coders initially had an overall inter-coder reliability of 83.3% ($\kappa = .68$); however, when looking at individual inclusion criteria or coded variables, reliability ranged widely, from 50% to 100% ($\kappa = .24$ to 1). In order to address this issue, the coders met to discuss the areas where reliability fell below 80% and determine if agreement could be reached. After this discussion, reliability increased to 99.29% ($\kappa = .99$).

Calculating Effect Sizes

Two different effect size estimation methods were used, both a nonparametric and parametric method. The nonparametric approach was utilized to obtain an overall effect size by means of Baseline Correct Tau. In contrast, a parametric multilevel modeling methodology was employed to investigate the performance feedback moderator, as well as to conduct a secondary analysis of overall significance.

Data Extraction. In order to conduct effect size calculations, the data from the sample studies first had to be prepared for analysis. Each applicable graph from the studies was uploaded as a JPG file. These images were then opened in *GetData Graph Digitizer 2.26* (Fedorov, 2013), a software program used to digitize graphs. *GetData* was used to accurately identify the x- and y-values for each data point in the graphs. The y-values for baseline and treatment phases were the basis on which effect sizes were then computed.

Some studies in this sample measured increases in appropriate student behaviors, while others measured decreases in inappropriate behaviors. For instance, one study might measure an increase in on-task behavior, while another measured a decrease in off-task behavior. In the first instance, a positive effect would indicate the consultation had the intended impact, but in the

second, a negative effect would be expected. To unify the metric, all effects were coded so that a positive effect size represented an improvement in behavior after treatment. This was accomplished by changing the sign for inappropriate-behavior data (i.e., a negative y-value was transformed into a positive one, or a positive value into a negative one). These transformed y-values were then used for effect size calculations.

Baseline Corrected Tau. Baseline Corrected Tau (BC-Tau) was calculated (Tarlow, 2017) to evaluate the overall impact of consultation on externalizing student behavior. This variant of Tau retains a few advantages over other nonparametric effect size calculation methods. Namely, it does not exceed the conventional lower and upper bounds of a correlation (i.e., -1 and 1), is distribution free (i.e., not assuming either a linear relationship or normal distribution, or requiring interval-level data), and is robust to autocorrelation (Tarlow, 2017). BC-Tau was calculated using the *Baseline Corrected Tau Calculator* created by Tarlow (2016). Data from all baseline-treatment (i.e., AB) phases were analyzed, excluding follow-up or maintenance phases. In studies where more than one AB phase was reported, a separate BC-Tau was calculated for each AB contrast (i.e., A1/B1 and A2/B2).

Each BC Tau result was transformed into a Fisher's *z* score before averaging these effects together using the methods described by Walker (2003), and a corresponding standard error (SE) was calculated using the formula: $SE=1/\sqrt{(n-3)}$, where *n* is the total number of data points within the combined baseline and treatment phases (Lehmann, 2011). The variance of correlation-based effect sizes, such as BC-Tau, depends strongly on the correlation itself. Thus, this conversion process was done so that when effect sizes were averaged together, the newly calculated Fisher's *z* variance could be used as a more uniform and pure metric, not impacted by the strength of the correlation, but only by the number of observation points within the AB phase

(Borenstein, Hedges, Higgins, & Rothstein, 2009). When averaging effects, the Fisher's *z* score was weighed by the inverse of its variance; in other words, this method of calculating the SE penalized AB pairs with fewer data points, irrespective of their effect size.

These weighted averages were obtained by use of the *WinPepi* program (Abramson, 2011). Specifically, the following commands were chosen: (a) Compare2, (b) Meta-Analysis, (c) Other, and (d) Standard Error. The Fisher's *z* results and SE's were then entered into the program to acquire an average effect size for each student. Student-level effects were combined to gain overall averages for each study, and then these study-level effects were averaged to obtain the omnibus effect, using the same *WinPepi* commands. A random effects model was chosen for these calculations, as student characteristics and goals varied and thus the true effect was conceptualized to differ between participants and studies. Finally, each averaged effect and its corresponding standard error was converted back to BC-Tau for presentation purposes within this paper. Confidence intervals for individual effects were also calculated to provide additional information about the reliability of these estimates.

Multilevel modeling. A parametric approach using multilevel modeling was employed to assess whether the presence of performance feedback moderated the relationship between consultation and student behavioral outcomes. The use of multilevel modeling, following the recommendations of Baek and colleagues (2014), allowed for an additional omnibus effect to be calculated, as well as more sophisticated methods of evaluating the proposed moderating variable. Only 21 of the 26 studies were included in this analysis. While the majority of the studies reported student outcomes in percentages (e.g., percentage of on-task behavior measured in 20-second intervals), five of the studies reported results as frequency counts. With the above nonparametric methods, study results could be converted into a common metric before being

compared (i.e., BC-Tau and Fisher's *z*). In contrast, when using MLM methods, individual data points are compared across studies and must therefore be presented in the same metric for analysis at the observation level. Thus, the studies that were not reported as percentages were removed from the sample for this procedure. The studies excluded were: Angell (2005), Barnhouse (1979); Beckman (2008), Borenstein, et al. (1977), and Mendicino (2015).

Three levels of data were considered: (a) level-1, or the observation level, which consisted of individual data points, (b) level-2, or the level where individual student results could be compared, and (c) level-3, or the level where results could be compared across studies. Figure 3 provides an illustration of how these levels are conceptualized. In order to assess the data at these various levels, a series of four models were developed and the *SAS 9.4* software was used to run each model (SAS Institute Inc., 2015).

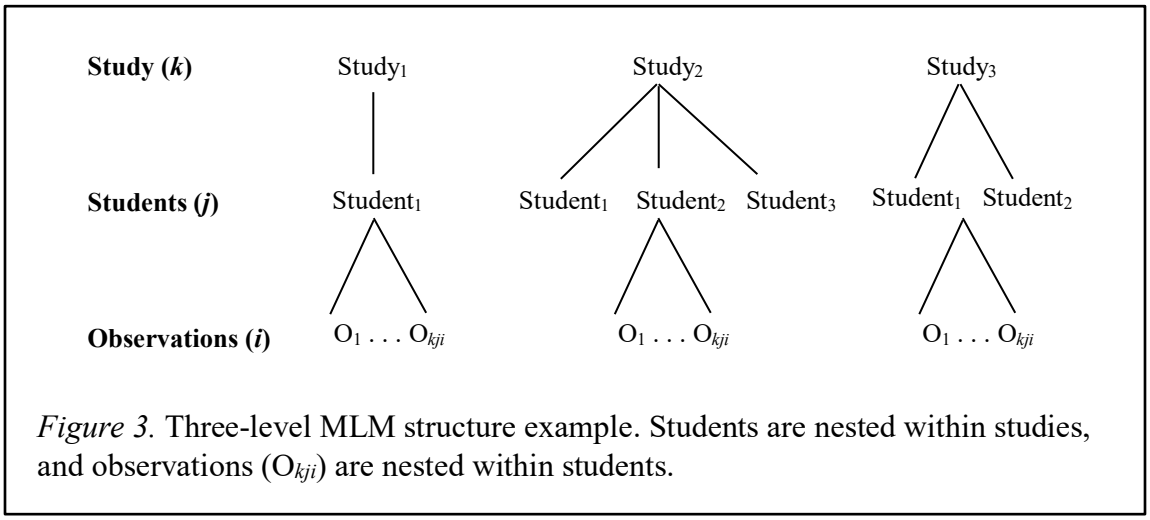

Figure 3. Three-level MLM structure example. Students are nested within studies, and observations (O_{kji}) are nested within students.

The first model assessed, Model 1, was created to provide results for calculating an Intraclass Correlation Coefficient (ICC). This statistic provided information about how much of the variation in behavioral outcomes was explained by differences between students or studies. The more variation explained by these differences, the more reason to investigate what variables might be moderating these differences. The following ICC equations were used (Bell, Ene, Smiley, & Schoeneberger, 2013):

$$\text{Student-Level ICC} = \frac{\sigma^2_{Student}}{\sigma^2_{error} + \sigma^2_{Student} + \sigma^2_{Study}}$$

$$\text{Study-Level ICC} = \frac{\sigma^2_{Study}}{\sigma^2_{error} + \sigma^2_{Student} + \sigma^2_{Study}}$$

Next, Model 2 incorporated the dummy-coded variables of "phase" and "AB-pair." Phase indicated whether a data point was observed in baseline or treatment (i.e., 0 = baseline, 1 = treatment), while AB-pair was created to denote which baseline-treatment data points were paired. This was particularly important if a student had multiple AB-pairs for a behavior (i.e., a reversal design), or if more than one behavior was observed for a single student. For instance, a student might be assessed for off-task behavior, aggression, and academic engagement. In such a situation, each baseline-treatment condition for the various behaviors would receive its own dummy coded number for the AB-pair variable (i.e., 1 = off-task, 2 = aggression, 3 = academic engagement). This allowed the model to account for the relationship between these baseline-treatment pairs.

In the second model, the error and intercepts were allowed to fluctuate, but the slopes were held constant. Meaning that variation at baseline was acknowledged, as well as differences related to random error; however, the shift in behavior from baseline to treatment (i.e., slope or treatment effect) was assumed to be uniform. In comparison, Model 3 implemented a random slope model, thereby allowing for differences in treatment effect between students and studies.

Finally, Model 4 was developed to incorporate the previously discussed variable of performance feedback. This last model was used to examine what impact the addition of this potential moderator had on consultation effects. It was with this model that the second research question could be answered.

Throughout this process, autocorrelation was controlled for using the first-order autoregressive method (i.e., AR(1)) discussed by Baek and Ferron (2013), using a non-varying model. This was accomplished in *SAS* through a "repeated" command: "repeated / subject=student type = AR(1)". Model fit was continually checked as well, to ensure the most appropriate model was chosen for analysis. Model fit indices included the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), two model fit statistics commonly used in MLM analyses (Bell et al., 2013). For coding examples and model syntax please see Appendix A and B respectively.

Assessing Bias and Homogeneity

To assess for sampling issues, a funnel plot was created. A funnel plot is a scatter plot, with one axis representing the effect size and the other signifying some measure of the effect's precision (i.e., BC-Tau and its corresponding standard error; Sterne et al., 2011). Skew in a funnel plot can represent various problems such as bias, heterogeneity, and other sampling issues (Sterne et al., 2011). Asymmetry in a funnel plot is therefore undesirable. An Egger's test was conducted in *WinPepi* to gain a statistical estimate of skew (Egger, Smith, Schneider, & Minder, 1997).

Furthermore, the H and I^2 statistics, of Higgins and Thompson (2002), were run in *WinPepi* to specifically assess heterogeneity. The presence of significant heterogeneity suggests the effect size of a meta-analysis may be impacted by differences between studies, rather than simple random error. In cases where heterogeneity is high, a random effects model can be used to help account for these differences (Sterne et al., 2011).

4. RESULTS

Descriptive Statistics

The 26 studies used in this analysis ranged on several variables. The earliest investigation was conducted in 1970 and research continued up through 2016. A majority of the studies were unpublished dissertations (n = 15, 58%), but published works still made up a considerable portion of the sample (n = 11, 42%). Only two design types were used: multiple baseline (n = 22) and reversal designs (n = 4). Table 3 provides additional information about the individual studies in this sample.

When examining student-level characteristics, several trends emerged. The sample was predominantly male. Although one study did not report student gender, the remaining research included 81 males and 22 females. Most of the students were in elementary school (n = 95), but there were a few secondary students as well (n = 14). Only 12 studies noted the disability status of the participants. The most common disabilities were Autism Spectrum Disorder (n = 11), Attention Deficit/Hyperactivity Disorder (n = 6), and Learning Disability (n = 5). A majority of the consultees were teachers (general education teacher: n = 39, special education teacher: n = 12, teacher's aide: n = 1, certification unreported: n = 61), but there were a few parents involved as consultees as well (n = 21).

During the consultation process, student behaviors frequently targeted included off-task behavior, aggression, inappropriate verbalizations, out-of-seat behavior, and noncompliance. Interventions used to address these concerns involved contingency plans, praise, self-monitoring, and peer-mediated strategies, among others.

Table 3
Studies in Meta-Analytic Sample

Author (Year)	Published	N CL (CO, CE)	CL Age(s)	Target Behavior(s)	Intervention(s)
Angell (2005)	No	6 (1,18)	9-13	Throwing objects, perseveration, getting out of seat, tearing up work	Visual/verbal cues, contingencies
Barnhouse (1979)	No	6 (5,6)	5-11*	Non-attending	Not Reported
Beckman (2008)	No	3 (1,6)	8-12*	Blurt outs, talking back, out of seat, noncompliance, off-task, verbal/physical aggression	Visual/verbal cues, contingencies
Bellinger, Lee, Jamison, & Reese (2016)	Yes	3 (1,9)	6-7	Noncompliance, off-task, inappropriate body movements/noises, lack of problem solving	Praise, peer-mediated, self-monitoring, school job, contingencies, breaks, choices, visual schedule, simple directions, timers
Bice-Urbach & Kratochwill (2015)	Yes	6 (1,6)	5-10	Tantrums, out-of-seat, inappropriate verbalizations, physical aggression, inattention	Setting, antecedent, & consequence strategies
Bornstein, Hamilton, & Quevillon (1977)	Yes	1 (1,1)	9	Out-of-seat behavior	Contingencies, positive practice, self-monitoring
Coffee (2007)	No	4 (1,4)	7-9	Off-task behaviors	Contingent praise
Colton & Sheridan (1998)	Yes	3 (1,6)	8-9	Self-control, inappropriate peer interactions	Coaching, role play, self-monitoring, contingencies, home-school communication

Note. N=sample size, CL=client, CO=consultant, CE=consultee, NR=not reported
* Age estimated based on provided grade-levels

Table 3

Studies in Meta-Analytic Sample

Author (Year)	Published	N CL (CO, CE)	CL Age(s)	Target Behavior(s)	Intervention(s)
Dee (1998)	No	4 (1,4)	8-12*	Negative/aggressive behaviors	Debriefing, problem solving, goal-setting
Emery (1986)	No	3 (1,3)	15-17	Off-task behaviors	Contingent praise
Engelhardt, Sulzer, & Altekruse (1971)	Yes	1 (1,1)	6	Out-of-seat behavior	Contingent attention
Garbacz & McIntyre (2016)	Yes	3 (3,6)	6-8	Tantrum, off-task, noncompliance	Antecedent/instructional/consequence strategies, home-school notes
Geil (1998)	No	8 (5,8)	6-11*	Disruptive or off-task behavior, self-stimulation, inappropriate play/verbalizations/locale	Not Reported
Hazzard (1977)	No	3 (1,3)	6-7	Out-of-seat, disruptive behavior, aggression, off-task, blurt outs, following directions	Praise, contingencies
McPhail (2005)	No	6 (1,6)	6-8	Off-task behaviors	Praise, pre-correction, assignment modification, peer modeling, replacement behaviors
Mendicino (2015)	No	1 (1,1)	11	Touching hair, flapping, screaming	Prompts, contingencies, teacher/material proximity, teaching strategies
Meyers (1975)	Yes	1 (1,1)	8-9*	Out of seat, inappropriate verbalizations	Planned ignoring, non-emotional discipline, contingent attention
Munton (2004)	No	9 (1,12)	6-11*	Disruptive behaviors, out-of-seat, noncompliance, physical/physical aggression	Praise, contingencies, pre-correction

Note. N=sample size, CL=client, CO=consultant, CE=consultee, NR=not reported

* Age estimated based on provided grade-levels

Table 3
Studies in Meta-Analytic Sample

Author (Year)	Published	N CL (CO, CE)	CL Age(s)	Target Behavior(s)	Intervention(s)
Rahn (2008)	No	9 (NR,9)	5-9*	Inappropriate verbalizations, out-of-seat, off-task	Group/individual contingencies, self-monitoring
Robbins (1990)	No	3 (1,3)	7	Off-task, work completion/accuracy	Praise, contingencies, self-monitoring, reduced environmental distractions, changed schedule, peer mediated interventions
Sanetti, Collier-Meek, Long, Byron, Kratochwill (2015)	Yes	4 (2,4)	7-8	Off-task behaviors	Antecedent manipulation, teaching strategies, contingencies
Shipley (2013)	No	12 (1,6)	5-11*	Off-task behaviors	Praise, contingencies (class-wide/individual), behavior contract, self-monitoring
Sterling-Turner, Watson, & Moore (2002)	Yes	4 (1,4)	6-16	Self-injury, inappropriate verbalizations, incomplete work, out-of-seat, off-task, noncompliance	Praise, planned ignoring, prompts
Tillman (1999)	No	3 (1,3)	7-8	Off-task behaviors	Contingent praise
Whitley & Sulzer (1970)	Yes	1 (1,1)	9	Out-of-seat, inappropriate verbalizations	Praise, contingent attention, planned ignoring
Wilkinson (1997)	Yes	3 (1,3)	6-9	Disruptive behaviors	Contingencies, behavior contract

Note. N=sample size, CL=client, CO=consultant, CE=consultee, NR=not reported

* Age estimated based on provided grade-levels

Research Question One: Nonparametric Effect Results

To answer the first research question, regarding the impact of consultation on student behavioral goals, the methods for calculating BC-Tau were implemented and an omnibus effect of .46 was obtained, with a 95% confidence interval extending from .34 to .58 (Pearson's r = .66, Cohen's d = 1.76). Individual study effects ranged from a BC-Tau of -.15 to .8. These results are displayed by means of a forest plot diagram in Figure 4, along with their corresponding 95% confidence intervals. Higher effects represent greater positive improvements in student behavior after consultation. Visual analysis of this forest plot reveals several outliers, but in general the individual study effects appear to cluster around the omnibus effect.

Research Question Two: Parametric Effect Results

The second research question was answered through a multilevel modeling approach, used to evaluate the impact of performance feedback on the relationship between the consultation process and behavioral outcomes. After the first model was run (as detailed in the Method section), an ICC was calculated for the study- and student-levels (.86 and .04 respectively). These results suggest a high level of variability in student behavioral outcomes was accounted for by differences between studies (i.e., 86%). The ICC results justified a hierarchical model that would investigate the differences between studies further; thus, the additional models were assessed. Table 4 displays the results of this analysis.

Model fit results were used to determine the best fitting model. For both the AIC and BIC indicators, a lower number suggests a better fitting model. Base on model fit indices, Model 4 was the best fitting model, and the results of this model were therefore used to assess the impact of each variable on student behavior.

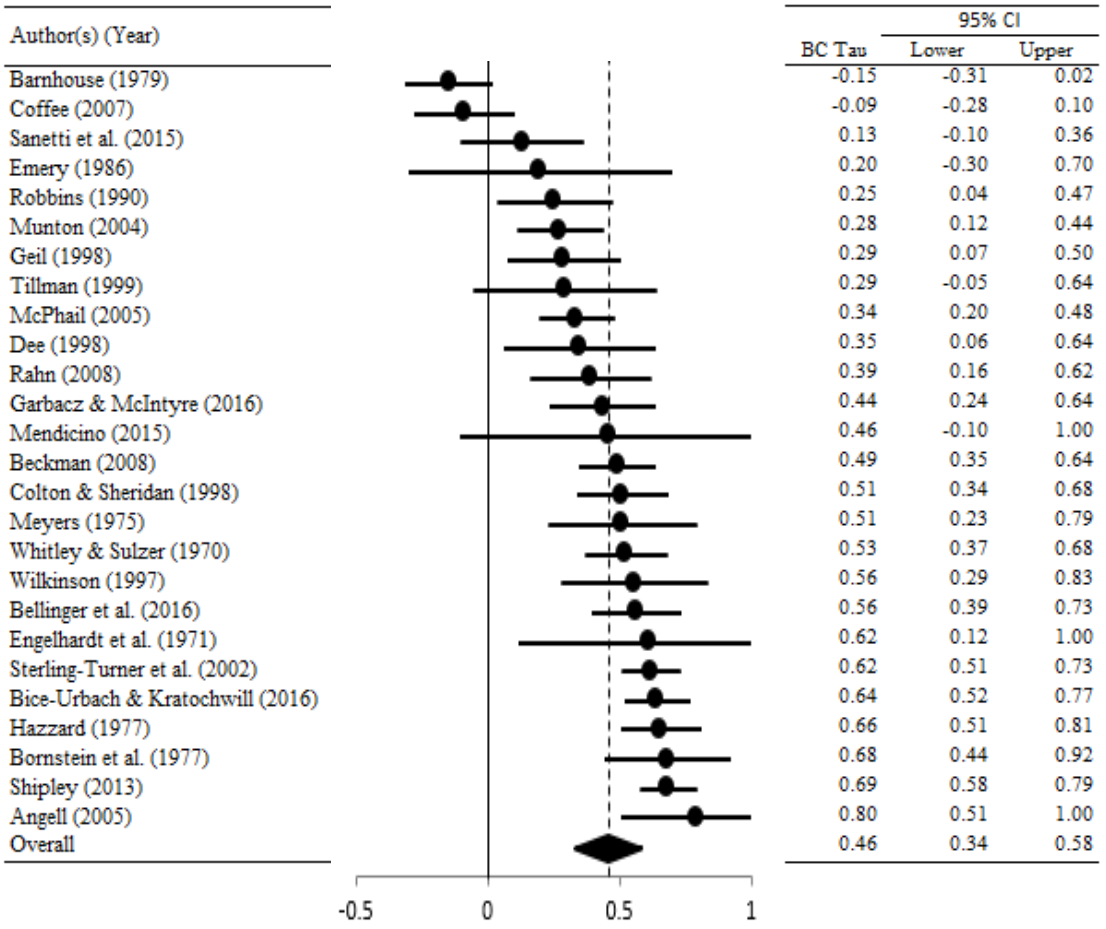

Figure 4. Forest plot of individual study effects and omnibus effect. Corresponding confidence intervals (CI) included.

Table 4

Multilevel modeling results. Displayed as parameter estimates (standard errors).

	Model 1	Model 2	Model 3	Model 4
Fixed Effects				
Intercept	22.36*	12.69	11.15	9.15
(Standard Error)	(9.47)	(12.31)	(13.00)	(13.13)
Phase		16.96*	16.71*	16.51*
(Standard Error)		(1.21)	(1.76)	(1.78)
Performance Feedback				5.45*
				(2.09)
Variance Components				
Level-1 Residual	1047.88*	656.69*	197.04*	195.98*
(Standard Error)	(69.73)	(36.18)	(6.39)	(6.35)
Autocorrelation	0.79*	.72*	0.18*	0.17*
(Standard Error)	(0.01)	(.02)	(0.02)	(0.02)
Random Intercept (Student)	25.08	32.11	16.06	17.62
(Standard Error)	(28.22)	(20.10)	(17.19)	(18.20)
Random Intercept (Study)	1776.55*	1913.13*	834.00	875.50
(Standard Error)	(583.13)	(635.77)	(759.85)	(772.83)
Random Slope (Phase-Student)			10.15	9.45
(Standard Error)			(8.04)	(7.93)
Random Slope (Phase-Study)			47.30*	48.95*
(Standard Error)			(20.03)	(20.45)
Model Fit				
AIC	21440.3	20912.7	20009.7	19999.6
BIC	21432.3	20904.7	19993.7	19983.6

The fixed intercept examines the average amount of appropriate behavior at the start of the baseline phase (i.e., 9.15, meaning that during about 9% of observed intervals, students could be expected to exhibit appropriate behaviors at baseline on average). The random intercepts provide insight into the amount of variation in student behavior at baseline for both the student-level (17.62) and study-level (875.5). While the random intercepts were initially significant in Model 1, by Model 4 neither were statically significant, suggesting that there was no notable variation between students or studies at baseline when other variables were included into the model.

The "phase" variable, which measures the treatment effect (i.e., change in level between baseline and treatment phases) was positive and statistically significant ($p < .0001$). This result indicates that after the implementation of the consultation process, student behavior was immediately and notably improved. The parameter estimate of 16.71 suggests student behavior improved by about 17% at the start of consultation. These results demonstrate not only statistical significance, but practical significance as well. Furthermore, the significant p-values of the "random slope" of phase at the study-level show that the coefficients associated with phase varied across studies. In other words, the amount of improvement a student exhibited changed based on the individual study in which they participated. These results provide further justification for a random effects model, as they indicate variation in true effects sizes may not occur completely at random, but may also be dependent on study features.

Of additional note are the performance feedback results. The presence of performance feedback was relatively evenly split between studies, with 10 studies including and 11 studies excluding feedback. The relationship of performance feedback with student behavior was a positive one and resulted in a parameter estimate of 5.45, suggesting student behavior improves

by a little over 5% when performance feedback is used. These results were also statistically significant (p = .009).

The significant results for the residual variance indicated there are variables related to consultation not included in this model, but which may account for a substantial amount of the remaining variance in behavioral outcomes. Of further interest are the autocorrelation results. It is not unexpected that single-case studies would have some amount of autocorrelation; however, the statistical significance of these results suggests the sample contained a notable amount of autocorrelation-related variance. Thus, the use of first-order autoregressive methods likely helped to account for autocorrelation issues which could have biased the results.

It should be noted that results were returned for the "AB-pair" variable, but this was a dummy-coded, control variable that did not provide logically interpretable results, thus these statistics were excluded from Table 4.

Bias and Homogeneity

A funnel plot was created to assess publication bias. As illustrated by Figure 5, most of the published literature reported effect sizes above the omnibus, while dissertations demonstrated smaller effects. If the unpublished literature was excluded, the left half of the funnel would be missing data, suggesting publication bias is an issue in this body of literature. The BC-Tau for unpublished and published literature was .39 (95% CI = .19 to .59) and .54 (95% CI = .42 to .66) respectively. In order to statistically assess the level of difference between these effects, a reliable difference formula was used (Bowman-Perrott, Burke, Zhang, & Zaini, 2014):

$$\text{Z-score} = (Tau_1 - Tau_2) / \sqrt{[(SE_{Tau1}^2) + (SE_{Tau2}^2)]}$$

In this case, Tau_1 was the BC-Tau for published studies and Tau_2 was the result for unpublished dissertations. The resulting z-score was 1.29 (p = .20). These results suggest that although there

was a marked difference between published and unpublished results, it was not a statistically significant one.

In order to assess for other forms of bias in the sample, an Egger test was conducted. This resulted in a non-significant *p*-value ($p = 0.157$), indicating bias is not a prevalent concern in this sample when both published and unpublished studies are included. In contrast, the *H* and I^2 analyses indicated the sample does have notable heterogeneity. The value for *H* was 3.3 (95% CI = 1.8 to 3.8), with values above 1.5 demonstrating significant heterogeneity. Furthermore, the I^2 result indicated 90.6% of the variance may be attributed to heterogeneity (95% CI = 87.5 to 93%). The presence of notable heterogeneity again emphasized the importance of implementing a random effects model in this meta-analysis, which allowed the researcher to account for this heterogeneity in the BC-Tau and MLM analyses.

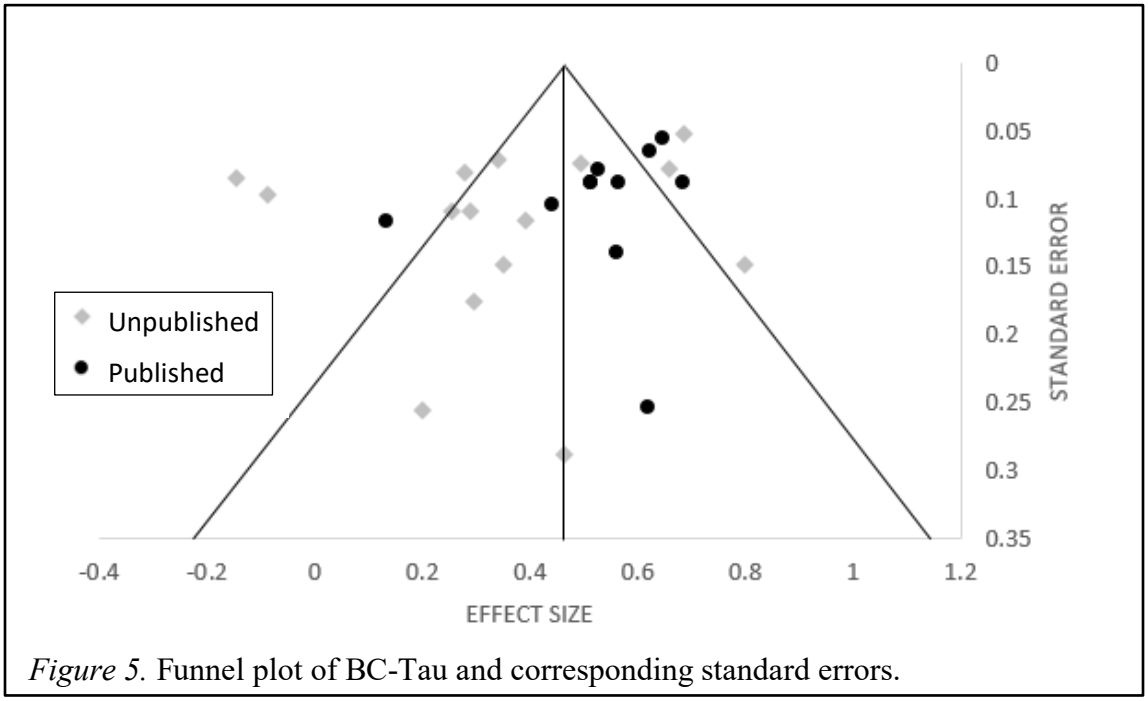

Figure 5. Funnel plot of BC-Tau and corresponding standard errors.

5. CONCLUSIONS

There were several findings from this meta-analysis. First, results indicated that the school-based, behavioral consultation process is generally an effective model when intervening with externalizing behaviors. A large BC-Tau overall effect size of .46 was found (Pearson's r = .66, Cohen's d = 1.76). When examining the BC-Tau effects for individual studies, there was some variability (with effects ranging from -.15 to .8), but most studies found that consultation improved externalizing behaviors to some degree. These results were further confirmed by the MLM analysis, which demonstrated that the intervention had a statistically significant impact on student behavior, as well as notable practical significance, with a parameter estimate indicating student behavior improved by about 17% after consultation. The results of the current paper are consistent with past meta-analyses whose authors have generally reported large effects when examining client behavior changes. Previous meta-analyses have found behavioral changes ranging from a Hedge's g of .50 to a Glass' Δ of 4.06 (Davis, 2012; Jackson, 1986). The results of the current analysis fall in the middle, between these two extremes.

Second, a study-level moderator was evaluated: performance feedback. This variable was included in an attempt to address treatment integrity concerns in the consultation literature base, as performance feedback provides both a means of measuring and addressing treatment fidelity concerns. In fact, when consultants utilized performance feedback, student behavior exhibited statistically significant changes. This suggests that treatment integrity differences may account for a notable amount of the variability between studies. However, the practical significance of such practices may be debatable, as this feedback accounted for only about a 5% increase in appropriate behavior. Performance feedback is a process which requires additional time and

effort from both the consultant and consultee as they ensure the treatment is being delivered with the utmost fidelity. Such a small practical change in behavior may bring this practice into question. Future studies should consider other study-level variables to determine if performance feedback is an area that would justify this extra level of attention, or if other changes to the consultation process might result in a greater amount of behavior change. Teachers and consultees have a finite amount of time and attention, thus components of the consultation process should be carefully considered to determine where these resources could best be invested.

Unpublished studies in the current meta-analysis tended to demonstrate lower effects, as compared to published ones; although the difference between these two groups was not statistically significant. This is consistent with findings from previous meta-analyses which incorporated unpublished literature into their analysis (Batts, 1987; Sibley, 1986). It is thus pivotal for future meta-analytic authors to consider the inclusion of unpublished works in their sample and how the exclusion of such studies might bias their results. Notably, the current sample did not demonstrate significant skew (i.e., bias) with the inclusion of unpublished works.

Autocorrelation is another limitation for researchers to contemplate, particularly those interested in single-case study designs. When using the same individual for both treatment and control conditions, variables not included in the model may impact the results (Baek & Ferron, 2013). For instance, if a student experiences a traumatic event, such as the loss of a pet, or is feeling ill over several observation periods, these variables can impact the trend in the data and result in autocorrelation issues. This meta-analysis utilized a nonparametric method robust to autocorrelation and accounted for these issues in the parametric analysis through the use of a first-order autoregressive matrix technique. When looking at the MLM results, autocorrelation

accounted for a significant amount of the variance in student behavioral outcomes, further highlighting the importance of incorporating autocorrelation control methods in meta-analytic studies featuring single-case designs.

Limitations and Future Research

In general, there was considerable between-study variance within this sample, which was demonstrated through high ICC and heterogeneity statistics. This issue was addressed by the use of a random effects model, which helped to account for differences between studies that appeared to be due to more than simple random error. As the literature base grows, future studies may be able to further correct this issue by creating more specific inclusion criteria to help create a more uniform sample and therefore less heterogeneity.

Another limitation of the current study is that the MLM method of moderator analysis restricts the number of variables which can be considered. When too many variables are included in a model, the model will fail to converge and results cannot be provided. The solution is to limit the number of variables, or reduce the severity of the inclusion criteria so more studies can be included in the analysis (i.e., with additional data points, more variables can often be incorporated in a model). The current analysis only found 21 studies which met inclusion criteria for the MLM analysis, thus the moderator variable analysis was limited to performance feedback; however, there are many additional variables that would be of interest.

For example, authors could investigate moderators identified in previous research. Many of the moderators in past meta-analyses have only been examined once, and replication studies could provide confirmation of these results. For instance, consultee job title, consultant experience level, and the duration of the consultation have each only been evaluated by a single author, but have resulted in interesting conclusions meriting additional investigation. Even when

multiple studies have evaluated the same moderator, there have often been divergent results, thus further research should be done to resolve these differences. In particular, findings on the optimal student-age has been widely disparate. It may prove particularly interesting to consider age in conjunction with other client variables, such as gender and disability status.

The results of this analysis suggest study-level moderators should receive special focus in future research, as most of the variation in student behavior was attributed to differences between studies. There are a plethora of study-level or process-related variables which have not yet been considered. These could include, conjoint consultation as compared to parent/teacher-only consultations, the number of meetings with the consultant, how collaborative the consultation approach was, and if teacher and parent satisfaction with the intervention was assessed, among many other variables.

Implications for Practice

Consultation appears to be a promising method of intervention delivery for use in the schools to address externalizing student behavioral issues. In the various studies examined, teachers and parents were able to effectively and independently improve client outcomes after meeting with a consultant. These results demonstrate a sizable amount of behavior change for a wide range of goals.

In comparison, the presence of performance feedback was related to statically significant changes, but only resulted in a small amount of behavioral improvement. Thus, consultants should assess whether the potential benefits of this practice outweigh the costs, which may include additional time and effort on both the teacher and consultant's part, finite resources that could conceivably be spent consulting on other students who also might need support. While performance feedback is considered a beneficial practice, its use should be done with careful

consideration. In general, behavioral consultations seem to be well suited for meeting the need of students who struggle to manage their behaviors in the school setting.

REFERENCES

Note. *References marked with an asterisk indicate studies included in the meta-analysis.

Abramson, J. H. (2011). WINPEPI updated: Computer programs for epidemiologists, and their teaching potential. *Epidemiologic Perspectives & Innovations, 8*(1). doi: 10.1186/1742-5573-8-1

*Angell, S. L. (2005). *Using Conjoint Behavioral Consultation with Children Who Have Autism: Effectiveness, Acceptability, and Generalization of Skills.* (Unpublished doctoral dissertation). Oklahoma State University, Stillwater, OK.

Baek, E. K., & Ferron, J. M. (2013). Multilevel models for multiple-baseline data: Modeling across-participant variation in autocorrelation and residual variance. *Behavior Research Methods*, *45*(1), 65-74.

Baek, E. K., Moeyaert, M., Petit-Bois, M., Beretvas, S. N., Van den Noortgate, W., & Ferron, J. M. (2014). The use of multilevel analysis for integrating single-case experimental design results within a study and across studies. *Neuropsychological Rehabilitation*, *24*(3-4), 590-606. doi: 10.1080/09602011.2013.835740

*Barnhouse, L. L. (1979). *The effects of collaborative consultation on the nonattending classroom behaviors of school children.* (Unpublished doctoral dissertation). University of Cincinnati, Cincinnati, OH.

Batts, J. W. (1987). *The effects of teacher consultation: A meta-analysis of controlled studies.* (Unpublished doctoral dissertation). University of Kentucky, Lexington, KY.

*Beckman, T. J. (2008). *Treatment integrity and child outcomes: Conjoint behavioral consultation in an urban setting with clients of ethnic minority status*. (Unpublished doctoral dissertation). Oklahoma State University, Stillwater, OK.

Begeny, J. C., & Martens, B. K. (2006). Assessing pre-service teachers' training in empirically-validated behavioral instruction practices. *School Psychology Quarterly, 21*(3), 262-285.

Bell, B. A., Ene, M., Smiley, W., & Schoeneberger, J. A. (2013). A multilevel model primer using SAS PROC MIXED. *SAS Global Forum*, *433*, 0-19.

*Bellinger, S. A., Lee, S. W., Jamison, T. R., & Reese, R. M. (2016). Conjoint Behavioral Consultation: Community–School Collaboration and Behavioral Outcomes Using Multiple Baseline. *Journal of Educational and Psychological Consultation*, *26*(2), 139-165. doi: 10.1080/10474412.2015.1089405

*Bice-Urbach, B. J., & Kratochwill, T. R. (2016). Teleconsultation: The use of technology to improve evidence-based practices in rural communities. *Journal of School Psychology*, *56*, 27-43.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* Hoboken, NJ: John Wiley & Sons, Ltd.

*Bornstein, P. H., Hamilton, S. B., & Quevillon, R. P. (1977). Behavior modification by long-distance: Demonstration of functional control over disruptive behavior in a rural classroom setting. *Behavior Modification*, *1*(3), 369-380.

Bowman-Perrott, L., Burke, M. D., Zhang, N., & Zaini, S. (2014). Direct and collateral effects of peer tutoring on social and behavioral outcomes: A meta-analysis of single-case research. *School Psychology Review*, *43*(3), 260-285.

Campbell, S. B., Shaw, D. S., & Gilliom, M. (2000). Early externalizing behavior problems: Toddlers and preschoolers at risk for later maladjustment. *Development and Psychopathology*, *12*(3), 467-488.

*Coffee, G. (2007). *Generalization programming: an examination of the role of prevention in school-based problem-solving consultation*. (Unpublished doctoral dissertation). University of Wisconsin—Madison, Madison, WI.

*Colton, D. L., & Sheridan, S. M. (1998). Conjoint behavioral consultation and social skills training: Enhancing the play behaviors of boys with attention deficit hyperactivity disorder. *Journal of Educational and Psychological Consultation*, *9*(1), 3-28. doi: 10.1207/s1532768xjepc0901_1

Cooper, H. (2017). *Research synthesis and meta-analysis: A step-by-step approach* (5th ed.). Los Angeles, CA: Sage Publications.

Dart, E. H., Cook, C. R., Collins, T. A., Gresham, F. M., & Chenier, J. S. (2012). Test driving interventions to increase treatment integrity and student outcomes. *School Psychology Review*, *41*(4), 467-481.

Davis, C. (2012). *A Meta-analysis of school-based problem-solving consultation outcomes: A review from 1986 to 2009.* (Unpublished doctoral dissertation). Texas A & M University, College Station, TX.

*Dee, C. C. (1998). *The effectiveness of behavioral consultation combined with competency-based training for the acquisition of teacher skills.* (Unpublished doctoral dissertation). The University of Utah, Salt Lake City, UT.

DiGennaro, F. D., Martens, B. K., & McIntyre, L. L. (2005). Increasing treatment integrity through negative reinforcement: Effects on teacher and student behavior. *School Psychology Review*, *34*(2), 220-231.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*, 629-634. doi: https://doi.org/10.1136/bmj.315.7109.629

*Emery, D. L. (1986). *The effect of behavioral consultation on teacher and mainstreamed student behaviors in a secondary classroom setting.* (Unpublished doctoral dissertation). Lehigh University, Bethlehem, PA.

*Engelhardt, L., Sulzer, B., & Altekruse, M. (1971). The counselor as a consultant in eliminating out-of-seat behavior. *Elementary School Guidance & Counseling*, *5*(3), 196-204.

Erchul, W. P., & Martens, B. K. (2010). *School consultation: Conceptual and empirical bases of practice*. New York, NY: Springer Science & Business Media.

Fedorov, S. (2013). *GetData Graph Digitizer*. Retrieved from http://getdata-graph-digitizer.com/

Freeman, J., Simonsen, B., Briere, D. E., & MacSuga-Gage, A. S. (2014). Pre-service teacher training in classroom management: A review of state accreditation policy and teacher preparation programs. *Teacher Education and Special Education: The Journal of the Teacher Education Division of the Council for Exceptional Children, 37*(2), 106-120. doi: 10.1177/0888406413507002

*Garbacz, S. A., & McIntyre, L. L. (2016). Conjoint Behavioral Consultation for children with Autism Spectrum Disorder. *School Psychology Quarterly*, *31*(4), 450-466.

*Geil, M. E. (1998). *Solution-focused consultation: An alternative consultation model to manage student behavior and improve classroom environment.* (Unpublished doctoral dissertation). University of Northern Colorado, Greeley, CO.

Greene, R. W., 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*(2), 79-89.

*Hazzard, J. C. (1977). *Effects of behavioral consultation on teacher application and transfer of behavior management principles.* (Unpublished doctoral dissertation). University of Arizona, Tucson, AZ.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*(11), 1539 –1558. doi: 10.1002/sim.1186

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, *71*(2), 165-179.

Horner, R. H., & Kratochwill, T. R. (2012). Synthesizing single-case research to identify evidence-based practices: Some brief reflections. *Journal of Behavioral Education, 21*(3), 266-272. doi: 10.1007/s10864-012-9152-2

Jackson, E. P. (1986). *A meta-analysis of school-based consultation outcome research.* (Unpublished doctoral dissertation). University of Pittsburgh, Pittsburgh, PA.

Kaufman, D., Codding, R. S., Markus, K. A., Tryon, G. S., & Kyse, E. N. (2013). Effects of verbal and written performance feedback on treatment adherence: Practical application of two delivery formats. *Journal of Educational and Psychological Consultation*, *23*(4), 264-299. doi: 10.1080/10474412.2013.845494

Kennedy, C. H. (2005). *Single-case designs for educational research.* Boston, MA: Allyn and Bacon.

Kratochwill, T. R., & Bergan, J. R. (1990). Behavioral consultation in applied settings: An individual guide. New York, NY: Springer Science & Business Media.

Kratochwill, T. R., Elliott, S. N., & Callan-Stoiber, K. (2002). Best Practices in School-Based Problem-Solving Consultation. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV*. Washington, DC, US: National Association of School Psychologists.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M. & Shadish, W. R. (2010). *Single-case designs technical documentation.* Retrieved from: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf.

Lehmann, E. L. (2011). *Fisher, Neyman, and the creation of classical statistics*. New York, NY: Springer Science & Business Media.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage Publications.

Maggin, D. M., Chafouleas, S. M., Goddard, K. M., & Johnson, A. H. (2011). A systematic evaluation of token economies as a classroom management tool for students with challenging behavior. *Journal of School Psychology*, *49*(5), 529-554. doi:10.1016/j.jsp.2011.05.001

Maggin, D. M., Swaminathan, H., Rogers, H. J., O'Keeffe, B. V., Sugai, G., & Horner, R. H. (2011). A generalized least squares regression approach for computing effect sizes in single-case research: Application examples. *Journal of School Psychology*, *49*(3), 301-321. doi: 10.1016/j.jsp.2011.03.004

McClelland, M. M., Cameron, C. E., Duncan, R., Bowles, R. P., Acock, A. C., Miao, A., & Pratt, M. E. (2014). Predictors of early growth in academic achievement: The head-toes-knees-shoulders task. *Frontiers in psychology*, *5*, 1-14.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, *22*(3), 276-282.

*McPhail, J. M. M. (2005). *The effectiveness of using didactic and direct training methods within school-based consultation.* (Unpublished doctoral dissertation). The Pennsylvania State University, State College, PA.

Medway, F. J., & Updyke, J. F. (1985). Meta-analysis of consultation outcome studies. *American Journal of Community Psychology*, *13*(5), 489-505. doi: 10.1007/BF00923263

*Mendicino, M. (2015). *The Effects of Partnership-based Collaborative Consultation on Treatment Integrity, Increasing Skill Acquisition, and Reducing Problematic Behaviors*. (Unpublished doctoral dissertation). West Virginia University, Morgantown, WV.

*Meyers, J. (1975). Consultee-centered consultation with a teacher as a technique in behavior management. *American Journal of Community Psychology*, *3*(2), 111-121.

*Munton, S. M. (2004). *Uncovering the most effective and cost beneficial methods of consultation in terms of disruptive behavior, treatment integrity, and social validity*. (Unpublished doctoral dissertation). University of California, Riverside, CA.

National Center for Education Statistics. (n.d.). *Fast facts: Students with Disabilities.* Retrieved from https://nces.ed.gov/fastfacts/display.asp?id=64

Nelson, J. R., Benner, G. J., Lane, K., & Smith, B. W. (2004). Academic achievement of K-12 students with emotional and behavioral disorders. *Exceptional Children*, *71*(1), 59-73.

Noell, G. H. (2008). Research examining the relationships among consultation process, treatment integrity, and outcomes. *Handbook of research in school consultation*, 323-341.

*Rahn, H. E. (2008). *The effects of direct and indirect school consultation methods on teacher treatment integrity and student behavior change in a general education setting.* (Unpublished doctoral dissertation). Capella University, Minneapolis, MN.

Reddy, L. A., Barboza-Whitehead, S., Files, T., & Rubel, E. (2000). Clinical focus of consultation outcome research with children and adolescents. *Special Services in the Schools*, *16*(1-2), 1-22. doi: 10.1300/J008v16n01_01

Reinke, W. M., Lewis-Palmer, T., & Martin, E. (2007). The effect of visual performance feedback on teacher use of behavior-specific praise. *Behavior Modification*, *31*(3), 247-263. doi: 10.1177/0145445506288967

*Robbins, J. R. (1990). *Consultee and client behavior change following consultation.* (Unpublished doctoral dissertation). University of Nebraska, Lincoln, NE.

Sanetti, L. M. H., & Kratochwill, T. R. (2007). Treatment integrity in behavioral consultation: Measurement, promotion, and outcomes. *International Journal of Behavioral Consultation and Therapy, 4*(1), 95-114. doi: 10.1037/h0100835

*Sanetti, L. M. H., Collier-Meek, M. A., Long, A. C., Byron, J., & Kratochwill, T. R. (2015). Increasing teacher treatment integrity of behavior support plans through consultation and implementation planning. *Journal of school psychology*, *53*(3), 209-229.

SAS Institute Inc. (2015). *SAS/IML® 14.1 User's Guide.* Cary, NC: SAS Institute Inc.

*Shipley, S. C. (2013). *The Effects of a Strength-Based Model of Behavioral Consultation on Student Behavior, Teachers' Use of Praise Statements, and Measures of Social Validity*. (Unpublished doctoral dissertation). Temple University, Philadelphia, PA.

Sibley, S. L. (1986). *A meta-analysis of school consultation research*. (Unpublished doctoral dissertation). Texas Woman's University, Denton, TX.

*Sterling-Turner, H. E., Watson, T. S., & Moore, J. W. (2002). The effects of direct training and treatment integrity on treatment outcomes in school consultation. *School Psychology Quarterly*, *17*(1), 47-77.

Sterne, J. A., Sutton, A. J., Ioannidis, J. P., Terrin, N., Jones, D. R., Lau, J., ... & Tetzlaff, J. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomized controlled trials. *BMJ*. doi: 10.1136/bmj.d4002.

Tarlow, K. R. (2016). *Baseline Corrected Tau Calculator*. Retrieved from http://www.ktarlow.com/stats/tau

Tarlow, K. R. (2017). An improved rank correlation effect size statistic for single-case designs: Baseline Corrected Tau. *Behavior Modification, 41*(4), 427-467.

*Tillman, T. C. (1999). *Generalization programming and school-based consultation: The effects of a generalization training program on teachers' consultation-related skills.* (Unpublished doctoral dissertation). Syracuse University, Syracuse, NY.

Walker, D. A. (2003). JMASM9: converting Kendall's tau for correlational or meta-analytic analyses. *Journal of Modern Applied Statistical Methods*, *2*(2), 525-530.

Walker, H. M., Nishioka, V. M., Zeller, R., Severson, H. H., & Feil, E. G. (2000). Causal factors and potential solutions for the persistent underidentification of students having emotional or behavioral disorders in the context of schooling. *Assessment for Effective Intervention*, *26*(1), 29-39.

*Whitley, A., & Sulzer, B. (1970). Reducing disruptive behavior through consultation. *Journal of Counseling & Development*, *48*(10), 836-841.

*Wilkinson, L. A. (1997). School-based behavioral consultation: Delivering treatment for children's externalizing behavior in the classroom. *Journal of Educational and Psychological Consultation*, *8*(3), 255-276.

Wilkinson, L. A. (2006). Monitoring treatment integrity: An alternative to the 'consult and hope' strategy in school-based behavioural consultation. *School Psychology International*, *27*(4), 426-438.

Zins, J. E., Kratochwill, T. R., & Elliott, S. N. (1993). *Handbook of consultation services for children: Applications in educational and clinical settings*. San Francisco, CA: Jossey-Bass Publishers.

APPENDIX A

EXAMPLE CODING FOR PARAMETRIC MLM ANALYSIS

Study	Feedback	Student	Phase	ABPair	Y
19	1	1	0	0	71.0065
19	1	1	0	0	63.3372
19	1	1	0	0	65.7308
19	1	1	1	0	5333794
19	1	1	1	0	71.1554
19	1	1	1	0	77.6889
19	1	1	1	0	62.9211
19	1	2	0	0	80.7819
19	1	2	0	0	73.8759
19	1	2	0	0	61.3292
19	1	2	1	0	75.1106
19	1	2	1	0	76.6881
19	1	2	1	0	79.8343
19	1	2	1	0	91.0254
19	1	3	0	1	65.1386
19	1	3	0	1	57.6047
19	1	3	0	1	77.3964
19	1	3	1	1	77.4225
19	1	3	1	1	57.6774
19	1	3	1	1	77.4672
19	1	3	1	1	79.8207
19	1	3	1	1	76.9622
19	1	3	0	2	98.7342
19	1	3	0	2	44.3038
19	1	3	0	2	72.1519
19	1	3	1	2	94.9367
19	1	3	1	2	83.5443
19	1	3	1	2	82.2785
19	1	3	1	2	75.9494
19	1	3	1	2	95.6842

APPENDIX B

MODEL SYNTAX FOR SAS MLM ANALYSIS

Model 1

```
proc mixed data=dissertation covtest;
class student study;
model y=/solution dfm=kenward;
random intercept/sub=study;
random intercept/sub=student (study);
repeated / subject=student type=AR(1);
run;
```

Model 2

```
proc mixed data=dissertation covtest;
class student study abpair;
model y=phase abpair /solution dfm=kenward;
random intercept /sub=study;
random intercept /sub=student(study);
repeated / subject=student type = AR(1);
run;
```

Model 3

```
proc mixed data=dissertation covtest;
class student study abpair;
model y=phase abpair /solution dfm=kenward;
random intercept phase abpair /sub=study;
random intercept phase abpair /sub=student(study);
repeated / subject=student type = AR(1);
run;
```

Model 4

```
proc mixed data=dissertation covtest;
class student study abpair feedback(ref='0');
model y=phase abpair pf /solution dfm=kenward;
random intercept phase abpair /sub=study;
random intercept phase abpair /sub=student(study);
repeated / subject=student type = AR(1);
run;
```