Treatment Integrity in Behavioral Consultation: Measurement, Promotion, and Outcomes

Lisa M. Hagermoser Sanetti & Thomas R. Kratochwill

Abstract

The purpose of this article is to provide an overview of the current state of research in the area of treatment integrity as it relates to behavioral consultation. To this end, four primary topics are discussed. First, a rationale for why treatment integrity is an important professional and methodological concern is presented. Second, conceptualizations of treatment integrity as a dimensional construct and methods for assessing treatment integrity are reviewed. Third, variables hypothesized to moderate treatment integrity and the empirical basis for several strategies to promote treatment integrity are summarized. Finally, the relationship between treatment integrity and outcomes is discussed. Research from related fields that may influence thinking and practice related to treatment integrity within behavioral consultation are discussed throughout.
Keywords: behavioral consultation, treatment integrity, treatment fidelity

The purpose of this article is to provide an overview of the current state of research in the area of treatment integrity as it relates to behavioral consultation. To this end, the following areas are discussed: (a) why treatment integrity is an important professional and methodological concern, (b) measurement of treatment integrity, (c) promotion of treatment integrity, and (d) the relationship between treatment integrity and treatment outcomes. Research from related fields that may influence thinking and practice related to treatment integrity within behavioral consultation are also discussed throughout.

Behavioral consultation (BC; Bergan & Kratochwill, 1990; Sheridan & Kratochwill, 2008) is an effective model of delivering intervention services for a wide range of academic and behavioral problems (Gutkin, 1996; Kratochwill & Stoiber, 2000) and is often cited by practitioners as a preferred role. BC involves providing indirect services wherein a consultant works with a consultee (e.g., teacher or parent) with the goal of producing change in a client (e.g., student). Typically, the BC model is described as a four-stage problem–solving process. In the first stage, *problem identification*, the consultant and consultee operationally define the target behavior(s), identify environmental conditions related to the target behavior(s), and develop a method of collecting baseline data. In the second stage, *problem analysis*, the consultant and consultee review the baseline data and develop an intervention plan to remediate the target behavior(s). In the third stage, *treatment implementation*, the consultee proceeds with implementation of the intervention and collection of outcome data with brief contacts from the consultant. In the final stage, *treatment evaluation*, the consultant and consultee meet to determine whether the intervention was effective in remediating the target behavior(s) (see Bergan & Kratochwill, 1990).

Treatment Integrity as a Professional and Methodological Concern

Over the past decade, our understanding regarding how to develop assessment-driven intervention plans and identify evidence-based interventions has greatly improved. However, identifying an appropriate intervention via the BC process is necessary but insufficient for producing behavior change (Wickstrom, Jones, LaFleur, & Witt, 1998). To produce behavior change, an intervention must actually be implemented as planned. Thus, among the many factors that have been identified as influencing the

efficacy of treatments, treatment integrity, the extent to which an intervention is implemented as planned, is the most fundamental (Gresham, 1989; Yeaton & Sechrest, 1981).

Without formative assessment of treatment integrity, researchers using experimental designs are not able to make valid conclusions about the functional relationship between the independent variable (i.e., intervention) and the dependent variable (e.g. student outcomes; Shadish, Cook, & Campbell, 2002). Likewise, if an intervention cannot be operationally defined, due to potential variations in implementation or poor understanding of the intervention, the external validity of the study is threatened (Moncher & Prinz, 1991; Shadish, Cook, & Campbell, 2002). For example, without an assessment of treatment integrity, a student that shows no decrease in problem behaviors at the end of an intervention may be recommended for a more intensive intervention, when in fact the less intensive intervention would have been effective but was poorly implemented. Furthermore, the consultant and consultee may deem an intervention, which would have been effective if properly implemented, to be ineffective reducing the likelihood of their using it with another student.

The BC model directly addresses the need to collect formative treatment outcome data throughout the treatment implementation phase. Yet, the BC model, as outlined by Bergan and Kratochwill (1990), does not directly address treatment integrity (Erchul & Schulte, 1996; Wilkinson, 2006; Witt, Gresham, & Noell, 1996), but rather implies that it should be assessed during brief consultation meetings conducted during treatment implementation. Several researchers have made recommendations regarding shortcomings of BC, such as this lack of attention to treatment integrity. For example, the use of social influence strategies to produce change in consultees' beliefs and attitudes regarding interventions (Erchul & Martens, 2002) and the provision of ongoing, systematic implementation support to consultees (Erchul & Martens, 2002; Witt, Gresham, & Noell, 1996) have been encouraged. Despite these recommendations, such differential attention to the specification and measurement of dependent (treatment outcomes) variables and disregard for the specification and measurement of independent variables is common in the BC literature, has been common in the treatment outcome literature across several disciplines (e.g., consultation, behavior analysis, child therapy), and was coined a "curious double standard" by Peterson, Homer, and Wonderlich (1982, p. 478). A review of treatment integrity within the treatment outcome literature illuminates the pervasive nature of this double standard.

Treatment Integrity within the Treatment Outcome Literature

Several reviews of treatment integrity in the treatment outcome literature (Gansle, 2005; Gresham, Gansle, & Noell, 1993; Gresham, Gansle, Noell, Cohen, & Rosenblum, 1993; Gresham, MacMillan, Beebe-Frankenberger, & Bocian, 2000; McIntyre, Gresham, DiGennaro, & Reed, 2007; Peterson et al., 1982; Snell, Chen, & Hoover, 2006; Wheeler, Baggett, Fox, & Blevins, 2006) have been conducted in the past two decades. Although each provides valuable information regarding the prevalence and quality of treatment integrity data in outcome literature, the results of these reviews are limited by their inability to be directly compared to one another due to differing methodologies and areas of focus; discussion of the reviews is organized by area of focus.

Journal of Applied Behavior Analysis. Peterson and colleagues (1982) conducted one of the earliest and most influential reviews of treatment integrity assessment in which they reviewed all of the experimental articles published in the *Journal of Applied Behavior Analysis (JABA)* from the journal's insemination in 1968 to 1980. This review has been updated twice (Gresham, Gansle, & Noell, 1993; McIntyre, Gresham, DiGennaro, & Reed, 2007) and together the three reviews provide a comprehensive review of the prevalence of treatment integrity data published in *JABA* from 1968 to 2005.

In their initial review, Peterson et al. (1982) categorized the precision of the intervention definition (reported, not reported but deemed unnecessary, not reported and deemed necessary) and the

type of treatment integrity assessment reported (assessment reported, no assessment reported but assessment deemed unnecessary, no assessment reported and deemed necessary) in each article. In both the Gresham and colleagues (1993) and the McIntyre and colleagues (2007) reviews, articles were considered to include an operational definition of the intervention if it could be replicated based on the description provided (references to more extensive sources were also accepted in the McIntyre et al. review). In the Gresham et al. and McIntyre et al. reviews, treatment integrity assessment was rated in terms of three categories: (a) yes (quantitative data provided); (b) monitored (mentioned assessment but no quantitative data provided), and (c) not mentioned. Figure 1 presents results from all three *JABA* reviews. Note that Peterson et al. did not address the number of studies in which treatment integrity was monitored or the mean level of treatment integrity for studies that did report quantitative treatment integrity data.

School-based studies. Using the same coding criteria, Gresham, Gansle, Noell and colleagues (1993) reviewed experimental, behaviorally-focused, school-based studies published in multiple journals between 1980 and 1990. Just under half (48.6%) of the school-based intervention studies reviewed utilized two or more interventions. Of the 181 studies reviewed, 35% (64 studies) provided an operational definition of the intervention(s). Only 14.9% (27 studies) of the studies were reported as having formally assessed and reported levels of treatment integrity for the independent variables, while 10% (18 studies) included treatment integrity monitoring but provided no quantitative data. The mean level of treatment integrity for the 27 studies that did include integrity assessment was 96.92% (range 75%-100%).

Learning disabilities. In an effort to further extend this line of research and evaluate the extent to which treatment integrity is assessed in the learning disabilities intervention research, Gresham and colleagues (2000) reviewed studies published in three learning disabilities journals between January 1995 and August 1999. Of the 65 intervention articles reviewed, 18.5% (12 studies) both assessed and reported treatment integrity data. Approximately half of the intervention studies reviewed alluded to or qualitatively described treatment integrity, but provided no numerical data regarding the number of treatment components implemented. Treatment integrity was not addressed in over 30% of the learning disabilities intervention articles reviewed.

Anger management. Gansle (2005) provided the first review of treatment integrity data in the anger management intervention literature. In a review of 20 outcome studies on school-based anger interventions and programs, published between 1984 and 2003, only 10% (2 studies) both assessed and reported treatment integrity data. Thirty percent (6 studies) mentioned treatment integrity as an important construct but did not measure it, and 60% (12 studies) did not mentioned treatment integrity at all. The average treatment integrity for the two studies that reported it was 92.5%.
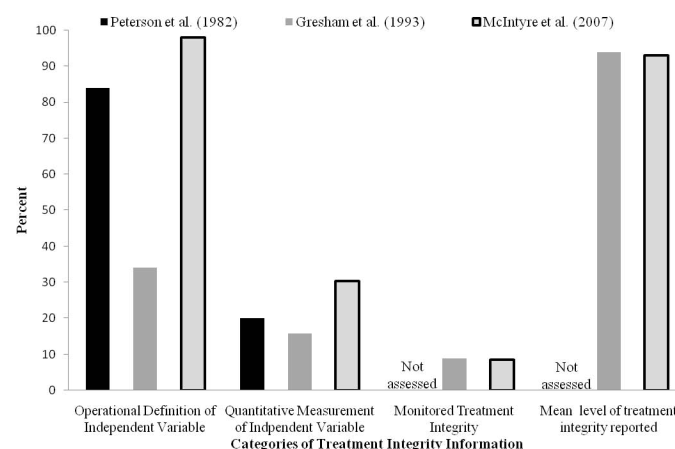

Figure 1. Categorized treatment integrity information from JABA reviews

Autism and alternative communication. Intervention research in the areas of autism and augmentative and alternative communication (AAC) have also been reviewed with respect to treatment integrity. Wheeler and colleagues (2006) evaluated the frequency of treatment integrity assessment in behavioral intervention studies conducted with children with autism. Of the 60 intervention studies reviewed (published between 1993-2003), results indicated that 92% (55 studies) reported operational definitions of the intervention, 18% (11 studies) operationally defined the intervention and assessed treatment integrity, 5% (3 studies) reported that they monitored treatment integrity but did not provide data, and 68% (41 studies) did not mention treatment integrity. In a review of 40 treatment outcome studies focused on AAC for individuals with severe disabilities, Snell and colleagues (2006) found that 30% (12 studies) both assessed and reported treatment integrity data. Of the 12 studies that reported treatment integrity data, 92% (11 studies) reported treatment integrity to be 80% or better.

Although the reviews of treatment integrity in the BC literature used different methods, the results suggest that it may be more common for researchers to operationally define independent variables than quantitatively measure them. The discrepancy between defining and measuring the independent variable may be influenced by the theoretical basis of the interventions studies in more recent reviews. Most of the studies in the Wheeler et al. (2006) and McIntyre et al. (2007) articles, for example, were behaviorally based. Operationally defining dependent variables has been a long-standing practice within the field of behavior analysis and BC and therefore, might be more readily applied to the independent variable. At best, only 39% of the articles in a review reported treatment integrity data. The dearth of treatment integrity assessment may be influenced by publication practices and the level of knowledge we have as a field regarding how to operationalize treatment integrity. With regard to publication practices, a majority of articles that did provide quantitative treatment integrity data reported high (80% or greater) levels of treatment integrity; peer-reviewed journals would likely be less inclined to publish a treatment outcome study with a low level of treatment integrity. Thus, it may be the case that researchers do not include quantitative treatment integrity data in articles if they illustrate a low level of implementation knowing that it will jeopardize their likelihood of publication, which is a requirement for most researchers. Alternatively, researchers may not be collecting treatment integrity data due to the fact that our understanding of treatment integrity within BC is in its infancy and it appears other fields are in a similar position (Power, Blom-Hoffman, Clarke, Riley-Tillman, Kellerher, & Manz, 2005). The measurement of treatment integrity is one of the greatest challenges to our understanding of treatment integrity.

Measurement of Treatment Integrity

Measurement of treatment integrity may be the single most significant barrier to our understanding treatment integrity (Noell, 2007). Our ability to measure treatment integrity is hindered by a lack of consensus regarding the dimensions of treatment integrity and limited data regarding the utility of behavioral assessment methods when applied to treatment integrity.

Dimensions of Treatment Integrity

Within BC, treatment integrity has often been conceptualized and assessed as a one-dimensional construct (Gresham, 1989). This focus is evident in the reviews of treatment integrity within the treatment outcome literature presented above. Results of these reviews illustrate that when BC researchers report treatment integrity data, they provide one metric of implementation: percentage of components implemented as planned. A brief review of conceptualizations of treatment integrity across other disciplines, such as prevention and psychotherapy, suggests that treatment integrity may be a multidimensional construct.

Prevention. In their review of primary and early secondary prevention programs, Dane and Schneider (1998) provided the first multidimensional conceptualization of treatment integrity. More specifically, they defined the following dimensions as being potentially important in the assessment of treatment integrity: (a) adherence, (b) exposure, (c) quality of delivery, (d) program differentiation, and (e) participant responsiveness. Adherence was defined as the extent to which specified program components were delivered as prescribed in treatment manuals. Exposure was an index that could include (a) the number of sessions implemented, (b) the length of sessions, or (c) the frequency with which the treatment techniques were implemented. Quality of delivery referred to the qualitative aspects of the intervention, such as interventionist effectiveness, enthusiasm, and preparedness. Participant responsiveness reflected the degree to which the intervention participants were engaged in the program. Program differentiation was defined as a manipulation check that is performed to safeguard against the diffusion of treatments. Research evaluating these different dimensions of treatment integrity is emerging in the prevention literature (Dusenbury, Brannigan, Falco, & Hansen, 2003); additional research is needed to clarify whether these are critical dimensions of treatment integrity and how each of these factors relates to treatment outcomes.

Power and colleagues (2005) combined the dimensions of treatment integrity proposed by Dane and Schneider (1998) with treatment integrity assessment guidelines proposed by Gresham (1989) into a framework for treatment integrity assessment. The framework suggests that two estimates of integrity and five dimensions of integrity should be considered in a thorough assessment of treatment integrity. The two estimates of integrity are component integrity (i.e., an average rating for a specific intervention component) and daily integrity (i.e., an average rating for all intervention components combined). Power et al. conceptualized each dimension as being either content or process related. According to this conceptualization, adherence, exposure, and program differentiation, address the content of the intervention and answer the question "How much of the intervention was implemented?" The remaining two dimensions of treatment integrity, quality of delivery and participant responsiveness, address the process of implementing the intervention and answer the question "How well was the intervention implemented?" Power and colleagues suggest that to execute a thorough assessment of treatment integrity that includes both estimates of and all of the dimensions of treatment integrity, it is essential to use a partnership model in which the responsibility for planning and executing treatment integrity assessment is shared by researchers and interventionists. This framework for assessing treatment integrity is the most comprehensive proposed to date. With regard to the estimates of integrity within the BC literature reviewed above, researchers who have provided quantitative data in their reports have focused solely on daily integrity. However, for many interventions, some intervention components are conceptualized as being essential for behavior change while implementation of other components is ideal but not essential (Waltz, Addis, Koerner, & Jacobson, 1993). Thus, assessing and reporting component integrity will allow data-based (as opposed to theory-driven) determinations as to which components of an intervention are truly essential to positive treatment outcomes. As mentioned above, research regarding the dimensions of treatment integrity is just emerging and additional studies are needed to confirm this conceptualization.

Psychotherapy. Waltz and colleagues (1993), in the context of psychotherapy trials, provided a conceptualization of treatment integrity that further expands the dimension of treatment differentiation. More specifically, they recommended that for a treatment integrity measure to address treatment differentiation in a useful manner, the measure should assess (a) treatment agent behaviors that are unique to the treatment and essential to it; (b) behaviors that are essential to the treatment but not unique to it; (c) behaviors that are compatible with the specified treatment, and therefore not prohibited, but neither necessary nor unique; and (d) behaviors that are proscribed. Data available from studies using this conceptualization in comparing two treatments suggest that it may be a useful framework for seeing the extent to which two treatments are different and therefore, how likely differential outcomes are. This level of measurement appears as though it could bolster the validity of conclusions drawn from the many studies that compare the comparative efficacy of treatments within a consultation framework.

Methods to Assess Treatment Integrity

There are a number of assessment methods consultants can use to assess treatment integrity; however, it is important to note that due to the fact that some treatments can occur at various times and locations during a day, all of the available methods for data collection will not be equally feasible for all interventions. Methods include: (a) permanent products; (b) direct observation; (c) self-monitoring, self-reporting, and behavioral interviews; and (d) manualized treatments and intervention scripts (Gresham, 1997; Lane, Bocian, MacMillan, & Gresham, 2004).

Permanent products. Permanent products, products generated as a result of the intervention, can be used by consultants as a method of assessing treatment integrity. The primary benefits of utilizing permanent products are that (a) they may not result in any additional work or responsibility for the consultee, (b) they can sample all occasions of the target behavior with minimal reactivity, and (c) they have been used successfully in multiple applied studies (Lane et al., 2004; Noell, 2007). The products can be collected by the consultant on a regular basis and evaluated to determine the extent to which components of the intervention were implemented. One drawback to permanent product review is that often all intervention components are not represented by the permanent products naturally created by the intervention. For example, consider an intervention in which a child self-monitors his behavior throughout the day by giving himself smiley faces for meeting behavioral expectations. If the child meets his goal for the day, his parents provide the appropriate reinforcement. The consultant could monitor whether the self-monitoring aspect of the intervention was implemented by collecting the smiley face sheets; however, there is no natural permanent product associated with the parents providing reinforcement as planned. Such situations may result in consultants needing to either create a permanent product for each intervention component that does not naturally have one or use multiple assessment methods (e.g., permanent products and consultee self-report) to ensure that treatment integrity data are collected for all intervention components.

Direct observation. Direct observation is the most commonly used direct method to assess treatment integrity (Lane et al., 2004). Directly observing treatment integrity is identical to systematically observing any behavior in applied settings and, therefore, factors such as the number of observation sessions and the reactivity of observations need to be taken into consideration (see Foster & Cone, 1986). Direct observation of treatment integrity requires (a) an operational definition of the intervention components, (b) a record of the occurrence or nonoccurrence of each component, and (c) a calculation of the percentage of treatment components (see Lane et al., 2004 or Hagermoser Sanetti & Kratochwill, 2005 for examples of direct observation forms).

Self-reporting, self-monitoring, and behavioral interviews. Consultee self-reporting requires the consultee to record his or her level of treatment integrity subsequent to intervention implementation. Self-report measures can be completed at the end of the session or day and are often used to assess treatment integrity. Such measures require (a) an operational definition of each intervention component, (b) a framework for rating each intervention component (e.g., likert scale, free response), and (c) the completion of the scale after implementation. Consultee self-report measures, however, may result in reactive effects resulting in either an over- or and under-estimation of the level of treatment integrity (see Lane et al., 2004 or Hagermoser Sanetti & Kratochwill, 2005 for examples of self-report forms).

Consultee self-monitoring (i.e., simultaneously implementing an intervention and assessing integrity) has not received a lot of attention from researchers as a viable direct measure of treatment integrity, as it may be incompatible with the implementation of an intervention.

The results of a study by Wickstrom and colleagues (1998) demonstrate the potential pros and cons of assessing treatment integrity via observation and self-report. This study examined the treatment

integrity of 27 teachers participating in BC. Treatment integrity was assessed with three measures of varying methodological rigor, (a) the Baseline and Intervention Record Form (BIRF), a teacher completed rating scale; (b) a stimulus product; and (c) direct observation. Mean integrity for the least rigorous method, BIRF, was 54%. The presence of stimulus materials increased the mean integrity to 62%. Observations of treatment use however, indicated that the mean percentage of target behaviors followed by the appropriate consequence was only 4%. The authors note that while these disparate results could be influenced by issues such as measurement error or missed observations, it is also important to consider that they could be dependent on the method of assessment utilized.

Other options include using behavioral rating scales or interviews. Wilkinson (2006) proposed adding a *treatment monitoring interview* (TMI) to the BC process to facilitate treatment integrity assessment. During the TMI, consultants would (a) review client outcomes, (b) identify barriers to intervention implementation, and (c) determine whether the intervention plan requires modification. The consultant would then provide performance feedback by (a) analyzing treatment integrity data, (b) providing praise for correctly implemented intervention components and corrective feedback for omitted or incorrectly implemented intervention components, (c) addressing consultee questions and concerns, (d) determining whether additional training or support is necessary, and (e) confirming the consultee's commitment to plan implementation. Expanding the BC process to directly address treatment integrity seems promising; however the labor-intense nature of such behavioral interviewing and performance feedback may make routine assessment of treatment integrity using this method difficult.

Manualized treatments and intervention scripts. Interventions that include explicit components such as treatment manuals and scripted lessons may facilitate increased levels of treatment integrity by providing operationally defined behavioral expectations for the consultee (Lane et al., 2004). Some manualized interventions include treatment integrity assessment measures (e.g., Webster-Stratton & Reid, 2003) to assess adherence to the manual. Even when the authors of a manualized treatment have not provided a treatment integrity assessment, it is relatively easy to develop a treatment integrity assessment if program procedures are delineated clearly (Power et al., 2005). Intervention scripts, collaboratively developed by a consultant and consultee were found to be an effective and acceptable method of measuring treatment integrity (Ehrhard, Barnett, Lentz, Stollar, & Reifin, 1996).

Promotion of Treatment Integrity

Across multiple school-based consultation studies, teachers often demonstrated perfect (100%) treatment implementation immediately following training and all of the teachers except for two displayed decreasing or low adherence levels (range: 0%-65%) within 0-10 days after training (Mortenson & Witt, 1998; Noell, Witt, Gilbertson, Ranier, & Freeland, 1997; Hagermoser Sanetti & Kratochwill , 2007; Witt, Noell, LaFleur, & Mortenson, 1997). These results illustrate the need to understand both variables that may moderate treatment integrity as well as strategies to increase treatment integrity in BC.

Variables that May Moderate Treatment Integrity

Much of the conceptual work related to treatment integrity in BC has focused on variables that may moderate treatment integrity (Yeaton & Sechrest, 1981; Gresham, 1989). Yeaton and Sechrest (1981) hypothesized that there is an inter-relationship between treatment strength, integrity, and effectiveness. More specifically, they hypothesized that choosing a theoretically strong intervention, if implemented with a high level of treatment integrity, would increase the probability of obtaining the desired outcomes. Conversely, Yeaton and Sechrest hypothesized that when treatment integrity decreased, the strength of the intervention was compromised.

Gresham (1989) hypothesized that consideration of the following factors would increase treatment integrity: (a) complexity of the treatment, (b) time required to implement the treatment, (c) access to materials or resources necessary to implement the treatment, (d) number of consultees needed, (e) motivation of the consultee(s), and (f) perceived and actual effectiveness of the treatment. More specifically, Gresham hypothesized that there is an indirect relationship between treatment integrity and the first four factors and a direct relationship between treatment integrity and the latter two factors. In addition to their individual influences on treatment integrity, Gresham, like Yeaton and Secrest (1981), recognized interactions among these factors. For example, an interaction between the complexity of an intervention and the time required for its implementation is likely—the more complex the intervention, the more time it will require to implement.

Treatment acceptability has also been hypothesized to be related to treatment integrity (Eckert & Hintze, 2000; Witt & Elliott, 1985; Reimers, Wacker, & Koeppl, 1987). Conceptual models of treatment acceptability suggest a reciprocal relationship between treatment acceptability, integrity, and effectiveness (Witt & Elliott, 1985; Reimers et al., 1987). The hypothesized relationship between acceptability and treatment integrity prompted a large amount of research. Resulting data do not support a direct relationship between treatment acceptability and integrity (Noell et al., 2005; Hagermoser Sanetti & Kratochwill, 2007), but rather support the view that human behavior is determined by multiple factors that may vary over time. For instance, evaluating an intervention as acceptable will not be sufficient to change a consultee's behavior if he or she does not have the resources or skills required for implementing the intervention. Conversely, a consultee may not find an intervention acceptable, but will implement the intervention because he or she is accountable to another stakeholder (e.g., principal, parent).

Consultant competence may also moderate treatment integrity (Schulte, 2007). It has been suggested that consultant competence within BC might include adapting procedures to the (a) consultee's knowledge or skill level, (b) consultee's level of distress, or (c) client's particular problem (Schulte, 2007). Several methods have been used to assess consultant competence in the small number of studies that have evaluated the variable (Bergan & Tombari, 1976; Hughes, Hasbrouck, Serdahl, Heidgerken, & McHaney, 2001; Kratochwill, Elliott, & Busse, 1995). The *Consultant Evaluation Rating Form* (Hughes & Hasbrouck, 1997) appears to be the most promising measure of consultant's competence in the following areas (a) task and interpersonal dimensions of consultation, (b) intervention plan development and effectiveness, (c) use of nonverbal behavior, and (d) sensitivity to consultee needs during consultation (Schulte, 2007).

Currently, with the exception of treatment acceptability, there are more hypotheses than data regarding the variables that may moderate treatment integrity within BC. Unfortunately, these variables are often written about as if there is strong empirical support for their role in moderating treatment integrity (Noell, 2007). While the hypothesized relationships discussed above are well reasoned and intuitively appealing, there is little programmatic research empirically supporting them. For example, therapist competence has been hypothesized to be an important moderating variable in the psychotherapy literature for over a decade (Waltz et al., 1993), but consultant competence as it relates to treatment integrity has just been proposed within the BC literature. Furthermore, there is no empirical support that these are the only or the most important moderating variables related to treatment integrity. For example, the level of home-school collaboration or degree of administrator support for intervention efforts may also be moderating variables. Until research on these variables and their hypothesized relationship with treatment integrity is conducted, the hypothesized relationships between these variables and treatment integrity must remain just that, hypotheses. Assuming such relationships exist will likely hinder, rather than help, the evolution of our understanding of treatment integrity within BC.

Strategies to Promote Treatment Integrity

Performance feedback. A programmatic line of research indicates that performance feedback can be utilized as an intervention to increase the level of treatment integrity of school-based interventions. When using performance feedback as a strategy, consultants analyze treatment integrity and client outcome assessments and provide the consultee with written and/or verbal constructive feedback. The feedback is intended to shape the consultee's implementation of the intervention and result in the consultee being able to implement the intervention with increased proficiency and effectiveness.

Witt, Noell, LaFleur, and Mortenson (1997) evaluated the relationship between performance feedback and treatment integrity. In a multiple baseline design study, four teachers were trained to implement an intervention for a student with an academic performance deficit. The intervention was designed so that each step of the intervention resulted in a permanent product, allowing the researchers to track the teacher's treatment integrity without the bias that may be present in self-report or direct observation methods. Teachers were provided with all of the necessary materials to carry out the intervention and were provided with both didactic and in vivo training on how to implement the intervention. The results of the study demonstrated that teachers' treatment integrity was perfect immediately following training but decreased rapidly during post-training baseline, with no teacher maintaining treatment integrity above 80% for greater than 2 days. Performance feedback included the consultant presenting the teacher with a graph of both the student's score on a daily worksheet and the teacher's treatment integrity as well as a brief discussion of missed intervention steps. When daily performance feedback was implemented, both treatment integrity and student outcomes improved. After performance feedback was removed, treatment integrity again declined, but did so more slowly than during the post-training baseline. All of the students' academic performance increased with the introduction of the intervention, and 3 of the 4 students' performance increased further with improved treatment integrity.

These results have been replicated several times using similar single-case research designs. Noell and colleagues (1997) found that teachers' treatment integrity was perfect for only a few days after training and then decreased. Furthermore, results indicated that treatment integrity increased immediately when performance feedback was introduced. Mortenson and Witt (1998) extended this line of research with a follow-up study in which the schedule of performance feedback was decreased from once daily to once weekly. Results indicated that all but one teachers' implementation dropped several days after training, and for those whose integrity declined, treatment integrity increased immediately when performance feedback was provided. Codding, Feinberg, Dunn, and Pace (2005) demonstrated that biweekly performance feedback was effective in increasing the level of treatment integrity with which teachers implemented behavior support plans in a private school for students with acquired brain injury. Interestingly, Noell and colleagues (2000) found that brief, daily meetings typically did not maintain high levels of treatment implementation similar to the more structured performance feedback meetings used in previous research. However, the brief meetings were moderately to highly effective for some teachers; therefore, the type and intensity of follow-up needed to increase and maintain treatment integrity may vary across intervention agents.

DiGennaro and colleagues (2005, 2007) built upon and extended the above lines of research by combining performance feedback with negative reinforcement. More specifically, DiGennaro, Martens, and McIntyre (2005) provided public school teachers with written and graphic performance feedback regarding their implementation of a behavior intervention plan. If the teacher demonstrated less than perfect implementation of the intervention, a meeting with the consultant to review and practice missed intervention steps was scheduled for the following day. However if the teacher demonstrated 100% treatment integrity, the teacher was not required to meet with the consultant. Treatment integrity increased across all participants. DiGennaro, Martens, and Kleinmann (2007) replicated these results in a private school for students with acquired brain injuries and demonstrated that performance feedback with a

negative reinforcement contingency was superior to goal setting in increasing teachers' level of treatment integrity.

Hagermoser Sanetti, Luiselli, and Handler (2007) contributed to this line of research by comparing the effects of two forms of performance feedback (verbal and graphic plus verbal) on the implementation of an individualized behavior support plan. Results indicated that treatment integrity declined within the first week of implementation and that weekly graphic plus verbal performance feedback resulted in higher levels of treatment integrity than weekly verbal feedback. These results suggest that graphic representation of treatment integrity data across time may be a critical element in delivering performance feedback.

Noell and colleagues (2005) conducted the first randomized field trial of three performance feedback procedures intended to increase the level of integrity with which teachers implemented treatment plans following consultation. The performance feedback procedures included brief weekly interviews, weekly interviews combined with an emphasis on the commitment to implement the treatment, and performance feedback. Brief weekly interviews were structured around questions about (a) the extent to which the intervention plan was implemented that week, (b) the extent to which the student was improving, and (c) the concerns of the teacher, and were conceptualized as corresponding to best practice in consulting in schools. The weekly interviews with commitment emphasis consisted of a consultant introducing a social influence procedure at the initial meeting with the teacher and reviewing specific points with the teacher that were designed to enhance correspondence between commitment to implement the intervention and actual implementation at subsequent weekly meetings. The performance feedback condition was similar to that used by Witt et al. (1997). The study was conducted in six urban elementary schools. Participants in the study were the first 48 teachers who made a referral to the building consultation team in their school. Students who completed the study included 32 boys and 13 girls, in kindergarten through fifth grade, referred for (a) academic skills problems, (b) challenging behaviors (c) academic work habits, and (d) challenging behaviors plus work habits. Participants were randomly assigned to one of the three follow-up conditions with two constraints. First, the number of participants was equal across conditions. Second, all consultants served in all conditions as equally as the first constraint and random assignment permitted. There were not statistically significant differences between treatment groups on student grade level or referral concern. Treatment plans were individualized for each student, but all incorporated self-monitoring. Results indicated that performance feedback was associated with superior treatment implementation and child outcomes when compared to the other conditions. Treatment implementation did not differ for the weekly meetings with and without commitment emphasis.

Hagermoser Sanetti and Kratochwill (2007) completed the first evaluation of the *Treatment Integrity Planning Protocol* (TIPP), a three-step process for planning and creating a treatment integrity assessment for any treatment. The process was derived from both guidelines for the assessment of treatment integrity (Gresham, 1989; Moncher & Prinz, 1991; Yeaton & Sechrest, 1981) as well as results of experimental research investigating factors that increase and maintain high levels of treatment integrity (Mortenson & Witt, 1998; Noell et al., 1997, 2000, 2005; Witt et al., 1997). The treatment integrity assessments that result from completion of the TIPP provide intervention agents with on-going process and outcome data. A multiple baseline across participants design was used to evaluate the effect of the TIPP on treatment integrity, student academic outcomes, and the accuracy of teachers' self-report of their intervention implementation.

Participants in the study were six elementary school teachers (five general education teachers, one reading specialist) and a student each teacher referred for consultation services due to the student's inconsistent performance in mathematics. Four teacher-student dyads were randomly assigned to the TIPP with assistance condition, one dyad was randomly assigned to the TIPP without assistance condition, and

one dyad was randomly assigned to the control condition. In the TIPP with assistance condition, the consultant assisted the teachers in completing the protocol, whereas in the TIPP without assistance condition the teacher was given the TIPP manual and protocol to complete independently. In the control condition, the teacher was not given the TIPP. Results indicated that although all teachers implemented the intervention with 100% integrity during an analogue training session, all but one of teachers in this study exhibited either low levels of or decreasing trends in their implementation of the intervention during the first week. When the TIPP was introduced and treatment integrity assessments were completed daily by teachers, both visual analyses of teachers' levels of treatment integrity and effect size statistics indicated that the TIPP, with and without assistance, was an effective intervention to increase teachers' levels of intervention implementation. The data also suggest that introduction of the TIPP decreased the variability of teachers' intervention implementation. Finally, the results indicated that teachers provided accurate feedback about their implementation on the treatment integrity assessments.

Consultee training. Although hypothesized to be a factor related to treatment integrity (Gresham, 1989) and considered in many of the studies described above, there is relatively little research about how consultees should be trained to maximize maintenance and generalization of learned intervention skills (Kratochwill, Sheridan, Carrrington Rotto, & Salmon, 1992). Consultee training procedures can be separated into direct methods and indirect methods. Didactic instruction and written descriptions of an intervention are examples of an indirect approach while modeling, role-playing, rehearsal and feedback are considered direct training procedures.

Sterling-Turner, Watson, Wildmon, Watkins, and Little (2001) investigated the relationship between training method and treatment integrity. Consultants (undergraduate students) were trained to implement a treatment protocol for a confederate seeking treatment at a university clinic. The consultants received 5 minutes of either didactic training, modeling training, or rehearsal feedback training and then conducted a therapy session with the confederate. All sessions were videotaped and treatment integrity was evaluated. Results indicated that modeling and rehearsal feedback, the direct training procedures, were associated with higher levels of treatment integrity.

Extending this line of research, Sterling-Turner, Watson, and Moore (2002) investigated the influence of direct and indirect training procedures on treatment integrity and treatment outcomes in an applied setting. In a multiple baseline across consultees design, three teachers implemented an intervention for a student with problem behaviors. During the first phase, (a) consultation goals were established, (b) target behaviors were defined, (c) baseline data were collected, (d) specific antecedents and consequences of behavior were collected, and (e) treatment plans were developed. Each treatment plan had multiple components and required the teacher to perform specific behaviors to implement the treatment. During the second phase, teachers received didactic instruction on how to implement their treatment plan and then implemented the plan. Direct observation data were collected using the same procedures used for the baseline data. Consultees were given no information regarding their implementation. During the third phase, rehearsal and feedback training, the teacher was first presented with baseline and phase two data after which consultant pointed out areas of strength and areas in need of improvement in the teacher's implementation. The didactic training from phase two was repeated and followed by role-play in which the consultant modeled each step of the treatment while the consultee role-played the part of the child. In addition, there was an opportunity for each consultee to rehearse and receive feedback on his or her implementation of the treatment plan with the client. After training was complete, direct observation data were again collected. Results indicated that treatment integrity increased slightly after didactic training and sharply after direct training in all cases. Three of the four cases implied a direct relationship between high treatment integrity and successful outcomes. In addition, the results of one case indicate that direct instruction was insufficient to either change teacher behavior or increase student outcomes.

Overall, results from experimental research indicate that (a) performance feedback is an empirically-supported method for increasing consultees' level of treatment integrity, (b) training might improve treatment integrity levels, however, as there has only been one study completed in a naturalistic setting (Sterling-Turner et al., 2002), more research is needed; and (c) the TIPP might be a useful tool for creating a treatment integrity assessment and providing performance feedback to a consultee.

<center>Treatment Integrity and Treatment Outcomes</center>

Intuitively, it is appealing to believe that increased treatment integrity results in increased treatment outcomes, with perfect implementation resulting in optimal client outcomes. Many of the studies described above have evaluated this relationship secondary to the effectiveness of strategies to promote treatment integrity. Results from these secondary evaluations are variable with some indicating a direct correlation between treatment integrity and outcomes (e.g., DiGennaro et al., 2005, 2007; Noell et al., 2000; Witt et al., 1997) and others finding an inconsistent relationship (e.g., Hagermoser Sanetti & Kratochwill, 2007; Noell et al., 1997).

Several reserachers have empirically examined the relationship between treatment integrity and treatment outcomes as a primary research question. McEvoy, Shores, Wehby, Johnson, & Fox (1990) investigated 17 teachers' levels of implementation of a social skills intervention for 12 students with moderate to severe disabilities. After the intervention, the authors divided the teachers into thirds based on their levels of treatment integrity. Then, the third of teachers who implemented the social skills treatment with the highest level of integrity was compared to the third who implemented the treatment with the lowest level of integrity. Across the outcome measures, students of the teachers with the highest levels of treatment integrity had better outcomes during the direct instruction phase than did student of teachers with the lowest levels of integrity.

Greenwood, Terry, Arreaga-Mayer, and Finney (1992) assessed the implementation of classwide peer tutoring (CWPT) to increase spelling achievement in five elementary school classrooms. Results indicated that variability in teacher implementation was associated with variability in student improvements. In addition, results suggested that low treatment integrity was not predictive of treatment failure in all cases. Rather, low levels of treatment integrity seemed to decrease the probability of treatment success for a student.

Holcombe, Wolery, and Snyder (1994) examined the effects of two levels of treatment integrity for constant time delay instruction. Preschoolers with developmental delays were taught to identify or select pictures with either continuous delivery of the delayed prompt (high integrity) or intermittent delivery of the delayed prompt (low integrity). One student learned equally well under both integrity conditions. Three students achieved mastery in both conditions but learned less efficiently in the low integrity condition. A fifth student achieved mastery under the high integrity condition, but not the low integrity condition. When this student was reintroduced to the high integrity condition it was ineffective unless combined with reinforcement. The authors hypothesized that previous exposure to low integrity instruction decreased the effectiveness of subsequent high integrity instruction.

Gansle and McMahon (1997) expanded this line of research by implementing three levels of treatment integrity of a self-monitoring intervention for elementary school children. The three levels of implementation were (a) self-monitoring, reward, and graphing of behavior (100% integrity), (b) self-monitoring and reward (83.3% integrity), and (c) self-monitoring with recording only (66.7% integrity). According to teacher reports, all three groups displayed similar levels of improvement after four weeks of plan implementation. Students using the more inclusive levels of implementation recorded more positive

behaviors when self-monitoring. These results indicate that the implementation of all treatment components may not be equally important.

Vollmer, Roane, Ringdahl, and Marcus (1999) found similar results when they evaluated the effectiveness of differential reinforcement of alternative behavior (DRA) at varying levels of treatment integrity with three students with significant disabilities who were referred for severe behavior problems. Results indicated that when exposed to the DRA at 100% implementation (all appropriate behavior was reinforced and no aberrant behavior was reinforced) inappropriate behavior was almost completely replaced by appropriate behavior. Subsequent exposure to the DRA in conditions that mimicked decreased treatment integrity levels (25%, 50%, or 75% of appropriate and/or inappropriate behaviors reinforced) resulted in decreased treatment effectiveness. These results provide further evidence that higher levels of treatment integrity are associated with better client outcomes and suggest that subsequent exposure to lower levels of treatment integrity can compromise intervention effects.

Noell, Gresham, and Gansle (2002) investigated three levels of computer delivered math instruction. The computer delivered a prompt to use a specific strategy to solve the math problems for all, two-thirds, or one-third of the instructional trials. Students exhibited a 45% mean percentage improvement during the lowest level of treatment integrity and a 91% mean percentage improvement during perfect implementation. Results during the middle integrity condition were variable.

Wilder and colleagues (2006) extended this line of research by evaluating the effects of three levels of treatment integrity (100%, 50%, and 0%) on two students' compliance when implementing a three-step prompting procedure. Each of the three directives with which the two students often did not comply was randomly assigned to one of the three levels of treatment integrity for each participant. Results indicated a direct relationship between treatment integrity and outcomes; directives assigned to the 100%, 50%, and 0% integrity levels were associated with substantial, moderate, and no improvement respectively.

van Otterloo and colleagues (2006) evaluated the impact of treatment integrity on the outcomes of home-based phonological awareness intervention in The Netherlands. The intervention focused on letter-sound correspondence and was designed to take approximately 10 minutes per day, 5 days a week for 14 weeks. Results indicated that approximately two-thirds of the parents completed the intervention. Of those that completed the intervention, mean treatment integrity was 66% and accounted for 36% of the variance in the dependent variable (pre-reading skills).

As each of the studies measured treatment integrity differently, examined different treatments, and varied in study population, the conclusions that can be drawn from their results are challenging. Considered together, the results suggest a probabilistic relationship between treatment integrity and treatment outcomes (Noell, 2007). Considered as individual studies, four conclusions can be drawn. First, assuming an effective treatment, high levels of treatment integrity appear to result in better client outcomes (McEvoy et al., 1990; Noell et al., 2002; Vollmer et al., 1999; Wilder et al., 2006). Second, lower levels of treatment integrity may make the treatment less effective (Wilder et al., 2006), less efficient (Holcombe et al., 1994), or produce less predictable responses (Greenwood et al., 1992; Noell et al., 2002). Third, positive treatment outcomes resulting from initially high levels of treatment integrity may be compromised as treatment integrity decreases over time (Vollmer et al., 1999). Fourth, it may not be necessary to be equally concerned with the implementation of all components of an intervention (Gansle & McMahon, 1997).

Behavioral Consultation and Treatment Integrity: Where are we?

It is commonly accepted that treatment integrity is an important methodological concern for any researcher or practitioner interested in making valid conclusions about the effectiveness of an intervention. Unfortunately, BC researchers have not embraced treatment integrity as a vital variable on which to collect and report data. This situation may be due to the relative lack of understanding of this variable, its dimensions, and how to assess it. Nevertheless, methods for assessing treatment integrity in BC have been identified and some, such as permanent products and the TIPP, are being empirically evaluated in applied settings. Until we have a better understanding of treatment integrity, best practices in assessment should be followed. Using multiple methods of assessment and examining assessment results for convergent findings will provide the richest information about how completely and consistently an intervention is being implemented.

To date, several variables that may moderate treatment integrity have been presented. The factors intuitively make sense, but as of yet there is little empirical support for their relationship to treatment integrity. By far, the greatest amount of systematic research has been related to how to increase treatment integrity in BC. Results of this research suggest that direct training in the problem analysis stage followed by weekly graphic and verbal performance feedback may be the most efficient and effective method for increasing treatment integrity levels in BC. Higher levels of treatment integrity have consistently resulted in better client outcomes. In general, lower levels of treatment integrity have resulted in poorer client outcomes; however, this relationship seems to be variable. The impact of low levels of implementation may vary due to a variety of factors (e.g., intervention, components implemented within an intervention, client, consultee, setting). Moreover, implementation of some treatment components may be more important to client outcomes than others. In these cases, low integrity may still result in better client outcomes when the essential treatment components are implemented.

As stated earlier, we are in our infancy in understanding treatment integrity. However, we aren't alone, researchers and practitioners from multiple disciplines are also working to develop an understanding of treatment integrity as a construct, how it can be assessed and promoted, as well as how it is related to treatment outcomes. It will be essential that BC researchers and practitioners collaborate with researchers and practitioners in other fields to identify (a) important dimensions of treatment integrity, (b) factors that moderate treatment integrity, (c) assessment methods that are psychometrically sound yet feasible, and (d) methods to promote treatment integrity.

Behavioral Consultation and Treatment Integrity: What can we do?

Despite the limited knowledge we have with relation to treatment integrity, it is still essential that all consultants are knowledgeable of, plan for, assess, and evaluate the treatment integrity of interventions developed and implemented via BC to the best of their ability. Consultants can plan for higher levels of treatment integrity by proactively addressing the many factors that are hypothesized to influence treatment integrity. A consultant can consider the complexity of and time required to implement the proposed intervention in light of the demands of the consultee's current context prior to implementation. For example, a consultee may be able to identify particular locations or times when implementation will be more challenging. Such information would enable a consultant and consultee to problem-solve what additional supports may be needed in these places or during these times to bolster treatment integrity. Furthermore, consultants can work to provide intervention options that are evidence-based, acceptable to the consultee, and believed by the consultee to be effective. When an intervention that meets all of these criteria is not available, the consultant can introduce an appropriate evidence-based intervention and employ social influence strategies (see Erchul and Martens, 2002) to improve the consultees' acceptance of the intervention, provide structured feedback about student outcomes, or both in an effort to change the

consultees' beliefs and attitudes. Regardless of the intervention, research results suggest consultants should provide direct training to consultees. Ensuring that a consultee is competent to implement the intervention in context at the onset of the intervention implementation may serve to increase the likelihood of intervention implementation (and hopefully positive client outcomes) and decrease consultee frustration.

With regard to assessing treatment integrity, a multi-method, multi-source approach that includes direct observation of intervention implementation appears to be the most sound at this time. As there is little to no research supporting several of the assessment methods outlined above, consultants should make time to conduct direct observations and utilize permanent product review whenever possible. Furthermore, use of tools such as the TIPP can facilitate developing treatment integrity assessments that incorporate the multiple conceptualizations of treatment integrity.

When a consultant formatively assesses client outcomes and treatment integrity, a comprehensive evaluation of the intervention can be conducted and data-based decisions made regarding whether the intervention should be continued, strategies to promote treatment integrity are needed, or another intervention should be considered. When treatment integrity data suggest a high level of implementation and student outcome data suggest positive outcomes, no changes are warranted with regard to the intervention. To maintain this level of implementation, it is advisable for the consultant to provide positive verbal feedback to the consultee and to positively reinforce this behavior in other ways appropriate to the setting (e.g., in a school setting, if positive attention from the principal or colleagues was desired by the teacher, the consultant could announce the consultee's high level of treatment integrity at a staff meeting). When treatment integrity data suggest a low level of implementation and student outcome data suggest neutral or negative outcomes, a consultant should implement strategies to promote treatment integrity. Research results suggest weekly graphic and verbal performance feedback effectively increases teachers' level of treatment integrity (Codding et al., 2005; Hagermoser Sanetti et al., 2007; Mortenson & Witt, 1997; Noell et al., 1997, 2000, 2005; Witt et al., 1997) . A consultant may also require consultation meetings to practice missing intervention steps when a consultee demonstrates a low level of treatment integrity and remove this requirement when a consultee demonstrates a high level of treatment integrity. When treatment integrity data suggest a high level of implementation and student outcome data suggest neutral or negative outcomes, a consultant should return to problem identification (DiGennaro et al., 2005, 2007). If the problem was correctly identified, another evidence-based intervention should be implemented and comprehensively evaluated. Despite the dearth of empirical knowledge regarding treatment integrity, consultants knowledgeable about why treatment integrity is a methodological concern, how to measure and promote treatment integrity, and how treatment integrity and treatment outcomes are related will be able to provide more effective services to their consultees and clients.

References

Bergan, J. R., & Kratochwill, T. R. (1990). *Behavioral consultation in applied settings.* New York: Plenum.

Bergan, J. R., & Tombari, M. L. (1976). Consultant skill and efficiency and the outcomes of consultation. *Journal of School Psychology, 14*, 3-14.

Codding, R. S., Feinberg, A. B., Dunn, E. K., & Pace, G. M. (2005). Effects of immediate performance feedback on implementation of behavior support plans. *Journal of Applied Behavior Analysis, 38*, 205-219.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18*, 23-45.

DiGennaro, F. D., Martens, B. K., & Kleinmann, A. E. (2007). A comparison of performance feedback procedures on teachers' treatment implementation integrity and students' inappropriate behavior in special education classrooms. *Journal of Applied Behavior Analysis*, 40, 447-461.

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, 220-231.

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health Education Research, 18*, 237-256.

Eckert, T.L., & Hintze, J.M. (2000). Behavioral conceptions and applications of acceptability: Issues related to service delivery and research methodology. *School Psychology Quarterly, 15,* 123-148.

Ehrhard, K. E., Barnett, D. W., Lentz, F. E. J., Stollar, S. A., & Reifin, L. H. (1996). Innovative methodology in ecological consultation: Use of scripts to promote treatment acceptability and integrity. *School Psychology Quarterly, 11*, 149-168.

Erchul, W. P., & Martens, B. K. (2002). *School consultation: Conceptual and empirical bases of practice* (2nd ed.). New York: Plenum Publishers.

Erchul, W. P., & Schulte, A. C. (1996). 'Behavioural consultation as a work in progress: A reply to Witt, Gresham, and Noell.' *Journal of Educational and Psychological Consultation, 7*, 345-354.

Foster, S. L., & Cone, J. D. (1986). Design and use of direct observation. In A. R. Ciminero, K. S. Calhoun, & H. E. Adams (Eds.), *Handbook of behavioral assessment* (pp. 253-324). New York: Wiley.

Gansle, K. A. (2005). The effectiveness of school-based anger interventions and programs: A meta-analysis. *Journal of School Psychology, 43*, 321-341.

Gansle, K. A., & McMahon, C. M. (1997). Component integrity of teacher intervention management behavior using a student self-monitoring treatment: An experimental analysis. *Journal of Behavioral Education, 7*, 405-419.

Greenwood, C. R., Terry, B., Arreaga-Mayer, C., & Finney, R. (1992). The classwide peer tutoring program: Implementation factors moderating students' achievement. *Journal of Applied Behavior Analysis, 25*, 101-116.

Gresham, F. M. (1989). Assessment of treatment integrity in school consultation and prereferral intervention. *School Psychology Review, 18*, 37-50.

Gresham, F. M. (1997). Treatment integrity in single-subject research. In R. D. Franklin, D. B. Allison, & B. S. Gorman (Eds.), *Design and Analysis of Single-Case Research*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Gresham, F. M., Gansle, K. A., & Noell, G. H. (1993). Treatment integrity in applied behavior analysis with children. *Journal of Applied Behavior Analysis, 26*, 257-263.

Gresham, F. M., Gansle, K. A., Noell, G. H., Cohen, S., & Rosenblum, S. (1993). Treatment integrity of school-based behavioral intervention studies: 1980-1990. *School Psychology Review, 22*, 254-272.

Gresham, F. M., MacMillan, D. L., Beebe Frankenberger, M. E., & Bocian, K. M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented? *Learning Disabilities Research and Practice, 15*, 198-205.

Gutkin, T. B. (1996). Patterns of consultant and consultee verbalizations: Examining communication leadership during initial consultation interviews. *Journal of School Psychology, 34*, 199-219.

Hagermoser Sanetti, L., & Kratochwill, T. R. (2005). Treatment integrity assessment within a problem-solving model. In R. Brown-Chidsey (Ed.), *Assessment for intervention: A problem-solving approach* (pp.304-325). New York: Guilford Press.

Hagermoser Sanetti, L. M., & Kratochwill, T. R. (2007). *Treatment integrity assessment in the schools: An evaluation of the Treatment Integrity Planning Protocol (TIPP)*. Manuscript submitted for publication.

Hagermoser Sanetti, L., Luiselli, J. K., & Handler, M. W. (2007). Effects of verbal and graphic performance feedback on behavior support plan implementation in a public elementary school. *Behavior Modification, 31,* 454-465.

Holcombe, A., Wolery, M., & Snyder, E. (1994). Effects of two levels of procedural fidelity with constant time delay on children's learning. *Journal of Behavioral Education, 4,* 49-73.

Hughes, J. N. & Hasbrouck, J. (1997). *Consultation Evaluation Rating Form (CERF) scoring manual.* Unpublished manuscript. College Station, TX: Texas A& M University.

Hughes, J. N., Hasbrouck, J. E., Serdahl, E., Heidgerken, A., & McHaney, L. (2001). Responsive systems consultation: A preliminary evaluation of implementation and outcomes. *Journal of Educational and Psychological Consultation, 12*, 179-201.

Kratochwill, T. R., Elliott, S. N., & Busse, R. T. (1995). Behavior consultation: A 5-year evaluation of consultant and client outcomes. *School Psychology Quarterly, 10*, 87-117.

Kratochwill, T. R., Sheridan, S. M., Carrrington Rotto, P., & Salmon, D. (1992). Preparation of school psychologists in behavioral consultation service delivery. In T. Kratochwill, S. N. Elliott, & M. Gettinger (Eds.), *Advances in School Psychology* (Vol. VIII, pp. 115-152). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kratochwill, T. R., & Stoiber, K. C. (2000). Empirically supported interventions and school psychology: Rationale and methodological issues—Part I. *School Psychology Quarterly, 15,* 75-105.

Lane, K. L., Bocian, K. M., MacMillan, D. L., & Gresham, F. M. (2004). Treatment integrity: An essential-but often forgotten-component of school-based interventions. *Preventing School Failure, 48*, 36-43.

McEvoy, M. A., Shores, R. E., Wehby, J. H., Johnson, S. M., & Fox, J. J. (1990). Special education teachers' implementation of procedures to promote social interaction among children in integrated settings. *Education and Training in Mental Retardation, 25*, 267-276.

McIntyre, L. L., Gresham, F. M., DiGennaro, F. D., & Reed, D. D. (2007). Treatment integrity of school-based interventions with children in Journal of Applied Behavior Analysis studies from 1991 to 2005. *Journal of Applied Behavior Analysis, 40*, 659-672..

Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review, 11*, 247-266.

Mortenson, B. P., & Witt, J. C. (1998). The use of weekly performance feedback to increase teacher implementation of a prereferral academic intervention. *School Psychology Review, 27*, 613-627.

Noell, G. H. (2007). Research examining the relationships among consultation process, treatment integrity, and outcomes. In W. P. Erchul & S. M. Sheridan (Eds.), *Handbook of research in school consultation: Empirical foundations for the field* (pp.315-334)*.* Mahwah, NJ: Erlbaum.

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

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

Noell, G. H., Witt, J. C., Slider, N. J., Connell, J. E., Gatti, S. L., Williams, K. L., Koenig, J. L., & Resetar, J. L. (2005). Treatment implementation following behavioral consultation in schools: A comparison of three follow-up strategies. *School Psychology Review, 34*, 87-106.

Noell, G. H., Witt, J. C., LaFleur, L. H., Mortenson, B. P., Ranier, D. D., & LeVelle, J. (2000). Increasing intervention implementation in general education following consultation: A comparison of two follow-up strategies. *Journal of Applied Behavior Analysis, 33*, 271-284.

Peterson, L., Homer, A., & Wonderlich, S. (1982). The integrity of independent variables in behavior analysis. *Journal of Applied Behavior Analysis, 15*, 477-492.

Power, T. J., Blom-Hoffman, J., Clarke, A. T., Riley-Tillman, T. C., Kellerher, C., & Manz, P. (2005). Reconceptualizing intervention integrity: A partnership-based framework for linking research with practice. *Psychology in the Schools, 42*, 495-507.

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

Schulte, A. C. (2007). Measurement in school consultation research. In W. P. Erchul & S. M. Sheridan (Eds.), *Handbook of research in school consultation: Empirical foundations for the field* (pp.29-57). Mahwah, NJ: Erlbaum.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton Mifflin.

Sheridan, S. M., & Kratochwill, T. R. (2008). *Conjoint behavioral consultation: Promoting family-school connections and interventions (2nd ed.).* New York: Springer.

Snell, M. E., Chen, L. Y., & Hoover, K. (2006). Teaching augmentative and alternative communication to students with severe disabilities: A review of intervention research 1997-2003. *Research and Practice for Persons with Severe Disabilities, 31*, 203-214.

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*, 47-77.

Sterling-Turner, H. E., Watson, T. S., Wildmon, M., Watkins, C., & Little, E. (2001). Investigating the relationship between training type and treatment integrity. *School Psychology Quarterly, 16*, 56-67.

van Otterloo, S. G., van der Leij, A., & Veldkamp, E. (2006). Treatment integrity in a home-based pre-reading intervention programme. *Dyslexia, 12*, 155-176.

Vollmer, T. R., Roane, H. S., Ringdahl, J. E., Marcus, B. A. (1999). Evaluating treatment challenges with differential reinforcement of alternative behavior. *Journal of Applied Behavior Analysis, 32*, 9-23.

Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. S. (1993). Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology, 61*, 620-630.

Webster-Stratton, C., & Reid, J. M. (2003). Treating conduct problems and strengthening social and emotional competence in young children: The dina dinosaur program. *Journal of Emotional and Behavioral Disorders, 11*, 130-143.

Wheeler, J. L., Baggett, B. A., Fox, J., & Blevins, L. (2006). Treatment integrity: A review of intervention studies conducted with children with autism. *Focus on Autism and Other Developmental Disabilities, 21*, 45-54.

Wickstrom, K. F., Jones, K. M., LaFleur, L. H., & Witt, J. C. (1998). An analysis of treatment integrity in school-based behavioral consultation. *School Psychology Quarterly, 13*, 141-154.

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

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

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

Witt, J. C., Gresham, F. M., & Noell, G. H. (1996). What's behavioral about behavioral consultation?, *Journal of Educational and Psychological Consultation, 7*, 327-344.

Witt, J. C., Noell, G. H., LaFleur, L. H., & Mortenson, B. P. (1997). Teacher use of interventions in general education settings: Measurement and analysis of the independent variable. *Journal of Applied Behavior Analysis, 30*, 693-696.

Yeaton, W. H., & Sechrest, L. (1981). Critical dimensions in the choice and maintenance of successful treatments: Strength, integrity, and effectiveness. *Journal of Consulting and Clinical Psychology, 49*, 156-167.

Author Contact Information:

Lisa Hagermoser Sanetti,
Department of Educational Psychology(U-2064),
University of Connecticut,
Storrs, Connecticut 06269-2064;
E-mail: lisa.sanetti@uconn.edu

Thomas R. Kratochwill,
Department of Educational Psychology,
University of Wisconsin—Madison,
1025 West Johnson Street,
Madison, Wisconsin 53706
E-mail: tomkat@education.wisc.edu