

UC Riverside

UC Riverside Electronic Theses and Dissertations

Title

Using Performance Feedback to Increase Treatment Integrity in Behavioral Interventions:
Meta-Analysis

Permalink

<https://escholarship.org/uc/item/81j7r0z7>

Author

Cipani, Alessandra

Publication Date

2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Using Performance Feedback to Increase Treatment Integrity
in Behavioral Interventions: Meta-Analysis

A Thesis submitted in partial satisfaction
of the requirements for the degree of

Master of Arts

in

Education

by

Alessandra Cipani

August 2016

Thesis Committee:

Dr. Cathleen Geraghty, Chairperson

Dr. Austin Johnson

Dr. Katherine Stavropoulos

Copyright by
Alessandra Cipani
2016

The Thesis of Alessandra Cipani is approved:

Committee Chairperson

University of California, Riverside

ABSTRACT OF THE THESIS

Using Performance Feedback to Increase Treatment Integrity
in Behavioral Interventions: Meta-Analysis

by

Alessandra Cipani

Master of Arts, Graduate Program in Education
University of California, Riverside, August 2016
Dr. Cathleen Geraghty, Chairperson

This study examined the effects of performance feedback on the treatment integrity of various behavioral interventions occurring inside a classroom setting. Nine studies were analyzed across seven moderator variables (school, classroom, intervention, latency of feedback, type of feedback, consultee position and consultee experience). This meta-analysis was driven by three research questions: a) What is the magnitude of performance feedback on the treatment integrity of behavioral interventions? b) Is performance feedback an effective intervention for more intensive behavioral concerns? c) What factors make performance feedback more beneficial? Major findings in this study included small to large effect sizes, with larger effects seen in special education classrooms and interventions for behavior-specific praise. These findings suggest that performance feedback can be used across school and classroom types as well as to address intensive behavioral concerns.

TABLE OF CONTENTS

Introduction.....	1
Methods.....	9
Results.....	16
Discussion.....	18
References.....	25

Using Performance Feedback to Increase Treatment Integrity in Behavioral Interventions: Meta-Analysis

After the renewal of the Individuals with Disabilities Education Improvement Act (IDEIA, 2004), schools are now mandated by law to address problem behaviors when they begin to impede learning. Unfortunately, this often happens as a reactive strategy, occurring only after the behavior has had a significant, negative impact on the student's education. Problem behavior is the single most common reason students with disabilities are removed from their school setting (Reichle, 1990). Extreme problem behaviors occurring in both general and special education classrooms can lead to a host of negative outcomes including poor teacher-student relations, a student's change of placement and teacher burnout. A change of placement for behavioral concerns occurs when all other avenues of support have been exhausted. Mandated by law, this includes providing access to mental health services and implementing a function-based intervention to help remediate the problem behavior. When the need for support becomes that intensive, it is assumed that the behavior is so pervasive, the student is not benefiting from their current placement. In these cases, a change of placement to a more restrictive environment occurs. It is surmised that the additional support provided to remediate these extreme behaviors should continue within the restricted setting. The need for evidence-based practices (EBP) and function-based interventions is widespread, occurring across grades and within both general and special education. Even when an EBP is implemented, evidence suggests that the effectiveness or the impact of an EBP is reduced when not implemented with high integrity or fidelity (Walrath, 2006). To implement these

strategies with fidelity, teachers must acquire behavior management skills that may not be in their repertoire. Self-report data show that teachers do not feel prepared to handle extreme problem behaviors and are reporting low levels of self-efficacy regarding behavior management (Rimm-Kaufman & Sawyer, 2004).

Treatment Integrity

Known as the “Curious Double Standard” (Peterson, Homer, & Wonderlich, 1982), researchers in applied behavior analysis fields continually focused on the dependent variable in research, but not the implementation of the independent variable. When we consider that the field would not accept a change in the dependent variable without sufficient data, it is troublesome that the independent variable is often accepted at face value regardless of the implementation data provided. Intervention success is typically measured by student outcomes as the dependent variable. Inadequate assessment and implementation may result in faulty conclusions about the relationship between the independent variable and dependent variable. Shadish (2002) posits that there are four types of experimental validity that allow researchers to draw conclusions: internal, external, construct and statistical conclusion. Internal validity is the extent to which the change in the dependent variable can be unequivocally attributed to a systematic change in the independent variable. External validity refers to the extent to which the causal relationship can be generalized. Construct validity refers to interpretation or explanation of the causal relation between the independent and dependent variables. Statistical conclusion validity refers to the conclusions drawn about the effectiveness of the intervention from quantitative analysis. A threat to statically

conclusion validity is the variability in which the treatments are implemented, also known as treatment integrity. Increased variability within the data can lead to a higher standard deviation and in turn a lower effect size when using single-case or single-subject designs. As previously stated, many factors contribute to student success but one that is not often measured is the extent to which the intervention is correctly being implemented (Noell, 1999). This is known as treatment fidelity or treatment integrity. More specifically defined to intervention research, it is the degree to which an independent variable is implemented as intended (Gresham, Gansle, & Noell, 1993). Treatment integrity is imperative to Shadish's (2002) four proposed types of validity. For example, if a reported student outcome is positive, but the intervention was not implemented with fidelity, can we still assume that the intervention was the cause of the improvement and not just natural growth or other extraneous variables? On the other hand, if student outcomes are not favorable, can we assume that the quality and application of the intervention was as intended? This leads to a host of problems including a lack of generalization and even a possible change in placement into special education or from special education into a more restrictive setting. Gresham, Gansle, and Noell (1993) reviewed studies published in the *Journal of Applied Behavior Analysis (JABA)* and found only 16% of studies published from 1980 to 1990 included data on treatment integrity. A more recent review of over 200 school-based intervention studies published between 1991 to 2005 in *JABA* revealed that only 30% included treatment integrity data (McIntyre, Greshman, DiGennaro, & Reed, 2007). Lastly, Sanetti, Gritter and Dobey (2011) reviewed over 200 intervention studies published across four school psychology

journals and found that approximately half of the studies did not report quantitative treatment integrity data. This phenomenon is seen in both research settings and when applied in practice. Gresham (1989) outlined several possible reasons on why a previously agreed upon or even mandated treatment plan is not being implemented as intended. They generally fall into two categories: variables related to intervention and variables related to the interventionist. These can include: complexity of the plan, number of treatment agents, time required of them to implement, resources required and the perceived effectiveness of the plan. As previously mentioned, teachers and staff are not equipped or properly trained to implement behavioral interventions with high fidelity.

Professional Development

Traditional models of professional development usually consist of a one-time meeting or training. Despite its popularity and the required attendance, professional development has shown little effectiveness in long-term gains and behavior change (Kinkead, 2007). Reviews have frequently pointed out its flaws and ineffectiveness (Kennedy, 2005; Wang et al., 1999). Professional development for educators is usually mandated by district policy and occurs over the summer or a few times throughout the year. When the professional development has concluded, ongoing support is not typically offered. While professional development leads to an increase in general knowledge via post-test data, follow up reports show little to no behavior change (Kinkead, 2007). Driscoll and colleagues (2010) reported that educators were 10 times more likely to implement an intervention with fidelity when given some type of additional or ongoing support. This finding is even more jarring when regarding behavior that is of a greater

concern. Therefore, it is critical to extend more support to those who are providing direct services to students exhibiting high levels of maladaptive behaviors (Erchul & Martens, 2010).

Behavioral Consultation

Consultation, or the interaction between two professionals to address an area of concern (Erchul & Martens, 2008), is frequently utilized within educational settings. The roles in consultation include the consultant, who is usually an expert in the area, a consultee, and a client. In education, the consultee is most commonly the teacher, who is looking to improve the behavior of their student, the client (Sheridan, Welch, & Orme, 1996). The consultant can range from an individual employed at the school, such as a certificated staff member or school psychologist to a researcher attempting to field a study.

Of the consultation models, the most widely used and empirically supported is the behavioral consultation (BC) model (Wilkinson, 2006). The BC model traditionally consists of four parts: (1) problem identification, (2) problem analysis, (3) treatment implementation, and (4) treatment evaluation (Kratochwill & Bergan, 1990). Similar to professional development, BC still takes place over a set period of time or until the four stages are completed. However, it is more in depth and specific to problem behaviors than traditional professional development. Although research has shown that BC is efficacious (Sheridan et al., 1996), teachers and staff still report low levels of treatment fidelity within 7-10 days after beginning an intervention (Mortenson & Witt, 1998; Noell

et al., 2000).

Performance Feedback

As an addition to behavioral consultation, performance feedback can be used to monitor the behavior of a consultee and client, as well as provide objective and specific feedback regarding plan implementation (Frank & Kratochwill, 2008). Stemming from the organizational psychology literature, performance feedback is a tailored method of ongoing consultation in which a consultant collects data on the components of an intervention, as well as the target behavior of a student. By presenting the implementation data alongside student outcomes, the consultee is made aware of the functional relationship between their behavior and their students (Mortenson & Witt, 1998). Performance feedback is utilized in traditional behavioral consultation during the implementation phase. Although treatment integrity is still relevant during other phases of BC that include the Problem Identification Interview (PII) and Problem Analysis Interview (PAI), it is assumed that those have been completed correctly when implementation begins. Ideally, performance feedback would occur after the target behavior has been identified and operationally defined through some type of functional assessment. The literature on performance feedback has shown a variety of different types including verbal, written or visual (graphical). Performance feedback can occur immediately after an observation or occur at a daily or weekly rate. Meetings typically happen in person but can be remote via email or video chatting. Once performance feedback has begun, it can be tailored to fit the needs of the consultant and more

importantly, the consultee. Reports indicate that performance feedback can be completed within three to five minutes, and is favorable to teachers.

Performance Feedback to Increase Intervention Implementation

Currently, performance feedback is the most research supported method for increasing treatment integrity in schools (Fallon, Collier-Meek, Maggin, Sanetti, & Johnson, 2015). Several meta-analyses have been conducted to examine the effectiveness of performance feedback to increase treatment integrity. Solomon, Klein, and Politylo (2012) reviewed 36 single-case studies that employed performance feedback to improve treatment integrity for various academic and behavioral interventions as well as teacher skills. The authors found moderate effect sizes across the combined sample that ranged from preschool through high school in both general education and special education settings. It was found to be more effective for academic interventions than behavioral interventions, which may be due to measurement error from the original studies. Immediate feedback had larger effects than weekly feedback and special education settings reported larger effect sizes than general education settings.

Fallon and colleagues (2015) applied the What Works Clearinghouse (WWC) guidelines (Kratochwill et al., 2010) to the body of performance feedback research. Of the 111 studies found in the literature search, 47 studies were included in the review. Of those, only 48% of studies met design standards, while 27% met design standards with reservation. The systematic review found that most often, PF was given in person, using visual (written or graphic) feedback.

Currently literature and systematic reviews have concluded that performance feedback is an evidence-based intervention that can be used across settings, populations, and context (Gilbertson, Witt, Singletary, & Vanderheyden, 2007). While various components of performance feedback have been analyzed to determine what is most commonly being used, it is still unknown which components work best for extreme behaviors and restricted settings. As previously stated, problem behaviors are the most common cause of students being removed from the current classroom either from expulsion or a change in placement.

Performance Feedback in an MTSS

In a multi-tiered model of support, reactive strategies can be employed at a Tier 1 level through classroom management techniques. These include evidence-based practices (EBP) such as behavior-specific praise, error correction and prompting. These are universal supports that are fairly easy to implement and can have positive effects on the entire class. When a student does not respond to Tier 1 accommodations and is displaying more severe problem behaviors, a Tier 2 intervention should be put in place. At this point, we are surmising that the student is not responding to behavioral cues and environmental changes within the classroom and is in need of greater support. Finally, at Tier 3, we are implementing a function-based behavior intervention plan (BIP) in hopes of remediating the behavior before more disciplinary actions or a change of placement is suggested. A functional relationship should exist between intensity of the behavior, the amount of services received, and the support given to the service providers. In other words, support for the teachers and staff should increase as the support for the problem

behavior increases. An important aspect of behavioral intervention plans is the teaching strategies required for the new desired behavior. If we are assuming there is a skill deficit, we are explicitly teaching the behavior and providing frequent and appropriate feedback to the student. If we are assuming that staff do not possess these advanced behavioral techniques, the same level of support should be given to those responsible for implementation.

Purpose of the Study

The purpose of this study is to examine the effects of performance feedback on the implementation of behavioral interventions for extreme problem behaviors. Currently, no other study examines the use of performance feedback on behavioral interventions for students with disabilities, in restricted setting and displaying high levels of maladaptive behavior. This critical review will examine specific components of performance feedback given to service providers of students with disabilities. Three research questions guided this study: a) What is the magnitude of performance feedback on the treatment integrity of behavioral interventions? b) Is performance feedback an effective intervention for more intensive behavioral concerns? c) What factors make performance feedback more beneficial?

Methods

A computer search was employed using PsychINFO, ERIC, and Google Scholar. The keywords used were “*performance feedback*” or “*consultation*” with “*intervention*” and “*treatment integrity*” or “*treatment fidelity*”. The results were constrained to 2004-

2014. A hand search of *JABA* and *Journal of Educational and Psychological Consultation* was also conducted from 2012-2014. The results of the searches yielded a total of approximately 95 studies. Of those, approximately 30 were in peer-reviewed journals. Twenty-seven of those studies employed single-case, multiple baseline design and nine met the criteria for inclusion in the meta-analysis.

Single-Case Design

Single-case design studies are frequently utilized in educational and behavioral research to demonstrate the effect due to the introduction of an independent variable (Kazdin, 1982; Horner et al., 2005). By using a multiple baseline design, the researcher can control for context and/or time allowing them to hypothesize a functional relationship between the independent and dependent variables. To increase the generalizability of these findings, meta-analysis is used to test moderators and weigh effect sizes across studies (Hedges & Olkin, 1985; 2014). Single-case designs can be combined and examined across different populations, contexts, and aspects of the independent variable.

Criteria for Inclusion in a Meta-Analysis

For a study to be included in the meta-analysis, the following criteria must be met:

- 1) The study must employ a single-case research design (e.g., multiple baseline, alternating treatment)
- 2) The study provided all necessary information to calculate effect size
- 3) The study must have at least three data points at baseline phase, as well as three data points at treatment (Performance Feedback) phase

- 4) The study must be measuring intervention implementation (treatment integrity) as the dependent variable
- 5) The study must implement performance feedback during treatment phases
- 6) The dependent variable must be measured in each treatment phase
- 7) Inter-rater agreement must be calculated in study
- 8) Study must be measuring a behavioral intervention strategy

Coding and Inter-rater Agreement

Categories for coding included treatment, quality of the study, and the type of dependent measures. Inter-rater agreement was conducted for 80 % of the studies with a peer graduate student in the School Psychology PhD program. Inter-rater agreement for inclusion was 92% and 89% for coding. This falls beyond the minimum threshold of 80% for inter-rater agreement.

Coding of Dependent Measures

To be included in this study, dependent measures or outcome measures must be clearly specified. For this study, this will be defined as the number or percentage of components of an intervention that were correctly implemented. This is also known as treatment integrity. The interventions being implemented serve as a moderator and are broken up into three categories: an established behavior support plan (BSP), class-wide behavior plan, or the amount of behavior-specific praise provided by the teacher. The intervention was coded as a BSP, if the intervention contained several different strategies (reactive, antecedent, consequence) and was currently being implemented within the classroom setting. A class-wide behavior plan was defined as an intervention strategy that

targets the behavior of an entire classroom. Behavior-specific praise statements were operationally defined to include a specific positive statement said after a desired behavior.

Coding of Treatment Variables

The treatment variables include the delivery of performance feedback. This was categorized as “immediately” or “same day”. Immediately is operationally defined as immediately after concluding the data collection interval. This does not include any type of in vivo coaching. Another aspect was the display of the performance feedback. This was categorized as “verbal”, “written” or “graphic”. A study was coded as “verbal” if no other supplemental information was given. Studies that implemented “written” feedback obviously used verbal strategies, but did not have any type of graphical representation.

Moderators

To examine variation between studies, seven moderators were chosen. For this study, the moderators included are:

- 1) Type of institution (public or private)
- 2) Classroom (special education or general education)
- 3) Type of behavioral intervention (BSP, class-wide & behavior-specific praise)
- 4) Participant position (special education teacher, general education teacher, paraprofessional & student teacher enrolled in master’s level program)
- 5) Participant’s experience in current role (0-4 years, 5-10 years, & 10+ years)
- 6) Latency of performance feedback (immediate or same day)

7) Type of performance feedback (verbal, written, & graphic)

Other aspects of the studies were examined but due to the lack of uniformity, they will not be included in the analysis. Those include: social validity measures, role of the consultant, and student outcome data. The sample includes students who are receiving services due to a heightened level of need. Disability information will not be used as a moderator in statistical analyses at this time due to logistical issues (comorbidity, teacher-student dyads). Disability was coded as “*none*” or “*yes*”. Extreme problem behavior or intensive behavioral concerns is defined as “*physical or verbal aggression, tantrums and inappropriate sexual behaviors*”. This was also coded as “*none*” or “*yes*”. Diagnostic information as well as a description of the behaviors will be provided in the discussion.

Effect Size Calculation

Along with visual analysis, quantitative procedures were also done. Visual analysis is used in single-case research to determine whether the presentation of an independent variable (e.g. an intervention) has altered the data of a participant. There are three components of visual inspection: level, trend, and variability. Level refers to the position of the data on the Y-axis, or in other words, where the observations are falling on the graph. The trend refers to the direction or pattern the data are pulling towards. Trends can be increasing, decreasing, cyclical or curvilinear. Variability refers to how divergent or spread out the data are within their own phases (i.e. baseline or treatment). If there is a lack of stability within the phases, this could effect the visual and statistical analysis of the data.

Kratochwill and colleagues (2010) suggest considering three additional variables when examining single-case research. These include: immediacy of the effect, overlap, and consistency of data patterns across phases. Kratochwill et al., (2010) also suggest following four steps when conducting visual analysis. The first step is documenting a predictable baseline pattern. If the baseline is following the same hypothesized trend of the treatment phase, it is not stable or predictable. The second step entails examining the within-in phase patterns, while the third step compares the data with an adjacent phase. This is done to assess whether the independent variable has demonstrated an effect in the treatment phase. The fourth and final step is to integrate all the information to determine if there were at least three demonstrations of an effect at three different points in time.

Percentage of Non-Overlapping Data (PND) was also used to evaluate the data. PND is calculated by taking the percentage of data points in phase B (treatment) that fall above the highest point in phase A (baseline). When a study has more than one sample, PND can be aggregated by taking the median percentage, which is less effected by outliers. PND is a nonparametric measure that takes baseline variability into account and does not assume that the reported data are independent of each other. The limitations of PND include the lack of sensitivity to slope changes as well as ceiling effects for highly successful interventions. When interpreting the range of PND (0-100%), Scruggs, Mastropieri, Cook, and Escobar (1986) have identified explicit criteria. It is stated that a percentage under 50% reflects an unreliable treatment, 50-70% reflects a questionable treatment, 70-90% represents fairly effective and anything above 90% is classified as a

highly effective treatment. Since PND is not an effect size, improvement rate difference (IRD) was also calculated.

Improvement rate difference (IRD) is a nonparametric measure that is found by calculating the percentage of improvement between the baseline and treatment phases (Parker, Vannest, & Brown, 2009). The use of IRD for single-case studies stems from group medical research that models the “risk difference” theory. To calculate IRD, you must identify the unimproved data in the baseline phase that do not overlap (equal or exceed) any data points in the treatment phase. This is also done in the treatment phase, where unimproved data should equal or fall below one or more baseline data points. The maximum IRD score is 100% or 1.00, which signals that all data points within the intervention exceeded all baseline scores. If an IRD score equals .50, the improvement from baseline to treatment phases is only chance-level. Parker and colleagues (2009) identified IRD scores at or about .50 and below to be small. Effects sizes of around .50 to .70 are considered moderate and effect sizes of .70 and higher are considered large. With that being said, IRD can easily be used alongside PND but allows for better sensitivity and an interpretable effect size. Robust IRD is comparable to robust Phi, a PAND coefficient. PND will be calculated by hand while data will be processed for IRD using Number Cruncher Statistical Software (NCSS, Hintze, 2002).

Results

Characterization of Studies

This review included nine studies in the analysis, which resulted in a total sample size of 30. The average number of baseline and treatment sessions were five and eleven, respectively. The average number of participants per study was three. For all study characteristics see table 1. Of the nine studies, the majority were published in the *Journal of Applied Behavior Analysis*. The other studies were found in: *Journal of Behavioral Intervention, Journal of Positive Behavior Education, Journal of Education and Treatment of Children and Behavioral Modification*.

The meta-analysis was guided by an overarching research questions. a) What is the magnitude of performance feedback on the treatment integrity of behavioral interventions? Overall, the effect sizes for each study ranged from small to large (.5-.92). Similar results were seen for PND (62% to 88%). This suggests that overall, performance feedback has a positive functional relationship with treatment integrity. Reinke, Lewis-Palmer and Martn (2007) produced the largest effect size. The study employed a multi-base line design to test the effects of visual performance feedback on behavior-specific praise. The smallest effect size (Minor et al., 2014) employed a multiple-base line design to conduct written feedback on BSP. For all PND and IRD scores and effect sizes see table 2.

Diagnoses and Problem Behaviors

The second research question examined whether performance feedback an effective intervention for more intensive behavioral concerns? Medium to large effect

sizes were seen in studies with samples including individuals with Autism Spectrum Disorders (Minor, et al., 2014), conduct disorder (Coddling, et al., 2008), emotional/behavioral disorder (Duchaine, et al., 2011; Hawkins, 2010; Sutherland, 2000), traumatic brain injury (Coddling, et al., 2005; Digennaro, 2007), and attention hyper deficit disorder (Duchaine, 2011; Coddling, 2008). Due to high comorbidity and unclear teacher-student dyads, it was not possible to use diagnosis as a moderator. Eight of the nine studies reported extreme behaviors such as physical aggression, verbal aggression or outburst, tantrums and inappropriate sexual behavior. Their effect sizes ranged from .65 to .90.

Characterization of Moderators

This review examined the relationship between moderators and effect size through the third research question: What factors make performance feedback more beneficial? Moderator variables included: institution (private vs. public), classroom (special education vs. general education) type of behavioral intervention (BSP, class-wide & Specific Praise), teacher experience (0-4 years, 5-10 years, and 10+ years), teacher position (special education teacher, general education teacher, paraprofessional & student teacher enrolled in master's level program), latency of feedback (immediate vs same day), and type of feedback (verbal, written or graphic). Teacher experience ranged from less than one year to 29 years, with an average of seven years. Effect sizes aggregated by moderators are shown in table 3.

Discussion

Overall the effect sizes yielded from this study were small to large. Of the moderators examined, delivery of feedback had a significant impact on effect sizes with graphic displays producing the largest effects, followed by written and verbal. Since only five of the nine studies provided student outcome data, it is unclear whether or not displaying both TI data and student outcome data simultaneously would have produced larger effect sizes. Of the student outcome data reported, four studies included either on-task behavior or disruptive behavior. Another aspect of performance feedback to investigate is the impact of reporting negative student behaviors versus on-task or replacement behaviors. The effect sizes for latency of feedback did not differ. As the studies were categorized as *immediately* or *same day*, it is not surprising that there was not a difference between the two. The behavioral intervention that produced the largest effect size was behavior-specific praise. The unit of measurement for these studies was the amount of behavior-specific praise statements said by the teacher/staff during the observation session. These results are not surprising, as initiating behavior-specific praise statements is by far easier to do than other complex interventions (Cavanaugh, 2014; Stormont & Reinke, 2014). Several of the baseline data for these studies reported zero rates of behavior-specific praise, leading to very large PND and IRD sizes. Class wide behavior plans also produced large effect sizes. In this specific study (Coddling, Livianis, & Pace, 2008) three general procedures took place: token economy, neutral statements/planned ignoring, and time-out. It is unclear whether “time-out” was referring to time-out from reinforcement, or in the more traditional sense of being removed from a

situation. The authors also included the aspect of observer reactivity, by collecting data through a one-way mirror for 34% of the observations. There were no significant differences between the observer-present and observer-absent phases. The third option for this moderator was an established behavior support plan (BSP), currently referred to as a behavior intervention plan (BIP). This showed the smallest effect size and there are several hypothesized reasons why. BSPs are the most complex and should have several different components (antecedent strategies, consequence strategies, differential reinforcement of alternative behavior). Using the treatment integrity sheet designed by Coddling et al., (2005), TI data for a BSP is calculated by divided the number of correctly implemented components by the total number of components. Several studies reported that the BSP was put in place after the results of a functional behavioral assessment (FBA). However, it is unknown if the correct function was found in the FBA and therefore addressed in the BSP. If a teacher or service provider does not target the correct function, the student outcomes will not improve, regardless of whether or not the treatment integrity improves. Hypothetically, this could cause the service provider to stop implementing the BSP, if they are not seeing a difference in the student's behavior. With that being said, effect sizes across all three behavioral interventions ranged from medium to large.

Environmental Factors

The effect sizes for private and public schools did not differ. This was added as a moderator under the impression that there may be different types of systems-level pressure from the either a district in the case of public school or the stakeholders when

addressing private schools. While private schools are not held to the same to the same standards under IDEA, often times special populations are more likely to enroll in a private school that can cater to their needs. In those cases, it is more likely that teachers and staff will receive a higher level of training. The effect sizes were larger for participants in special education or inclusion classrooms than in general education. This was also found by Solomon et al., (2012), who included both academic and behavioral interventions in their review. Class size was not reported in all studies so it could not be used as a moderator. However, special education classrooms always have less students, sometimes only a third of an average general education classroom. The studies in the current review included elementary schools, middle schools and high schools but due to the specific nature of the schools, that was not added as a moderator. For example, two of the studies took place at a school for traumatic brain injuries and another study occurred at a non-public community day school. These schools usually service students who range from 13-22 years of age, making it inappropriate to categorize using as traditional labels.

Consultee Factors

For this study, two moderators focused on the characteristics of the consultee: participant position and years of experience in said position. The positions reported across the studies were broken into four categories: special education teacher, general education teacher, paraprofessional, and student teacher enrolled in Master's level program. Results showed the largest effect sizes for paraprofessionals, which were included in one study (Coddling et al., 2008). The role of a paraprofessional or instructional aide varies across classroom settings. In this case, it may be safe to assume that the paraprofessionals

benefited most from performance feedback because the implementation was already an important aspect of their role inside the classroom. Special education teachers had larger effect sizes than general education teachers, which corresponds with the findings above regarding classroom type. One study (Minor et al., 2014) had a sample of student teachers enrolled in a Master's level program in special education. This study returned the smallest effect size. Consultee experience was measured in years and broken into three categories: 0-4 years, 5-10 years, and 10+ years. Effect sizes did not greatly differ, but did follow a decreasing trend based on experience. In other words, the largest effect sizes were reported for newer teachers (> four years) and the smallest for more experienced teachers, though not by a discernible amount. Several of the studies reported and described the type and amount of professional development the participants had received before or during the study. Unfortunately, not enough information was given across the studies for that to become a moderator. Additional support ranged from professional development, occurring once per year to monthly clinical meetings and extensive training in applied behavior analysis. It can be hypothesized that if the data were available, a functional relationship would be seen between the quality of training and treatment integrity.

Social Validity

Unfortunately, not every study used in this meta-analysis reported enough information regarding social validity so it could not be used as a moderator. Seven of the nine studies did employ a social validity measure, however most were researcher created and reported qualitatively. The Treatment Acceptability Rating Form-Revised (TARF-R:

Reimers & Wacker, 1988) and Intervention Rating Profile-15 (IRP-15) (Martens, Witt, Elliot, & Darveaux, 1985) were used in two separate studies. The TARF-R consists of 20 items in a Likert type format (1-7point scale). The IRP-15 is a 15-item single factor scale also reported in Likert type format. Studies that used researcher created forms ranged from 3-15 items, usually in a Likert type format as well. Social validity is an important aspect of all intervention studies where researchers are looking to promote and sustain behavior change. While some psychologists have stated that performance feedback is too intrusive or aversive to be used as a support, the body of research as whole finds teachers reporting it is favorable (Sanetti & Kratochwill, 2014).

Limitations

The current study had several limitations. To begin, as performance feedback is a relatively new aspect of consultation, there was a small body of research to choose from. This was compounded by not including studies that used coaching or other in vivo consultation methods. An issue regarding meta-analyses in particular is the file drawer effect (Rosenthal, 1979). Also known as publication bias, the file drawer effect refers to the sparse amount of published studies that report null results. Efforts to control for publication bias, such as down-weighting non-randomized studies were not used in this meta-analysis given the nature of performance feedback and behavioral intervention studies. This meta-analysis did not employ as stringent criteria as previous reviews such as Fallon et al., (2015) who used What Works Clearinghouse (WWC) Standards. According to WWC, single-case designs should be conducted by at least three different research teams with no overlapping authorship at three different institutions.

Unfortunately, this was not possible as several of authors did overlap. Another limitation exclusive to single-case research is the inflation of effect sizes. Within this type of research, there is only so much that can be done to control for inflated effect sizes.

Several limitations arose within the selected studies. For example, the role of the consultant across studies was primarily the researcher, which does not provide support for closing the research to practice gap. It would have been beneficial to assign the role of consultant to experts within the school and provide support for them as well. In a real-world application, performance feedback would need to be provided by a school employee who works alongside the consultee to implement and monitor the intervention. Unfortunately, we cannot draw conclusions about the maintenance period after the performance feedback has stopped since not all the studies reported follow up data.

Conclusions

This meta-analysis reviewed nine studies that employed performance feedback to improve treatment integrity across various behavioral interventions. The results provide support for previous meta-analyses (Solomon et al., 2012; Fallon et al., 2015) that found performance feedback to be effective across contexts and amongst various populations. Notable findings in this study include large effect sizes for simple behavioral strategies such as behavior-specific praise as well as large effect sizes for special education classrooms and teachers. From these results, we can conclude that performance feedback is a supported intervention for students with intensive behavior problems. The students in the sample met criteria for disabilities including Autism Spectrum Disorders, emotional and behavioral disorders, seizure disorders, traumatic brain injuries and attention hyper-

deficit disorder. The behaviors targeted ranged from inappropriate verbalizations, out-of-seat behaviors to inappropriate touching and self-injurious behaviors. In restrictive settings, performance feedback was shown to increase treatment integrity of the intervention, as well as student outcomes when provided. In less restrictive settings, services provided to staff should be matched to the level of need of the student, mirroring a multi-tiered system of support. To make performance feedback more acceptable, it should be embedded into the intervention as a reactive strategy, as opposed to a response to poor implementation. It is unknown which hypothesized functional property (positive reinforcement, negative reinforcement, discriminative stimulus) of performance feedback makes it more effective. Digennaro, et al., (2007) employed a meeting cancellation contingency (negative reinforcement), which was found to be effective. More research involving alternating treatments may shed some light on the functional properties of performance feedback as well as the importance of the consultant-consultee relationship. In effort to close the research-to-practice gap, more research is needed to generalize the consultant role from researcher to school personnel. By utilizing social power in behavioral consultation (Erchul & Raven, 1997; 2001), school psychologist can use both expert and referent power to improve on consultation skills and implement performance feedback. Overall, these results support previous research regarding the effectiveness of performance feedback and add new insight into the components of performance feedback and its use with extreme behavior concerns. Performance feedback aligns with the current paradigm shift in school psychology, from problem admiration to problem solving in a multi-tiered system of support

References

- Cavanaugh, B. (2013). Performance feedback and teachers' use of praise and opportunities to respond: A review of the literature. *Education and Treatment of Children, 36*(1), 111-136.
- *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*(2), 205-219.
- *Codding, R. S., Livanis, A., Pace, G. M., & Vaca, L. (2008). Using performance feedback to improve treatment integrity of classwide behavior plans: An investigation of observer reactivity. *Journal of applied behavior analysis, 41*(3), 417-422.
- *Digennaro, F. D., Kleinmann, A. E., & Hanley, G. (2007). A comparison of performance feedback procedures on teachers' treatment implementation integrity and students' inappropriate behavior in special education classrooms (B. K. Martens, Ed.). *Journal of Applied Behavior Analysis, 40*(3), 447-461. doi: 10.1901/jaba.2007.40-447
- Driscoll, M. (2010). *Web-based training: Creating e-learning experiences*. John Wiley & Sons.
- *Duchaine, E. L., Jolivette, K., & Fredrick, L. D. (2011). The effect of teacher coaching with performance feedback on behavior-specific praise in inclusion classrooms. *Education and Treatment of Children, 34*(2), 209-227. doi: 10.1353/etc.2011.0009

- Erchul, W. P., & Martens, B. K. (2010). *School consultation: Conceptual and empirical bases of practice- 3rd edition*. New York, NY: Springer.
- Erchul, W. P., & Raven, B. H. (1997). Social power in school consultation: A contemporary view of French and Raven's bases of power model. *Journal of School Psychology, 35*(2), 137-171.
- Erchul, W. P., Raven, B. H., & Ray, A. G. (2001). School psychologists' perceptions of social power bases in teacher consultation. *Journal of Educational and Psychological Consultation, 12*(1), 1-23.
- Erchul, W., & Sheridan, S. (2008). *Handbook of Research in School Consultation*. New York, NY: Lawrence Erlbaum Associates.
- Fallon, L. M., Collier-Meek, M. A., Maggin, D. M., Sanetti, L. M. H., & Johnson, A. H. (2015). Is performance feedback for educators an evidence-based practice? A systematic review and evaluation based on single-case research. *Exceptional Children, 81*(2), 227-246.
- Frank, J. L., & Kratochwill, T. R. (2008). School based problem-solving consultation: Plotting a new course for evidence-based research and practice in consultation. In W. Erchul, & S. Sheridan (Eds.), *Handbook of Research in School Consultation* (pp. 13-30). New York, NY: Lawrence Erlbaum Associates.
- Gilbertson, D., Witt, J. C., Singletary, L. L., & Vanderheyden, A. (2007). Supporting teacher use of interventions: Effects of response dependent performance feedback on teacher implementation of a math intervention. *Journal of Behavioral Education, 16*(4), 311-326. doi: 10.1007/s10864-007-9043-0

- Gresham, F. M. (1989). Assessment of treatment integrity in school consultation and prereferral intervention. *School Psychology Review*, 18, 37-50.
- Gresham, F. M., Gansle, K., & Noell, G. H. (1993). Treatment integrity in applied behavior with children. *Journal of Applied Behavior Analysis*, 26, 257-263.
- *Hawkins, S. M., & Heflin, L. J. (2010). Increasing secondary teachers' behavior-specific praise using a video self-modeling and visual performance feedback intervention. *Journal of Positive Behavior Interventions*.
- Hedges, L., & Olkin, I. (1986). Meta analysis: A review and a new view. *Educational Researcher*, 15(8), 14-16.
- Hedges, L. V., & Olkin, I. (2014). *Statistical methods for meta-analysis*. Academic press.
- Hintze, J. (2001). NCSS and Pass. *Number Cruncher Statistical Systems, Kaysville*.
Available at: www.ncss.com.
- 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.
- Individuals with Disability Education Act Amendments of 2002 [IDEIA]. (2004).
Retrieved from <http://thomas.loc.gov/home/thomas.php>
- Kazdin, A. E. (1982). Single-case experimental designs: Strategies for studying behavior change.
- [Kennedy, A. \(2005\). Models of continuing professional development: a framework for analysis. *Journal of in-service education*, 31\(2\), 235-250.](#)

- Kennedy, C. H. (2005). *Single-case designs for educational research*. New York, NY: Allyn and Bacon.
- Kinkead, S. (2007). *Improving instruction through coaching*. Olympia, WA: The Center For Strengthening the Teacher Profession, Superintendent of Public Instruction.
- Kratochwill, T. R., & Bergan, J. R. (1990). *Behavioral consultation in applied settings: An individual guide*. New York, NY: Plenum.
- Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., et al. (2010). *Single-case designs technical documentation*. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf
- *Mesa, J., Lewis-Palmer, T., & Reinke, W. (2005). Providing Teachers with Performance Feedback on Praise to Reduce Student Problem Behavior. *Beyond Behavior*, 15(1), 3-7.
- McIntyre, L. L., Gresham, F. M., DiGennaro, F. D., & Reed, D. D. (2007). Treatment integrity of school-based interventions with children in the Journal of Applied Behavior Analysis 1991–2005. *Journal of Applied Behavior Analysis*, 40(4), 659-672.
- *Minor, L., DuBard, M., & Luiselli, J. K. (2014). Improving intervention integrity of direct-service practitioners through performance feedback and problem solving consultation. *Behavioral Interventions*, 29(2), 145-156.
- 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*.

- Noell, G. H., Duhon, G. J., Gatti, S. L., & Connell, J. E. (2002). Consultation, follow-up, and implementation of behavior management interventions in general education.
- Noell, G. H., & Witt, J. C. (1996). A critical re-evaluation of five fundamental assumptions underlying behavioral consultation. *School Psychology Quarterly*, 11(3), 189-203.
- Noell, G. H., & Witt, J. C. (1999). When does consultation lead to intervention implementation? Critical issues for research and practice. *The Journal of Special Education*, 33(1), 29-35.
- 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(3), 271-284. doi: 10.1901/jaba.2000.33-271
- Parker, R. I., Vannest, K. J., & Brown, L. (2009). The improvement rate difference for single-case research. *Exceptional Children*, 75(2), 135-150.
- Peterson, L., Homer, A. L., & Wonderlich, S. A. (1982). The integrity of independent variables in behavior analysis. *Journal of applied behavior analysis*, 15(4), 477-492.
- Reichle, J. (1990). National working conference on positive approaches to the management of excess behavior: Final report and recommendations. *Minneapolis, MN: Institute on Community Integration, University of Minnesota.*

- *Reinke, W. M., Lewis-Palmer, T., & Martin, E. (2007). The effect of visual performance feedback on teacher behavior-specific praise. *Behavior Modification*, 31, 247-263.
- Rimm-Kaufman, S. E., & Sawyer, B. E. (2004). Primary-grade teachers' self-efficacy beliefs, attitudes toward teaching, and discipline and teaching practice priorities in relation to the "Responsive Classroom" approach. *The Elementary School Journal*, 104, 321-341.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological bulletin*, 86(3), 638.
- Sanetti, L. M., Fallon, L. M., & Collier-Meek, M. A. (2013). Increasing teacher treatment integrity through performance feedback provided by school personnel. *Psychology in the Schools*, 50(2), 134-150. doi: 10.1002/pits.21664
- Sanetti, L. M. H., Gritter, K. L., & Dobey, L. M. (2011). Treatment integrity of interventions with children in the school psychology literature from 1995 to 2008. *School Psychology Review*, 40(1), 72.
- Sanetti, L. M., & Kratochwill, T. R. (2014). *Treatment integrity: A foundation for evidence-based practice in applied psychology*. American Psychological Association.
- Sanetti, L. M. H., 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(4), 454-465.

- Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1986). The quantitative synthesis of single-subject research methodology and validation. *Remedial and Special Education, 8*(2), 24-33.
- Shadish, W. R. (2002). Revisiting field experimentation: field notes for the future. *Psychological methods, 7*(1), 3.
- Sheridan, S., Welch, M., & Orme, S. (1996). Is consultation effective? A review of outcome research. *Remedial & Special Education, 17*, 341-354.
- Solomon, B. G., Klein, S. A., & Politylo, B. C. (2012). The effect of performance feedback on teachers' treatment integrity: A meta-analysis of the single-case literature. *School Psychology Review, 41*, 160-175.
- Stormont, M., & Reinke, W. M. (2014). Providing performance feedback for teachers to increase treatment fidelity. *Intervention in School & Clinic, 49*(4), 219-224.
doi:10.1177/1053451213509487
- *Sutherland, K. S., Wehby, J. H., & Copeland, S. R. (2000). Effect of varying rates of behavior-specific praise on the on-task behavior of students with EBD. *Journal of Emotional and Behavioral Disorders, 8*(1), 2-8.
- Walrath, C. M., Sheehan, A. K., Holden, E. W., Hernandez, M., & Blau, G. (2006). Evidence-based treatments in the field: A brief report on provider knowledge, implementation, and practice. *The Journal of Behavioral Health Services & Research, 33*(2), 244-253.

- Wang, Y. L., Frechtling, J. A., & Sanders, W. L. (1999, April). Exploring linkages between professional development and student learning: A pilot study. In *Annual Meeting of the American Educational Research Association, Montreal*.
- Wilkinson, L. A. (2006). Monitoring treatment integrity: An alternative to the 'consult and hope' strategy in school-based behavioral consultation. *School Psychology International*, 29(4), 42

Appendix A

Table 2

PND and IRD

<u>Author</u>	<u>Median PND</u>	<u>Mean IRD</u>
Codding, Fineberg, Dunn & Pace, 2008	62%	.67
Codding, Livanis, Pace & Vaca, 2008	80%	.86
DiGennaro, Martens & Kleinmann , 2007	78%	.80
Duchaine, Jolivette, & Fredrick, 2011	85%	.90
Hawkins & Heflin, 2010	82%	.85
Mesa, Lewis-Palmer, Reinke, 2005	65%	.70
Minor, DuBard, Luiselli, 2014	55%	.65
Reinke, Lewis Palmer & Martin, 2007	88%	.92
Sutherland, Wehby, & Copeland, 2000	80%	.86

Appendix B

Table 3

Effect Size Categorized by Moderators

<u>Moderator</u>		<u>Effect Size (IRD)</u>
School	Public	.85
	Private	.87
Classroom	General Education	.70
	Special Education	.90
Intervention	BSP	.72
	Class-wide	.86
	Behavior-specific Praise	.90
Type of PF	Verbal	.60
	Written	.73
	Graphic	.89
Latency of PF	Same Day	.87
	Immediately	.90
Consultee Position	Graduate Student	.65
	General Education Teacher	.78
	Special Education Teacher	.85
	Paraprofessional	.90
Consultee Experience	10+ years	.82
	5-10 years	.85
	0-4 years	.89
