A Formative Program Evaluation of Treatment Integrity Practices, Assessments and Attitudes Within a Specialized School Setting

Lin Tang

University of Massachusetts Amherst

Follow this and additional works at: https://scholarworks.umass.edu/open_access_dissertations

Part of the Psychiatry and Psychology Commons

Recommended Citation
https://doi.org/10.7275/3533768 https://scholarworks.umass.edu/open_access_dissertations/669

This Open Access Dissertation is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.
A FORMATIVE PROGRAM EVALUATION OF TREATMENT INTEGRITY
PRACTICES, ASSESSMENTS AND ATTITUDES WITHIN A SPECIALIZED
SCHOOL SETTING

A Dissertation Presented
by
LIN TANG

Submitted to the Graduate School of the University of Massachusetts Amherst in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2012

School Psychology
A FORMATIVE PROGRAM EVALUATION OF TREATMENT INTEGRITY
PRACTICES, ASSESSMENTS AND ATTITUDES WITHIN A SPECIALIZED
SCHOOL SETTING

A Dissertation Presented

by

LIN TANG

Approved as to style and content by:

________________________________________
William J. Matthews, Chair

________________________________________
John M. Hintze, Member

________________________________________
Elizabeth Kohler, Member

________________________________________
Christine B. McCormick, Dean
School of Education
DEDICATION

I dedicate my dissertation to my family and friends.
ACKNOWLEDGMENTS

I would like to thank my advisor and chair of my dissertation, Professor William J. Matthews. Your support and encouragement was instrumental in this process. I would also like to thank Professor John M. Hintze, your expertise in the field of School Psychology inspired me to become a picky consumer of research. In addition, I would like to thank Dr. Elizabeth Kohler, you were my undergraduate professor and mentor, and my first introduction into the field of behavior. Having you on my dissertation committee, and having you share the last assignment of my graduate school career made this a meaningful transition for me.

I would also like to thank Dr. Shannon Kay, thank you for being a mentor, friend, and sister I never had. Your dedication and knowledge of all things behavior is truly inspiring, and I hope to continue learning from you and collaborating with you. A special thank you to Erica Webster, Jean Gao and Emily Green. Thank you for ensuring my treatment integrity was up to par. Also, to the teachers at the Behavior Day School, your skill and dedication serves as an inspirational model for our field.

To my LRC family, thank you for listening and being my cheering squad! To Dr. Susan B. Bronstein, our intrepid leader, mentor and friend; thank you for helping me celebrate the best days of graduate school, and put in perspective the worst days.

To my graduate school cohort, thank you for making this journey so much fun! A special thank you to Dr. Jessica Fede, Dr. James Cressey, and Dr. and Mrs. Benjamin and Emily Solomon; I am honored to have you as my friends and colleagues.

To my husband, thank you, thank you, thank you... for everything.
ABSTRACT

A FORMATIVE PROGRAM EVALUATION OF TREATMENT INTEGRITY PRACTICES, ASSESSMENTS AND ATTITUDES WITHIN A SPECIALIZED SCHOOL SETTING

SEPTEMBER 2012

LIN TANG, B.A., UNIVERSITY OF MASSACHUSETTS AMHERST
M.Ed., UNIVERSITY OF MASSACHUSETTS AMHERST
Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor William J. Matthews

The purpose of this study, a formative program evaluation, was to provide an overview of treatment integrity rates, attitudes, and assessment methods within a specialized school for children and adolescents with Autism Spectrum Disorder. The study utilized a partnership model whereby the evaluators worked collaboratively with primary stakeholders (school administrators) to determine areas of focus for the evaluation. The program evaluation was conducted between August 2009 and August 2010, spanning one school year. The evaluation participants were the school’s teachers. The evaluation was based on an exploratory case study design and included both quantitative and qualitative data. Quantitative data consisted of direct observations of treatment adherence to behavior plans in classrooms, teacher knowledge of treatment content assessed in quizzes, and observations of teacher engagement with students at recess. Qualitative data consisted of teacher surveys and focus groups. The evaluation yielded an in-depth assessment of treatment integrity within the school; an intervention to define and enhance teacher engagement during recess was also developed and implemented.
# TABLE OF CONTENTS

| ACKNOWLEDGMENTS | vi |
| LIST OF ABSTRACTS | vi |
| LIST OF TABLES | ix |
| LIST OF FIGURES | x |

## CHAPTER

### 1. PROBLEM BACKGROUND AND LITERATURE REVIEW

- Introduction ................................................................. 1
- Treatment Integrity: A Multidimensional Construct ........................................ 2
  - Defining the Construct of Treatment Integrity .......................................... 3
- Assessing Treatment Integrity ........................................................................ 6
  - Methods of Assessment ............................................................................... 8
- Relationship between Treatment Adherence, Treatment Differentiation and Interventionist Competence .......................................................... 12
- Relationship between Treatment Integrity to Treatment Outcome ................ 14
  - Treatment Outcome .................................................................................. 14
  - Integrity versus Adaptation ..................................................................... 14
- Program Evaluation ...................................................................................... 16
  - Participatory Intervention Model ............................................................... 17
  - Defining Program Evaluation .................................................................... 19
  - Formative Program Evaluation .................................................................. 20
- Applied Behavior Analysis ........................................................................... 22
  - Applied Behavior Analysis – Teaching Principles ...................................... 23
- Purpose of the Dissertation ......................................................................... 25

### 2. METHODS

- Setting and Participants ............................................................................... 27
- Evaluation Plan ............................................................................................ 31
  - Question 1 ............................................................................................... 32
  - Question 2 ............................................................................................... 34
  - Question 3 ............................................................................................... 37
  - Question 4 ............................................................................................... 38
  - Question 5 ............................................................................................... 42
  - Question 6 ............................................................................................... 45

### 3. RESULTS

- Question 1 ............................................................................................... 49
- Question 2 ............................................................................................... 50
- Question 3 ............................................................................................... 51
Question 4............................................................................................................................................ 52
Question 5............................................................................................................................................ 52
Question 6............................................................................................................................................ 59
4. DISCUSSION.................................................................................................................................... 66

Conducting a Program Evaluation within a Participatory Intervention Model (PIM) .......................................................... 67
Assessing Treatment Integrity ................................................................................................................ 69
Teacher Engagement during Recess ....................................................................................................... 72
Teacher Attitudes and Perceptions of Treatment Integrity (TI) .............................................................. 73
Summary of Program Evaluation Results ............................................................................................. 75
Limitations ............................................................................................................................................ 76
Future directions and Recommendations ............................................................................................. 77
APPENDIX: SURVEY ............................................................................................................................. 79

REFERENCES ....................................................................................................................................... 85
LIST OF TABLES

Table 1-1 A Comparison of Hierarchical and Partnership Models .........................................................18

Table 2-1 Summary of the three phases of the evaluation of treatment integrity. ..................30

Table 2-2 Management plan for evaluation Question 1: What is the average sample of teachers’ adherence to behavior plans within the school? .................................................................32

Table 2-3 Management plan for evaluation Question 2: What is the average sample score of the behavioral plan quizzes taken by the teachers? .................................................................34

Table 2-4 Management plan for evaluation Question 3: Is there a relationship between treatment adherence percentages and behavior plan quiz scores? ........................................37

Table 2-5 Management plan for evaluation Question 4: What is the average sample of teacher’s level of engagement during recess? ..................................................................................38

Table 2-6 Plan for evaluation Question 5: Is there a significant difference in the teacher’s level of engagement when utilizing a covert versus overt observation approach ................42

Table 2-7 Management for evaluation Questions 6: Do teachers understand the importance of treatment integrity, and what are the barriers to treatment integrity implementation as perceived by teachers? .................................................................45

Table 3-1 Descriptive Statistics for Treatment Adherence Percentages of All Behavior Plan Observations .................................................................................................................................49

Table 3-2 Behavior Plan Quiz Average (N=24) .................................................................................. 50

Table 3-3 Correlations between Average Behavior Plan Quiz Scores and Average Treatment Adherence (N = 24) ................................................................................................................... 50

Table 3-4 Descriptive Statistics for Teacher Engagement Percentages during Recess ........53

Table 3-5 Mean Item Ratings for Barriers to Treatment Integrity Implementation Survey (N=24) .......................................................................................................................... 59

Table 3-6 Survey Results: Domain Averages ................................................................................. 60
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 3-1 School-wide Teacher Engagement Averages during Recess at Two Time Points</td>
<td>54</td>
</tr>
<tr>
<td>Figure 3-2 Last Day of Phase 1 (covert) Observations versus First Day of Intervention Phase 2 (overt) Observations</td>
<td>56</td>
</tr>
<tr>
<td>Figure 3-3 Comparison of Graphs With and Without Absent Staff Member Datum Points</td>
<td>57</td>
</tr>
</tbody>
</table>
CHAPTER 1

PROBLEM BACKGROUND AND LITERATURE REVIEW

Introduction

With the reauthorization of Individuals with Disabilities Act (IDEA) in 2004, investigative and experimental rigor for educational researchers became a mantra in education and intervention-based research. The demand for empirically supported interventions with high efficacy within school settings reached a pinnacle with the paradigm shift that promoted Response to Intervention (RTI). Yet, most outcome-based research reports measures of inter-observer reliability for the dependent variable while paying very little attention to the implementation of the independent variable as intended by researchers (Peterson, Horner, & Wonderlich, 1982; Reschly, 2004). Billingsley, White, and Munson (1980) examined 108 research articles published in the Journal of Applied Behavior Analysis (1977-1978) and Behavior Modification (1978-1979), the two premier journals in the field of Applied Behavior Analysis (ABA), and found that only 5.6% of studies assessed treatment integrity. Billingsley et al. (1980) highlighted the persistent absence of treatment integrity in research studies from the 1960s to the early 1980s. Peterson et al. (1982) confirmed this trend in their review of studies published in the Journal of Applied Behavior Analysis between 1968 and 1980; they found that only 16% of publications provided information on the integrity of implementing the independent variable. Gresham (1989) also found that most intervention-based research conducted prior to 1989 failed to mention, much less emphasize, the issue of treatment integrity. Moreover, Gresham noted that, by the end of the 1980s and in the early 1990s, behavioral researchers were starting to recognize treatment integrity as an essential but neglected variable in behavioral research and programming.
In the field of School Psychology, behaviorally oriented interventions have had increasing use within school settings culminating in the amendment of Individuals with Disabilities Act in 1997, where the amendment recommended a strong focus on behaviorally oriented interventions. Along with the increase in outcome-oriented behavioral interventions comes the concurrent need for integrity in their implementation, and the ongoing assessment of their utility and success (Sulzer-Azaroff & Mayer, 1991). The reviews by Billingsley et al. (1980), Peterson et al. (1982), and Gresham (1989) have brought to light the meager attention previously given to treatment integrity in the research and has resulted in a new era of research concentration on the topic within the field of School Psychology the need for program-specific recommendations in the area of treatment integrity based on a partnership model is still pressing. The purpose of this chapter is to review the literature from education, psychology and prevention sciences on the multidimensional construct of treatment integrity, and on the concept of program evaluation imbedded within the parameters of a partnership model. Treatment Integrity (TI) is defined as the degree to which the independent variable in an intervention is manipulated as intended (Gresham, 1997; Gresham, Gansle, & Noelle, 1993; Yeaton & Sechrest, 1981). Other terms for treatment integrity include treatment fidelity, procedural reliability, procedural integrity, and adherence to intervention (Moncher & Prinz, 1991). The terms treatment fidelity and treatment integrity are often used interchangeably. In this dissertation, the term “treatment integrity” will be used to indicate the manipulation of the independent variable and inter-observer reliability of those manipulations (Peterson et al., 1982).

**Treatment Integrity: A Multidimensional Construct**

Over the past few decades, researchers and practitioners within the field of School Psychology have found increasing agreement on the importance of addressing treatment
integrity (Hagermoser Sanetti & Kratochwill, 2009). The paradigm shift within the field is due to multiple influences including federal legislation (e.g. reauthorization of the Individuals with Disabilities Act of 2004, and No Child Left Behind), professional organizations’ position, and the inclusion of treatment integrity data requirements in large funding agencies (Hagermoser Sanetti & Kratochwill, 2009). Despite the developing consensus in the field on the importance of treatment integrity, the task of conceptualizing this complex and multidimensional topic remains a challenge.

**Defining the Construct of Treatment Integrity**

Despite the relative popularity of the term and concept of treatment integrity, researchers are not necessarily in agreement as to what the term denotes (Dane & Schneider, 1998). The inconsistencies in conceptualizing treatment integrity are due partially to the influence of education and the related fields (e.g., clinical psychology, prevention sciences) all of which contribute to the broad construct of treatment integrity. Treatment integrity is an issue in a diverse range of fields in which the overarching idea of integrity is similar. For example in the field of nutrition, dietary adherence represents the concept of integrity; in the field of prevention science, it is program fidelity, while within the field of applied behavior analysis it is known as procedural fidelity (Gresham, 2009). Despite commonality in the overarching idea of integrity, there are also differences in integrity between and within related fields. These differences are not limited to terminology; each field has a different model of treatment integrity, with inherent differences in service delivery models, settings, and intervention targets. Contextual inconsistencies of treatment integrity may also relate to research insularity within each field. Because of this insularity, the definition, approach, implementation and documentation of treatment may vary greatly between related fields although the concept of
treatment integrity may overlap (Jones, Clarke, & Power, 2008; Hagermoser Sanetti & Krachtowill, 2009).

Despite these differences among related fields regarding the vocabulary and implementation of treatment integrity, there has been some consensus across disciplines regarding its primary goals and constructs. The basic definition of “implementation of an intervention as planned” does not adequately cover the construct’s complexity. Instead, the conceptualization of treatment integrity encompasses three primary aspects:

(a) Therapist treatment adherence, the degree to which the therapist utilizes prescribed procedures and avoids proscribed procedures;

(b) Therapist competence, the level of the therapist’s skill and judgment; and

(c) Treatment differentiation, whether treatments differ from each other among critical dimensions (Perepletchikova, Treat, & Kazdin, 2007; pg. 829).

Because these three aspects are so important to understanding treatment integrity, I will discuss each one.

_Treatment adherence_ is often explained as the extent to which the therapist (or, interventionists) uses interventions and approaches prescribed by the treatment manual while avoiding intervention procedures that the manual proscribes. Treatment adherence can be assessed quantitatively (e.g., crucial treatment components can be measured as implemented or not implemented).

_Therapist (interventionist) competence_ refers to the level of skill shown by the therapist in delivering the treatment. Skill itself is further defined as the extent to which the interventionist takes the relevant aspects of the therapeutic context into account and responds to them appropriately (Waltz, Addis, Koerner, & Jacobson, 1993). The concept of therapist competence is further complicated by client variables including (a) the degree of client
impairment, (b) the set of specific problems manifested by the client, (c) the client’s life situation and life stress, and (d) the client’s stage in therapy, degree of improvement already achieved, and sensitivity to the timing of the intervention within the intervention session (Waltz et al., 1993). Because defining interventionist competence faces such contextual challenges, Waltz et al. (1993) recommended that interventionists’ overall competence must be assessed within a specific field and theory of practice (e.g., a Rogerian therapist and a behavioral interventionist have vastly different competence profiles), rather than by the individual’s experience level and training. Hence, the definition of competence shifts away from the general, focusing instead on the specific needs of the intervention and the knowledge, skill and behaviors required to perform it. For example, during an intervention whose goal is teaching a student independence from teacher prompts, it may be very important for the teacher to refrain from engaging actively with a student; however, active teacher engagement may be very important for a program promoting play during recess. A teacher with low student engagement ratings would receive high competency ratings for the first intervention, but not for the second intervention. It is therefore unwise to assume that any interventionist behaviors represent universal expressions of competence across interventions (Waltz et al., 1993).

_Treatment differentiation_, the third conceptual aspect of treatment integrity is defined as whether treatments differ from each other theoretically and along critical dimensions (Perepletchikova et al., 2007, Waltz et al., 1993). Three suggested components of treatment integrity overlap with other conceptual models of treatment integrity, showing distinct similarities. For example, the constructs of content (i.e., adherence, differentiation), quality (competence), quantity (adherence), and process (competence) suggested by Hagermoser Sanetti and Kratochwill (2009) in their conceptual model of treatment integrity resemble the constructs of adherence, differentiation and competence described by Waltz et al. (1993) in their
conceptualization of treatment integrity. Dane and Schneider (1998) propose another model of
treatment integrity that elaborates on the one posited by Waltz et al. (1993), adding the
dimensions of dosage and participant responsiveness. In their seminal article, Dane and
Schneider (1998) outline five variables that influence treatment integrity, including (a)
adoherence, the degree to which the therapist utilizes prescribed procedures and avoids
proscribed procedures; (b) quality of delivery, the quality which specific program objectives are
of the interventionist's delivery of the implemented program; (c) exposure or dosage, the
number, length and frequency of intervention sessions and implementation; (d) participant
responsiveness, the level of participants' engagement in the implemented intervention; and (e)
program differentiation, the extent to which program components are implemented and
extraneous components are excluded during implementation. Dane and Schneider also
suggested that each of these components addresses either a content (or, quantitative)
construct, or a process (or, qualitative) construct of treatment integrity. The quantitative
dimension refers to how often the intervention is implemented; the quality dimension speaks to
how well the intervention is implemented. Conceptualizing treatment integrity with the
dimensions proposed by Dane and Schneider is an important contribution to defining and
assessing treatment integrity (Durlack & DuPre, 2008).

Assessing Treatment Integrity

When intervention effectiveness is the basis of educational practices, it is critical for
those implementing the interventions to do so with high treatment integrity (Taylor & Gregory,
2007). Treatment integrity data are important in determining not only intervention efficacy, but
also for the observed outcomes which are central to the highly individualized interventions
promoted within an RTI model. If a primary intervention is implemented with a high degree of
fidelity, a client’s non-responsiveness will not be misattributed to low procedural integrity or to
inaccurate intervention practices; rather, the reason for unresponsiveness can be determined with accurate data, and students can be provided with either extra support and/or alternative interventions (Gresham, 1989; Lane, 2007). Within data-based decision making, using reliable and validated measures implemented with fidelity will also heighten the intervention’s internal and external validity (Gresham, Gansle & Noelle, 1993).

From clinical and educational perspectives, it is difficult to determine whether nonsignificant results are due to a poorly conceptualized intervention or to an inadequate or incomplete delivery of prescribed services without assessing treatment integrity (Gresham & Kendall, 1987). If an intervention does not yield significant results, its utility may be questioned and it may be discarded. “Throwing out” an intervention that may have worked if the delivery had been implemented with integrity is a disservice to the field. On the other hand, the lack of treatment integrity data may lead to replicating a specific intervention due to seemingly positive effects on behavior (dependent variable). The lack of treatment integrity data makes it virtually impossible to separate real effects from environmental factors, thus making experimental or casual replication in other educational settings questionable. The reason for monitoring treatment integrity is to ensure that changes to the dependent variable are attributable to the independent variable only; this is important not only in research but also in determining behavioral changes in students (Gresham, 1989; Peterson et al., 1982). While the literature documents the need for assessing treatment integrity, it is difficult to define and implement the methodology. In the following paragraphs, I explore the various methods often utilized to assess treatment integrity, and then review the relationship between the constructs that are assessed.
Methods of Assessment

Intervention plans delivered in many educational settings adhere to the consultation model with its complex triadic relationship between the consultant, the interventionist (teacher/instructor), and the student; interventions are rarely delivered by the people who write them (Noell & Gresham, 1993). The overall challenges in conceptualizing the dimensions of treatment integrity; and the methods of assessing treatment integrity levels within the field of school psychology and related fields are reviewed below, along the challenges associated with the assessment of treatment integrity.

Within the field of school psychology, treatments are commonly conceptualized as behavioral interventions; the conceptualized dimensions of treatment integrity proposed by Waltz et al. (1993) (e.g., adherence, differentiation, interventionist competence, exposure, and participant engagement) and Dane and Schneider (1998) are once again invoked in assessing treatment integrity rates. In the field of school psychology, exposure, or hours of intervention implementation, and adherence (extent to which the therapist (or, interventionists) uses interventions and approaches prescribed by the treatment manual are typically measured to assess treatment integrity (Schulte, Parker, & Easton, 2009). Exposure and adherence are both quantitative dimensions: exposure is measured as the amount of time implementing the intervention, and adherence is generally measured as the percentage of treatment steps completed within a prescribed treatment or intervention plan.

Discussing the challenges associated with assessing treatment integrity, Noelle and Gansle (2006) created the construct of Treatment Plan Implementation (TPI) as a measurement of treatment integrity within schools. TPI is defined as the degree to which an intervention plan is implemented in practice as it was designed, and is a measure of educators. Focusing on and assessing educators instead of student’s presents a conceptual and practical challenge to school
Within schools, it is mostly students who are assessed as sources of concern and targets for evaluation, rarely the educators. And, assessing TPI adds to the complexity of executing intervention plans. The provisions of prescribed intervention plans have a wide range, from a microscopic level of detail that is nearly impossible to assess, to a global summary of core elements with unspecified details and process. Research examining TPI and the levels of prescribed detail within the intervention plan recommends an intermediate level, with a focus on readily measurable critical steps that can detect both variations in implementation (treatment differentiation) and links between implementation (adherence) and outcome (Noell & Gansle, 2006). The methods most commonly used for evaluating treatment integrity in school-based interventions include (a) direct observation; (b) feedback from consultants; (c) self-monitoring, self reporting, and behavioral interviews; (d) permanent products; and (e) manualized treatments and intervention scripts (Lane & Beebe-Frankenberger, 2004; Lane, Beebe-Frankenberger, Lambros, & Pierson, 2001). I will discuss each method, including definitions, examples, and the unique challenges associated with each dimension of integrity assessment.

*Direct observation* is the most straightforward method of assessing treatment integrity and usually involves an individual observing the intervention while it is being implemented; the individual records whether or not each step of the intervention occurs in adherence to the intervention plan. There are four key steps to direct observation: (a) Creating a detailed treatment integrity protocol with a specific list (or, task analysis) of prescribed intervention components. (b) Operationalizing definitions of intervention components, with the definitions addressing verbal, physical, temporal and spatial parameters. For example, “student using appropriate greetings” might be defined as students making either a verbal greeting (e.g., Hello, How are you?) or a physical gesture (e.g., waving with hands) to indicate an appropriate
greeting; (c) Documenting, the presence or absence of each component of the treatment or intervention; and (d) Computing component and session integrity which is reported as a percentage. For example, if six out of ten components of a behavior plan are implemented in a session, the integrity rating is 60%.

*Direction observation* is the method used most often for assessing TPI in schools. Assessing TPI with observation alone requires substantial time, especially since an intervention is usually implemented a number of times throughout the day. Because this method is both time and labor intensive, assessment by direct observation will likely result in a small sample; there is concern too that the observed person’s reactions to the situation may contaminate the observation data. Another problem associated with direct observation is that the data may not be representative of the sample across setting and time (Hintze & Matthews, 2004). Finally, this method is not appropriate for all interventions, given its necessary components.

*Feedback from consultants* involves an individual, the consultant, who does not implement the intervention but either developed it or was trained in it, observing the teacher and providing corrective feedback. The consultative dyadic relationship created by the teacher and consultant constitutes both the challenge and benefit of this assessment method. The social support associated with the relationship has been shown to increase teachers’ levels of treatment integrity (Noelle, LaFleur, & Mortensen, 1997). However, there is an ambiguous relationship between the frequency of feedback from the consultant and the overall integrity of implementation in relation to treatment integrity rates is a somewhat ambiguous relationship. Wickstrom, Jones, LaFleur, and Witt (1998) found that teachers often did not implement plans as indicated when they lacked ongoing consultation support. Consultant feedback is also time consuming, and the efficacy of feedback sessions depends on relational variables such as the
strength of the bond between the consultant and the teacher, and the teacher’s acceptance of consultant efficacy (Noelle et al., 1997).

*Self-monitoring, self-report and behavior interview* constitute a three-part method of assessing treatment integrity. In self-monitoring a teacher tracks his or her adherence to each component of the intervention, typically involving the same task analysis approach used in direct observation. The teacher is then asked to self-report his or her data to a consultant or other administrator. In the behavioral interview, the teacher is asked whether each step of the intervention was implemented as originally intended, and the interviewer documents the answers. Despite the popularity of self-monitoring, self-reporting, and behavior interviews in practice, these techniques have inherent problems. The primary difficulty is teachers’ tendency to overestimate their level of integrity in implementing the intervention as designed as compared with the lower levels of integrity reported by direct observation (Wickstrom et al., 1998).

*Permanent products* entail collecting intervention-related artifacts (e.g., class work completed by the student, videotape of class performances, or reinforcement data sheets) and evaluating them. The primary benefit of permanent products lies within the minimal disruption to class and teacher time. The clear limitation of this method is that it can be used only with interventions that yield permanent products.

In the *manualized intervention* method, the intervention is described step-by-step in a manual, and the teacher is instructed to follow the manual exactly as written. This method assumes that the intervention is implemented with high integrity. However, as Lane and Beebe-Frankenberger (2004) noted, scripted interventions do not guarantee high integrity. It is
therefore best to utilize intervention scripts in conjunction with one or more of the previously
described assessment methods.

Noelle and Gansle (2006) suggest that using various assessment methods together may
provide the most practical and effective strategy for evaluating treatment integrity. For
example, combining observations, self-report, and manualized interventions may capture a
more comprehensive profile of treatment integrity then any one assessment can. To
complement this review of integrity assessment methods, I now review the different constructs
of treatment integrity and the relationships among them.

**Relationship between Treatment Adherence, Treatment Differentiation and Interventionist
Competence**

The constructs of treatment adherence, treatment differentiation and interventionist
competence proposed by Waltz et al. (1993) are synonymous conceptualizations of treatment
integrity. The relationship between treatment adherence and treatment differentiation is
strong and lies in the intrinsic nature of assessing treatment adherence. For example, when an
interventionist’s adherence to prescribed intervention directions is evaluated with manipulation
checks, the distinction between interventions is also assessed de facto: as adherence to
prescribed implementation steps for intervention A is evaluated, avoidance of steps prescribed
for intervention B and other proscribed procedures is also evaluated (Perepletchikova et al,
2007). Since treatment adherence and treatment differentiation are so inter-related, “an
adherence measure is sufficient” when assessing for the differentiation of treatments (Waltz et
al., 1993, p. 625).

The relationship between treatment adherence and interventionist competence has been
researched extensively but results are inconclusive. McGlinchey and Dobson (2003) conducted
an empirical study of the relationship between treatment adherence and interventionist
competence and discussed the presumed relationship between these two constructs. The assumption is that an interventionist cannot achieve competence within a particular intervention without first adhering to the intervention procedures, and vice versa. However, while competence presumes adherence, adherence does not necessarily presume competence. An interventionist can perform prescribed behaviors exactingly, but not well (McGlinchey & Dobson, 2003). There is a conceptual distinction between adherence and competence related to measurement: adherence is defined as a quantitative construct of treatment integrity (i.e., how frequently the interventionist implements procedures that are prescribed by the manual, and avoids those that are proscribed) while competency is defined as a qualitative construct (i.e., how well prescribed procedures are implemented). Despite this distinction, the two constructs overlap significantly. This overlap presents troublesome challenges, especially within scripted measurements of treatment integrity. For example, consider an intervention session with the goals of having a student read from challenging text and encouraging attempts at new vocabulary. Teacher A asks a student to read a passage from a book; when the student responds ”I can’t read it”, Teacher A replies “Just try”, but relents when the student refuses. Teacher A followed the prescribed script for the intervention, and obtained a full score for adherence. However, should the teacher have encouraged the student again with different strategies? The answer is, “It depends” on context, and the student’s mannerisms and affect. The assessment dilemma presented here encapsulates the complexities caused by the conceptual overlap of treatment adherence and interventionist competence. The indistinct relationship between integrity and competence extends also to treatment integrity and treatment outcomes.
**Relationship between Treatment Integrity to Treatment Outcome**

**Treatment Outcome**

The relationship between treatment integrity and treatment outcomes is ambiguous at best. Although it is tempting to hope for a direct relationship between increased integrity in implementation and better client outcomes, the literature reveals conflicting conclusions regarding integrity and outcomes. The possibility of finding a definitive relationship between treatment integrity and outcomes is also clouded by the variable of adaptation to environment or client. For example, Durlack and DuPre (2008) confirmed the hypothesized relationship between implementation fidelity and program outcomes, finding strong support for the premise that effective implementation is associated with better outcomes. Citing the results of 500 studies evaluated through five meta-analyses, they maintained that extensive and persuasive evidence confirms the powerful influence of implementation on outcomes. However, Sterling-Turner, Watson, and Moore’s (2002) hypothesis that the degree of students’ behavior change would be associated with consultees’ treatment integrity was only partially supported; the literature has not demonstrated conclusively that total treatment integrity is necessary to promote behavioral change. But, if treatment integrity is less than 100%, by definition, an implied variable of treatment adaptation is introduced. The old debate between the necessities for treatment integrity versus adaptation of treatment rests on the fact that 100% integrity is not needed for behavioral change.

**Integrity versus Adaptation**

There is an active debate on the feasibility and need for treatment integrity as it pertains to treatment outcomes in the disciplines related to School Psychology, especially in the field of prevention research. The primary question “Is treatment integrity a necessity in
outcomes?” is a broad one, and must be addressed by exploring the integrity versus adaptation debate.

In Backer’s (2002) literature review of prevention programs, program adaptation was defined as the deliberate or accidental modification of the program, including (a) deletions or additions of program components; (b) modifications in the nature of the components that are included; (c) changes in the manner or intensity of administering program components called for in the program manual, curriculum, or core component analysis; and (d) cultural and other modifications required by local circumstances. The debate over integrity versus adaptation of systems-level intervention programs highlights these questions: Do changes in a science-based program dilute and even dissipate effectiveness? Does the inability to modify programs produce local resistance? Does rigid fidelity lead to programs that are irrelevant and inappropriate for meeting the program’s or client-specific needs?

This debate, including the proposed need for achieving a balance between fidelity and adaptation, has been an issue since the 1960s, when applied research in education and related fields became widespread, and the field of program evaluation was born. Program evaluation featured an increased effort to explain why evaluation results of previously validated programs were null or inconsistent across settings, thus raising the possibility of an effect from a program’s implementation rather than from the program itself. Over the past 40 years, research has provided a better understanding of the complexities of implementation fidelity; out of the field of program evaluation has come an understanding of the “top-down” approach to intervention and program implementation, and the “bottom-up” approach of adaptation. An intervention or program is often selected by a consultant or governing agency, but it is the practitioners who apply the components of the program or intervention who often make the adaptations. In the literature review, Backer (2002) concludes that attention to both fidelity and
adaptation is essential for successful implementation. The concept of fidelity versus adaptation is antiquated. It is currently held that fidelity and adaptation can co-exist, with the assumption that adaptation of programs or interventions will naturally occur with setting and interventionist changes. The key for optimal outcomes is achieving a balance between the degree of adaptation and the integrity of core components. While Backer (2002) focused on a single system wide program being implemented across settings by multiple interventionists; the fundamental concept can be extended to a school setting, in which multiple teachers may be systematically given the same intervention to be implemented with a specific student, and each teacher finds his or her adaptation of the scripted intervention. There is a dynamic and complex interaction between the program/intervention and the environment/student where the intervention is enacted; it is important to attend both to fidelity and the adaptation that occurs based on the unique needs of the environment or student (Backer, 2002). Next, I assess program-specific constructs related to treatment integrity, including the role of stakeholders, in order to explore the constructs of treatment integrity and adaptation of a particular program, institution, or school.

**Program Evaluation**

The justification for conducting a program evaluation stems from the complexity of the construct of treatment integrity, and from the importance of exploring treatment integrity within the environment where the intervention is being implemented (Backer, 2002; Gresham, 1997). Researchers have provided detailed accounts of the barriers to treatment integrity (e.g., McIntyre, Gresham, DiGennaro, & Reed, 2007), but they do not provide school administrators and teachers with detailed feedback about treatment integrity rates, perceptions, strengths, and weaknesses specific to their environment. The importance of exploring the multiple aspects of treatment integrity and adaptation within the host environment points to the pressing need
for a critical and comprehensive review of treatment integrity utilizing a partnership model. In a participatory research (i.e., partnership) model, the primary stakeholders (e.g., school administrators and teachers) partner with the researcher in determining the school’s primary needs and goals.

**Participatory Intervention Model**

While offering individualized programmatic input is difficult, the complexity of treatment integrity makes individualized evaluation and input essential. The concept of making research responsive to practice is based on the overarching theory of participatory research. At the core of a participatory intervention model is the active inclusion of all major stakeholder groups at each step of the research process (Nastasi, Varjas, Schensul, Silva, Schensul, & Ratnayake, 2000). The Participatory Intervention Model (PIM) integrates theory and research in developing context-specific interventions meant to promote ownership among stakeholders who are responsible for the long term use of the interventions after consultation ceases (Nastasi et al., 2000). Interventions created within the PIM model are culturally specific, reflecting the language, vocabulary, values and beliefs of the stakeholders who utilize the interventions. Other components of a PIM include the involvement and ownership of interventions by key stakeholders, and the fostering of sustainable and system-specific change. In order to achieve a PIM, researchers must devote time and energy to the development of partnerships, engage in formative research and conduct a comprehensive evaluation of the school (Nastasi et al., 2000).

Power, Blom-Hoffman, Clarke, Riley-Tillman, Kellerher, and Manz (2005) expanded upon the conceptualization of PIM, linking it with treatment integrity monitoring, and presenting the partnership framework as a method of integrity management. The prevailing model of integrity monitoring is based on the hierarchical model with a top-down framework, in which researchers control the creation of interventions, and treatment integrity rates are
monitored based on predetermined efficacy rates. The hierarchical model does not include the perceptions and beliefs of service providers or school based administrators in intervention planning or treatment integrity monitoring phases. This model fails to take school-specific needs into account, and fails to produce the opportunity for collaboration with site based administrators and staff. Based on a hierarchical model, treatment integrity is assessed in a summative manner that lacks the collaboration and input from primary stakeholders (Powers et al., 2005). A hierarchical and summative approach to monitoring treatment integrity limits opportunities for interventionists to review integrity data as it is being collected, and monitor for the individualized intervention needs of the school.

Alternatively, the partnership framework, with a strong foundation as an extension of PIM, acknowledges the dichotomous expertise of the researcher and the primary stakeholders within their respective domains. The researcher brings an understanding of evaluation and development, whereas the stakeholder understands the culture, history and specific needs of the school. Hence, the integration of evidence-based practice along with identified needs and priorities of stakeholders merge to form a partnership. Within a partnership model, treatment integrity is not only monitored by the researchers utilizing monomethod assessments, instead, researchers and school based stakeholders determine what gets monitored with multiple methods of monitoring treatment integrity (Powers et al., 2005).

Table 1-1. A Comparison of Hierarchical and Partnership Models (from Powers et al., 2005)

<table>
<thead>
<tr>
<th></th>
<th>Hierarchical model</th>
<th>Partnership model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Who develops the</td>
<td>Research team</td>
<td>Research and intervention teams with other stakeholders</td>
</tr>
<tr>
<td>intervention?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>What is the basis for</td>
<td>Evidence-based</td>
<td>Evidence-based practices and experiences/needs of</td>
</tr>
<tr>
<td>intervention development?</td>
<td>practice</td>
<td>stakeholders</td>
</tr>
<tr>
<td>Who conducts integrity</td>
<td>Research team</td>
<td>Multiple informants using</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
monitoring?  multiple methods  
What domains of integrity are assessed? Exposure, adherence, differentiation  Emphasis upon quality of intervention and participant responsiveness  
What is actually monitored? Highly specific intervention steps Critical components of intervention and intervention drift  
How are integrity data used? Summative evaluation Emphasis on formative evaluation  

Table 1-1 summarizes the differences between the hierarchical and partnership model. Within a PIM model and a partnership framework, the school or host environment dictates the goals of the program evaluation in cooperation with researchers, thus providing an overview of treatment integrity that is both needs driven and research supported.

**Defining Program Evaluation**

In order to fully define program evaluation, one must consider what a program is. Within a business model, programs are identified as overall goals that must be reached in order to accomplish a mission. In nonprofit or educational organizations, each goal often becomes a program intended to provide services or to improve a component of the overall function of a service (Fitzpatrick, Sanders, & Worthen, 2004). Program evaluation is the careful collection of data about a program, or various aspects of a program, based on previously determined evaluation questions in order to make necessary decisions about the program. Fitzpatrick et al. (2004) formalized the definition of program evaluation as “the clarification and application of defensible criteria to determine an evaluation object value (worth or merit) in relation to those criteria” (p. 7).

Program evaluations may be narrow or broad depending on the need of the evaluation. There are many reasons for utilizing an evaluation, including needs assessments, accreditation, cost/benefit analysis, and increasing effectiveness and efficiency. There are also many types of
evaluations, including, but not limited to, formative, summative, goal-based, process-oriented, and outcome-oriented. An evaluation of an organization is a tool for administrators to determine the extent to which a program or curriculum is effective; at the same time, it is a tool for administrators to remediate programs that are not effective and that do not contribute to the overall achievement of organizational goals.

Within an organization, program evaluation serves two primary functions, confirmation and diagnosis. A program evaluation can also guide training and professional development within an organization; once problems are identified, staff training and development programs can be designed or revised in an attempt to remediate the problems.

There are many definitions of “program evaluation” within the literature. Brainard (1996) described effective program evaluation as a “systematic process that focuses on program improvement and renewal and on discovering peaks of program excellence” (p. 5), while Posavac and Carey (2007) asserted that the goal of evaluation is “providing feedback from program activities and outcomes to those who can make changes in programs...” (p. 14). As the literature indicates, the primary goals of program evaluations are to utilize the data collected to structure feedback to administrators in an effort to produce data-based, sustainable, and systems-level changes.

**Formative Program Evaluation**

The type of program evaluation must address the overarching purpose of the assessment. There are two types of evaluation: formative and summative (Fitzpatrick et al., 2004). The purpose of a formative evaluation is to produce ongoing feedback and findings, conclusions and recommendations that will help to improve and refine the program, whereas the purpose of a summative evaluation is to utilize data to “summarize” the influence of the program. Summative evaluations focus on outcome data on the trends within a program and
include final determinations on the program’s efficacy by comparing trends in the observed data against criteria established to evaluate overall performance. Decisions made as a result of summative evaluations may include the adoption, discontinuation, and/or expansion of a program.

The form of evaluation used by an organization at a given time depends on the needs of the administrators and the goals of the organization. Over the life of an established program, there should be a balance of both formative and summative evaluations, but when a program is new or is adapting to its unique environment, a formative (ongoing) evaluation is preferable (Fitzpatrick et al., 2004). In this dissertation I describe a formative program evaluation, which was undertaken to address the administrators’ need for concrete and detailed feedback in a PIM model.

Formative evaluation allows exploring treatment integrity as a focal point within a program, and utilizing multiple methods to gather data on this complex construct. The evaluation provides data that will guide decisions about the future direction of the program, and provide baseline data of various conceptualizations of treatment integrity (Posavac & Carey, 2007). The systematic and comprehensive methods of formative program evaluation are in line with reflective practice and the integration of primary stakeholders into the assessment process as the partnership framework suggests. Formative program evaluations benefit administrators, and they also create teacher “buy-in” since teachers’ opinions regarding treatment integrity are solicited, and their ideas valued as part of the overall picture of treatment integrity within the school. This in turn may lead to team-building and cohesion among the staff (Fitzpatrick et al., 2004).

The program evaluation described in this study occurred in a specialized school setting for children who are diagnosed with Autism Spectrum Disorder (ASD). The school applies both
behavioral and educational interventions based on the principles of Applied Behavior Analysis (ABA). Since an understanding of ABA is essential to understanding the nature of interventions applied in the school and evaluated for treatment integrity in this dissertation, a review of the science of ABA follows.

**Applied Behavior Analysis**

Behaviorism is the overarching philosophy of the science of behavior and behavior change. One field that emerged from the philosophy of behaviorism became the contemporary Applied Behavior Analysis (ABA). ABA is defined as the scientific approach for discovering environmental variables that reliably influence socially significant behavior and for developing a technology of behavior change that takes practical advantage of those discoveries (Baer et al., 1968). ABA-based interventions have dramatically increased in popularity over the past half-century. Their popularity may be linked, in part, to the rise in Autism Spectrum Disorders (ASD) in the United States, since ABA has long been touted as a viable form of behavioral modification for individuals diagnosed with ASD. The shift in diagnostic criteria from the 1950s to the most recent revision of the DSM-IV-TR (American Psychiatric Association, 2000) has changed the incidence of autism: researchers have noticed an alarming increase in the number of clinical diagnoses of ASD (Wing & Potter, 2002). For the past century, autism has been considered a relatively rare disorder, with a low prevalence rate of 4 out of 10,000 children. According to recent statistics from The Center for Disease Control and Prevention (2012), there is an average diagnostic rate of 1 out of 110 children in the United States, thus making autism a widely recognized diagnosis in both clinical and popular culture. Along with the dramatic increase in ASD diagnoses and the related expansion of ABA, the principles and broad practice of ABA have become more prevalent within general education settings. Interventions based on behavioral
Modification strategies have found strong support because of changing special education legislation and a paradigm shift in the overall practice of school psychology (Reschly, 2004).

Within a three-tiered Response to Intervention (RTI) model, ABA falls into Tier 3, or, intensive interventions with individual students being served in a progress monitored continuum of support and providing procedures for lasting generalization of skills (Batsche, 2005). The characteristics of teaching from an ABA approach include: individualized instruction if administered to a group, consistent monitoring of student responses, data collection on the efficacy of the intervention, as well as utilizing the principles of ABA to teach educational and social skills (e.g., using positive reinforcement when a student answers a question correctly or ignoring an undesirable behavior in order to discourage it).

**Applied Behavior Analysis – Teaching Principles**

In the ABA model, learning is closely related to change: change in environmental stimulus, change in behavior, increase in adaptive behavior, and decrease in maladaptive behavior. This idea of learning is based on complex interactions between an organism and his or her environment (Baer et al., 1968). Thus, within a behavioral framework, the manipulation of antecedents and consequences of behavior is meant to teach new skills and eliminate maladaptive and excessive behaviors. An example of an ABA teaching strategy is the use of discrete trials. Discrete trials enable the learner to acquire complex skills and behaviors by first mastering the subcomponents of the targeted skill. Subcomponents are mastered gradually and are linked together to enable mastery of the more complex targeted and functional skill (Cooper, Heron, & Heward, 2007).

Modeling is another example of systematic teaching. With modeling, a new behavior is acquired through demonstration and through shaping strategies that differentially reinforce successive approximation of a target behavior, gradually building up to a more complex
behavior by teaching one skill at a time. The combination of systematic teaching strategies and functional relationships through the manipulation of environmental stimuli provide the foundation for most, if not all, behaviorally-based intervention plans (Cooper et. al., 2007).

The following principles of reinforcement describe a functional relationship between environmental events and resulting behavior:

- **Positive Reinforcement** – A functional relationship between two environmental events: a behavior (any observable action) and a consequence (a result of that action). Positive reinforcement is demonstrated when a behavior is followed by a consequence that increases the behavior’s rate of occurrence.

- **Negative Reinforcement** – A relationship among events in which the rate of a behavior’s occurrence increases when some (usually aversive or unpleasant) environmental condition is removed or reduced in intensity.

- **Punishment** – A relationship in which a behavior is followed by a consequence that decreases the behavior’s rate of occurrence. Distinct from the common use of the word “punishment,” in ABA terminology, an event is a punishment only if the rate of occurrence of the identified behavior decreases.

Along with these reinforcement principles, the concept of the function of behavior, (that is, why behavior occurs) is also a cornerstone to understanding client behavior. Within ABA, there are five primary motivations for students’ behaviors: (1) to gain attention from a peer or an adult; (2) to gain a tangible or other reinforcement; (3) to gain sensory stimulation whether visual, auditory, or olfactory; (4) to escape from tasks or other demanding interactions; and (5) to escape from painful or discomforting internal stimulations (Cooper et. al., 2007).
Thus, interventions based on ABA principles require reinforcement, systematic teaching strategies, and the assessment of the function of relationships. The innate complexity of these interventions aligns with the conceptual complexities of treatment integrity. The best method for evaluation, therefore, is a multi-pronged method that can fully explore treatment integrity.

**Purpose of the Dissertation**

A review of the literature makes it clear that treatment integrity is a complex construct. Realizing this complexity, evaluating a specific program or school-based intervention implemented within a partnership framework is essential to providing a comprehensive baseline of treatment integrity rates, and program-specific recommendations to teachers and administrators. A formative program evaluation provided a thorough evaluation of environment-specific constructs of treatment integrity, and the relationships among the constructs of treatment integrity. The nature of a specialized school founded on ABA principles also provides a rationale for program evaluation. The school day is founded on scripted interventions, with every minute of the day, excluding recess, based on a scripted behavioral or academic intervention. The integrity associated with implementation of these interventions is highly regarded due to the structured and intervention-driven nature of the school. Thus, a school founded on behavioral and academic interventions offers a unique opportunity to assess treatment integrity due to the sheer plethora of interventions that are being monitored for efficacy on a daily and hourly basis.

This evaluation assessed treatment integrity through a mixture of quantitative and qualitative methods. The methods included direct observation, permanent products, surveys, and focus groups. Through the evaluation, I sought to understand staff perceptions of treatment integrity (e.g., the barriers, the need and the conceptualization of treatment integrity), along with a sampling of various treatment integrity rates within the school.
The multiple assessment and observation methods used in an exploratory case study design highlighted described by Fitzpatrick et al. (2004) combine quantitative and qualitative data to provide a thorough understanding of the multiple aspects of treatment integrity within the school. The methods of assessment and observation used in this study are structured around the following evaluation questions.

1. What is the average sample of teacher’s adherence to behavior plans within the school?
2. What is the average sample score of the behavioral plan quizzes taken by the teachers?
3. Is there a relationship between treatment adherence percentages and behavior plan quiz scores?
4. What is the average sample of teacher’s level of engagement during recess?
5. Is there a significant difference in the teacher’s level of engagement when utilizing a covert versus overt observation approach?
6. Do teachers understand the importance of treatment integrity, and what are the barriers to treatment integrity implementation as perceived by teachers?
CHAPTER 2

METHODS

The previous chapter conceptualized the multiple constructs of treatment integrity through a review of school psychology, education, prevention and clinical literature. With the inherent complexities of treatment integrity and the importance of a collaborating with site based stakeholders when conducting field research, a program evaluation of treatment integrity within the school based on a partnership model was proposed. The evaluation questions include (1) What is the average sample of teacher’s adherence to behavior plans within the school? (2) What is the average sample score of the behavioral plan quizzes taken by the teachers? (3) Is there a relationship between treatment adherence percentages and behavior plan quiz scores? (4) What is the average sample of teacher’s level of engagement during recess? (5) Is there a significant difference in the teacher’s level of engagement when utilizing a covert versus overt observation approach? (6) Do teachers understand the importance of treatment integrity, and what are the barriers to treatment integrity implementation as perceived by teachers? Each evaluation question is presented in conjunction with a management plan, a description of evaluation measures and methodology, and data analytic methods.

Setting and Participants

The study was conducted in a specialized school setting for children and adolescents diagnosed with Autism Spectrum Disorder (ASD). The school, which will be referred to as the Behavioral Day School, is one of four children’s day schools belonging to one of the largest nonprofit agencies in the country that operates behaviorally oriented private schools. These schools are designed specifically to serve children with autism through academic and behavioral programs merging the techniques of Applied Behavior Analysis (ABA) and special education.
They serve children ranging in age from 3-18 years; children are usually referred to the school by school districts, advocacy groups or parent advocates. The reasons for referral vary but tend to cluster around the severity of the child’s behavioral problems, the inability of the home district to serve the needs of the child, and the school’s high reputation as one of the best resources in the state for children with behavioral disabilities, based on its observed results. The school where the research was conducted is in session for 222 days of the school year, broken down into a twelve-month, five-day-per-week year model.

There are four classrooms, with students grouped by age; each classroom consists of 5 or 6 students. During instruction time the teacher-to-student ratio in all classrooms is one-on-one; during recess time there is a one-on-two staff-to-student ratio. The teachers range in age from early 20s to mid-40s all teachers have at least a bachelor’s degree in education, psychology or a related field, and most either have a graduate degree or are currently enrolled in graduate programs in ABA. There are a total of 24 teachers in the school, and 23 students. The extra teacher would fill in if there was an absence; otherwise, she would aide with curriculum development. Each teacher is assigned to a specific classroom, and works only with the children in that room. The director of the school is a Ph. D.-level School Psychologist and a Board Certified Behavior Analyst; the education coordinator is a master’s level Board Certified Behavior Analyst.

The lead evaluator is a school psychology graduate student who held the position of research intern within the school. Due to this position, the evaluation was considered an internal program evaluation. However, at the time of the evaluation, the lead evaluator no longer worked directly with the students nor had any leverage or consequential power over the staff. The evaluator reported results directly to the school director. The evaluator made a point to ensure that staff members were not coerced to participate in the study, and that there were
no consequences for non-compliance with treatment plans, due to aggregate data reporting, in an effort to prevent bias. The evaluation was approved by the University of Massachusetts Amherst Institutional Review Board (IRB). The following procedure was used to obtain informed consent.

**Teaching Staff.** During a prescheduled morning staff meeting the evaluator presented a consent form to every teacher and explained the details of the form and the evaluation. The evaluator informed the teachers that participation was voluntary with no monetary or other incentives, and that there were no consequences of any kind should the teachers decide not to participate. The evaluator was available during and after the morning meeting to answer questions regarding the form and the teachers were allotted generous amounts of time to read it and ponder the decision to participate in the study. Teachers turned in the consent forms once they had made their decisions. All 24 staff members and 2 administrators agreed to participate. During the middle of the evaluation year, a new staff member joined the school; however, the data of the new staff member was not included in the evaluation project.

**Students.** The study evaluated teacher behavior, with no attempts to manipulate student programming. Students were a part of the observation periods, but all observations focused on staff behavior; students’ behavior was not recorded or analyzed. The nonprofit organization has a previously established policy of asking all parents to consent to their children’s participation in research prior to enrollment in each of the specialized day schools. The school consent forms allow for the observation of students during school hours, access to student records provided data is presented in the aggregate, and changes to a student’s behavioral or academic programming as supervised by the administrative staff. Since this study focused only on staff behavior and a school consent form for research participation already existed, the University of Massachusetts Institutional Review Board determined that the
The goal of this evaluation study is to assess treatment integrity within a participatory research framework, integrating the needs of administrators through a formative program evaluation. The evaluation was broken down into three primary phases (as summarized in Table 2-1).

**Table 2-1. Summary of the three phases of the evaluation of treatment integrity.**

| Phase 1 | Determination of assessment and data collection methods.  
Collection of data using established methods. |
|---------|---------------------------------------------------------|
| Phase 2 | Sharing Phase 1 data with administrators in a feedback meeting.  
Determining the need for additional data collection or interventions to address treatment integrity weaknesses. |
| Phase 3 | Follow up meeting to discuss interventions applied as a result of Phase 2 feedback meetings.  
Continued data collection of treatment adherence levels.  
Meetings to plan future direction of program evaluation. |

Phase 1 included determining the assessment/data collection methods, and collecting data based on methods agreed upon by both the lead evaluator and school administrators. In this first phase, the data collection was based on evaluation questions which were developed with the school administrators. Preliminary data meant to assess the evaluation questions were collected through a mixture of quantitative and qualitative methods, including surveys, systematic direct observations, focus groups, and observations. Phase 2 consisted of sharing the data collected in Phase 1 with administrators through a feedback meeting. The goal of the feedback meeting was to develop interventions to address areas of treatment integrity related to hypothesized weaknesses in the school, and to determine if additional data collection was require to follow up on the evaluation questions presented in Phase 1. Only one intervention was proposed during the feedback meeting in Phase 2, and this intervention was implemented
during the latter part of Phase 2. In addition to the intervention, it was decided that additional datum points observing treatment adherence to behavior plans would be collected. Phase 3 served as the follow up phase on the additional data that was collected and the intervention that was implemented as a result of Phase 2 feedback meetings. Phase 3 included a future directions component, wherein trainings on treatment integrity and data collection methods for the upcoming year were proposed to administrators.

As Scriven (1980) suggests, it is important in planning an evaluation to evaluate the “evaluand” (i.e., the goal or purpose). This planning phase helps to determine the kind of information that should be collected and the method of analysis most suited to the data that will support the evaluation questions. Accordingly, in Phase 1 of the program evaluation we strove to understand the overarching goals of the institution. These goals dictated that the study should have an exploratory single study design: decisions were to be made through the collection of both quantitative and qualitative data in an effort to determine the value or quality of the practices of the school in relation to treatment integrity (Fitzpatrick et al, 2004). The reason for using this design is that the single case exploratory design as a formative program evaluation method goes directly to the core of the research questions in ascertaining “what is working, what needs to be improved and how it can be improved” (Fitzpatrick et al, 2004).

**Evaluation Plan**

Based on the proposed evaluation questions, the school administrators and the lead evaluator created management plans to guide the program evaluation. Data reporting procedures that follow outline clearly the type of data required for each question and the method of reporting data to school administrators. Prior to the start of this evaluation, the evaluator met with the director and educational coordinator to discuss the evaluation needs of the school with regard to treatment integrity. Utilizing a partnership model, in which school
administrators (or primary stakeholders) have a say in the areas of concern, the questions and evaluation plan were agreed upon by both the evaluator and school administrators.

**Question 1.**

The management plan for the first evaluation question, “What is the average sample of teachers’ adherence to behavior plans within the school?” is presented in Table 2-2 below with details regarding the data collection instruments, and the methods and techniques used to analyze and interpret the data gathered.

**Table 2-2. Management plan for evaluation Question 1: What is the average sample of teachers’ adherence to behavior plans within the school?**

<table>
<thead>
<tr>
<th>Information required</th>
<th>Information source</th>
<th>Method for collecting information</th>
<th>Information collection arrangements</th>
</tr>
</thead>
<tbody>
<tr>
<td>School-wide average of teachers’ adherence to prescribed behavior plans.</td>
<td>Observation of teachers during class time</td>
<td>Direct observation: Six 1-hour observations per teacher in each of data collection periods</td>
<td>Whom: Evaluator and trained undergraduate data collectors. Inter-observer agreement calculated</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Analysis procedures</th>
<th>Interpretation procedures and criteria</th>
<th>Report of information</th>
</tr>
</thead>
</table>
| Descriptive statistics for calculating treatment adherence percentages | Indication of teacher’s adherence to behavior plan within a classroom environment | Whom: School administrators
How: Feedback meetings
When: Phase 2 and Phase 3 |

In order to address this evaluation question, the teachers were to be observed. In this context, it is important to describe the school’s system of student behavioral plans. In the Behavioral Day School where the evaluation took place, every student has a distinct and highly individualized behavior plan which outlines every aspect of a student’s school day, from toileting schedules in the morning, to receptive language programs in the afternoon, to scripted behaviors from the
teacher in reinforcing the student for saying goodbye at the end of the day. In these
Individualized Education Plans (IEPs) are imbedded academic goals oriented towards the
student’s annual behavioral goals; not a moment of a student’s day is left unscripted. The IEP is
an individualized document, created by a team usually consisting of parents, teachers, and
school administrators who determine the plan to address a student’s unique educational and
behavioral needs (IDEA, 2004). IEPs are reviewed annually to assess goal obtainment. While the
content of behavior plans is customized to the behavioral and academic needs of each student,
the basic structure of the plan follows a school-wide rubric. Since the plans are also written by
the same people (i.e., the school director who serves as the school’s clinician and the
educational coordinator), the behavior plans within the school show consistency in their
authorship, the language utilized, and the template used.

Every teacher is responsible for the behavior plans of the students in his or her class;
with classroom sizes ranging from 5 to 6 students, teachers are expected to know only 5 to 6
plans. For the evaluation study, two different behavior plans were selected randomly for
observation within each classroom. Each teacher in the school was observed six distinct times,
three observations per behavior plan with each observation lasting for one hour. The randomly
selected behavior plans, totaling eight behavior plans across the school, were broken down into
a treatment adherence checklist. Each distinct step of the plan (e.g., reinforce student for
successfully writing their name, or prompt student to return greetings) was assigned a value of 1
point. The occurrence or nonoccurrence of a teacher’s adherence to the plan was marked with
a 1 (for occurrence) or a 0 (for non occurrence). The number of observed occurrences,
signifying adherence to the behavior plan, was summed up and divided by the total number of
possible points on the behavior plan, yielding a treatment adherence percentage for an
individual observation session. During Phase 1 of the evaluation, 3 distinct 1 hour observations
were conducted per teacher, each teacher was observed with the first set of 4 behavior plans (1 behavior plan per classroom), yielding a total of 72 observations for Phase 1. The 72 observations provided a school wide average of treatment adherence to behavior plans.

Data was shared with administrators in the feedback meeting in Phase 2. During Phase 3, the second set of 4 behavior plans (1 per classroom) was utilized in a second round of observations, in which each teacher was once again observed for 3 distinct sessions, yielding a total of 72 observation sessions. The observations started in August of 2009 and ended in June of 2010.

This method of direct observation was repeated 144 times, with 30% of observations being conducted by both the principal evaluator and trained undergraduates or educational coordinator to obtain Inter-Observer Agreement (IOA) data. Interobserver agreement was calculated using overall agreement calculations and the kappa statistic ($\kappa$). Overall agreement was calculated by dividing the total number of agreed observations by the total number of agreements and disagreements (Gay & Airasian, 2003). The definition of Kappa, is that Kappa accounts for chance agreement, and is based on the difference between actual agreement and agreement due to chance (Kazdin, 1982; Viera & Garrett, 2005). According to Viera and Garrett (2005), values between .21 and .41 have fair agreement, values between .41 and .60 have moderate agreement, values between .61 and .80 have substantial agreement, and values between .81-.99 have almost perfect agreement. Interobserver agreement ranged from .7 to 1.

The average treatment integrity adherence percentage that was computed from these data (based on 24 teachers with 6 observations per teacher, and a total of 144 hours of observations) providing a school wide average of treatment adherence. Descriptive statistical measures were utilized to further analyze school-wide treatment adherence.

**Question 2**
The management plan for the second evaluation question, “What is the average sample of score of the behavioral plan quizzes taken by the teachers?” is presented in Table 2-3 with details regarding the data collection instruments, and the methods and techniques used to analyze and interpret the data gathered.

Table 2-3. Management plan for evaluation Question 2: What is the average sample score of the behavioral plan quizzes taken by the teachers?

<table>
<thead>
<tr>
<th>Information Required</th>
<th>Information Source</th>
<th>Method for Collecting Information</th>
<th>Information Collection Arrangements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher performance on behavior plan quizzes.</td>
<td>Permanent products, teacher percentages on monthly behavior plan quizzes.</td>
<td>Behavior plan quizzes based on prescribed behavior plans.</td>
<td>Whom – The evaluator will collect the quizzes. When-Phase 1 of the evaluation</td>
</tr>
</tbody>
</table>

Analysis Procedures | Interpretation Procedures and Criteria | Report of Information |
---|---|---|
Descriptive statistics. | Indication of teacher’s knowledge of behavior plans. | Whom- School administrators How-Feedback Meeting When-Phase 3 of the evaluation |

Behavior plan quizzes are an established method within the school to assess teacher knowledge of the behavior plans of the students in their classrooms. In each of the four classrooms, every teacher works with a specified student for 1 hour of the day, the teachers are on a rotating schedule, thus if there are 5 students and five teachers within a classroom, the teachers will rotate every hour so that each teachers will work with every student during a 5 hour period. Therefore, it is school procedure for teachers to know the behavior plans of every student within the classroom the teacher is assigned to. The behavior plans, as explained previously provide a script for the students’ day, with multiple steps and complicated academic and behavioral components. Behavior plans are routinely updated; the changes in behavior plans are dependent on the student’s academic and behavioral progress. During the pre evaluation meeting with administrators, it was revealed that teacher’s knowledge of behavior
plans is important, if teachers do not know the steps of the behavior plan, then the concept of adherence to the behavior plan cannot be broached. As a means of ensuring teacher knowledge of behavior plans, the administrators made behavior plan quizzes mandatory in every classroom. Teachers are required to take monthly behavior plan quizzes based on the behavior plans of the students they work with. Behavior plan quizzes are required monthly in the school, each month; teachers will take a behavior plan quiz of a different student, set on a rotation schedule. If there are 6 students in the classroom, teachers will take 6 different behavior plans over the course of 6 months, and at the end of the cycle start over again.

Behavior plan quizzes are meant to assess teacher knowledge of behavior plans, the complexity of the plan varies with each student, but it is universal across the school that every hour and activity of a student’s day, except for lunch and recess, is scripted in the behavior plan with ensuing academic and/or activities based programs combined with contingencies and complicated procedural sequences of how a program should be run and reinforced. With the overall complexity and intricacy of the behavior plans, the school wide policy already in place is to give out behavior plan quizzes to teachers in order to assess staff member knowledge of plan implementation and procedural sequences for the students they work with in the classroom. Behavior plan quizzes are given out to staff members every month, the quizzes are 20 questions each and are based directly on the behavior plan quizzes. Behavior plan quizzes are created by the director of the school, who also writes the behavior plans, with input from the educational coordinator. The behavior plan quizzes follows a similar template across the school, consisting of 20 questions, with 10 multiple choice questions, and 10 fill in the blank short answer questions. The quizzes are graded by both the evaluator and the writer of the plans (the director of the school) in order to assure that the quizzes are relatively similar in content and degree of difficulty. Each question is either marked right (or plus 1 point) or wrong (or 0 points).
divided by a total of 20 possible points, to yield a behavior plan quiz percentage. During the pre-evaluation meeting, evaluation of school-wide average performance of behavior plan quizzes and the utility of behavior plan quizzes along with the relationship between the quiz scores and treatment adherence was established.

In order to answer the evaluation question “What is the average sample of score of the behavioral plan quizzes taken by the teachers?” permanent products in the form of behavior plan quizzes were obtained. Since behavior plan quizzes were ongoing in the school, the behavior plan quizzes of the behavior plans that were observed in question 1 in relation to treatment adherence was gathered. Therefore, each teacher had two distinct scores on two different behavior plan quizzes, the quizzes were of the same behavior plans utilized in the evaluation of treatment adherence. A total of 48 behavior plan quizzes and scores were collected from August 2009 to June of 2009.

In keeping with previously established practices of the school, behavior plan quizzes were scored by the director of the school. The quiz scores were averaged into percentages and yielded a school-wide sample of behavior plan quiz performance. Data was further analyzed through descriptive statistical methods, and data was presented to administrators in the final feedback meeting at the end of the evaluation.

**Question 3**

The management plan for the second evaluation question, “Is there a relationship between treatment adherence percentages and behavior plan quiz scores?” is presented in Table 2-4 with details regarding the data collection instruments, and the methods and techniques used to analyze and interpret the data gathered.

Table 2-4. Management plan for evaluation Question 3: *Is there a relationship between treatment adherence percentages and behavior plan quiz scores?*
The information required for the assessment of Question 3 is the treatment adherence data gathered in Question 1 and the behavior plan quiz data gathered in Question 2. The evaluation question seeks to understand the existence of a relationship between treatment adherence and behavior plan performance. Treatment adherence data and behavior plan data was analyzed visually, through descriptive statistics, and correlated.

**Question 4**

The evaluation plan for Question 4, “What is the average sample of teacher’s level of engagement during recess?” is in Table 2-5.

Table 2-5. Management plan for evaluation Question 4: What is the average sample of teacher’s level of engagement during recess?
The fourth evaluation question, “What is the average sample of teacher’s level of engagement during recess?” centers around recess time, an area of concern for the primary stakeholders. During the planning meeting held by the principal evaluator and the school administrators prior to the start of the evaluation, the administrators identified recess as a ‘problem space’. They described recess as an unstructured time not covered by the behavior plans, where teachers were engaging in chatting and off-task behaviors. However, due to the concept of recess as a “free time” with a play dimension for students, putting in place set interventions or prescribed tasks for teachers and students was not appropriate. The stakeholders wanted teachers to stop engaging in off-task behaviors, but had trouble structuring recess with appropriate rules; furthermore, the stakeholders were not sure how to capture and measure the elusive concept of attention. From their own perspectives, when observing recess, the stakeholders reported teachers chatting, and playing on their cell phones while marginally monitoring students for safety.

Based on the needs of the school, the principal evaluator and the stakeholders came up with the concept of “engagement” to address the idea of attention. While there are many
descriptions of “teacher engagement” within the classroom, most definitions of engagement include action-oriented observable interaction between the teacher and student (Bussis et al, 1976). Engagement is also a familiar concept to the teachers in the school from professional development trainings: administrators would often refer to being “engaged” with students when describing basic classroom expectations. Combining the definitions already prevalent in the school with examples of what “engaged behavior” looks like, the operational definitions of engaged and unengaged teacher behavior, for the purpose of the program evaluation, are as follows:

**Engaged:** Talking to a student, redirecting self-stimulatory behavior, playing with a student actively (e.g., throwing a ball, playing chase, etc.), prompting a student to communicate or participating in a communicative interaction with the student, prompting students to participate in communicative interactions with each other or between student and staff, prompting student to use a play item appropriately, delivering social praise to students, and walking toward a student with the obvious intent of interaction.

**Unengaged:** Otherwise defined as not interacting with the student and/or not attending to the student. One examples of unengaged behavior is idle chatting, with chatting defined as conversation between two adults. The adults are usually standing within 5 feet of each other, either facing each other or with parallel bodies. Obvious dialogue between the two standing adults must occur in order to be counted as chatting. Another example is idle standing, defined as adults facing the students but chatting back and forth or adults looking over students and chatting to each other. However, dialogue that occurs between two adults and one or more students is not idle chatting if the student is also participating in the exchange, either verbally or behaviorally. A short greeting (no more than 2 sentences) exchange between 2 adults also does not constitute chatting.
Systematic direct observation is one of the most common methods of observing the behavior of individuals. The research of Volpe, Diperna, Hintze, & Shapiro (2005) presents the observation methods and their ensuing psychometric properties. As a method of observation, the focus on behaviors through systematic observation system and intervals, as exemplified in the Behavioral Observation of Students in School (BOSS; Shapiro, 2004), observes target behaviors such as active and passive engagement. Systematic direct observation is a reliable method of observing behavior. The engagement system presented for the observation of teacher engagement mirrors the BOSS system, wherein teachers are observed as either engaged or unengaged. Utilizing a 10-second partial-interval recording system, the evaluator and trained undergraduate data collectors recorded teacher behavior during 20-minute recess observations. Recess time in the school lasts for 30 minutes, but the first and last 5 minutes of recess were not observed to give students time to enter and/or leave. A portable cassette player with ear buds was used to delineate the start of each interval. Teachers were observed on a rotating schedule. For example, if there were six teachers at recess, they were numbered 1 through 6; the first 10-second interval was designated for observing teacher 1, the second 10-second interval went for observing teacher 2, and so on. With six 10-second intervals per minute, teacher 1 would be observed again during the first 10-second interval of minute 2. After a 20-minute observation, with a total of 120 10-second intervals, every teacher would have been observed 12 times, with teacher 1 always being observed during the first 10 seconds of a minute, and teacher 6 always observed during the last 10 seconds of a minute. If the teacher was engaged during any part of the 10 second observation, the behavior was coded as “e”, if the teacher was unengaged for the entirety of the 10 second observation; the observer would code the behavior as “u” or unengaged on the coding sheet.
Recess settings were similar within each observation, either inside during rainy or wintry days or outside at the playground on nice days. There were a minimum of 10 teachers during each recess observation, and a maximum of 14. During 50% of the recess observations, the evaluator and an undergraduate data collector or educational consultant simultaneously and independently observed recess. The percentage of engagement was derived from the number of intervals of “engaged” behavior by the staff divided by the total number of intervals, thus providing an engagement percentage. Prior to data collection, the lead evaluator conducted reliability checks in the application of the operational definitions to actual observed behaviors. The first 2 recess observation sessions were deemed as training sessions between the lead evaluator and 3 other data collectors. The lead evaluator and data collectors sat side by side, on a one-on-one basis, and simultaneously took engagement data based on observed behavior. The data from training sessions were then compared and discussed immediately after recess, any disagreements pertaining to interpretation of the operational definition of engaged versus unengaged was discussed. Data from the training sessions were not included in the final data set. The method of calculation will be based on Kazdin (1982), wherein the total number of points of agreements is divided by the total number of agreement points plus disagreement, multiplied by 100. Interobserver agreement was calculated using overall agreement calculations and the kappa statistic ($k$). The kappa statistic ($k$) ranged from .80 to 1, with an average ($k$) of .91.

**Question 5.**

The management plan for the first evaluation question, “Is there a significant difference in the teacher’s level of engagement when utilizing a covert versus overt observation approach?” is presented in Table 2-6 below with details regarding the data collection instruments, and the methods and techniques used to analyze and interpret the data gathered.
Table 2-6. Management plan for evaluation Question 5: *Is there a significant difference in the teacher’s level of engagement when utilizing a covert versus overt observation approach?*

<table>
<thead>
<tr>
<th>Information Required</th>
<th>Information Source</th>
<th>Method for Collecting Information</th>
<th>Information Collection Arrangements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nature of the shift in engagement when teachers were unaware of what was being observed (covert) versus knowledge of observation criterions (overt).</td>
<td>Systematic Direct observation during recess.</td>
<td>Systematic Direct Observation of Recess. 10 Second partial interval recording of Engaged versus Unengaged teacher behavior.</td>
<td>Whom – Evaluator and trained undergraduate data collectors. Conditions – systematic direct observation. When – 2 to 3 month intervals, 3 times a year.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Analysis Procedures</th>
<th>Interpretation Procedures and Criteria</th>
<th>Report of Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive statistics for teacher engagement levels during the covert-overt phase.</td>
<td>Analysis of difference scores on covert-overt engagement levels.</td>
<td>Whom- School administrators How-Feedback Meetings When-After the conclusion of Phase 1 and at the end of the evaluation.</td>
</tr>
</tbody>
</table>

The evaluation question “is there a significant difference in the teacher’s level of engagement when utilizing a covert versus overt observation approach?” evaluates the effectiveness of an intervention applied after Phase 2 feedback meetings with administrators. During Phase 1, engagement data during recess was collected and descriptive statistics of the data was utilized to assess school-wide recess teacher engagement levels during recess. Phase 1 of recess engagement data was collected during August to mid September, 2009. Data was presented to administrators during a feedback meeting. After the presentation of data, the lead evaluator and program administrators decided to apply an intervention to boost recess engagement percentages amongst teachers. During Phase 1 of data collection, the lead evaluator along with
data collectors observed recess, however teachers did not know what aspect of recess was being observed, hence Phase 1 was dubbed the “covert” observation phase. The intervention consisted of explaining clear expectations for recess engagement and posting a copy of those expectations; the expectations included the operational definition of engaged and unengaged teacher behavior during recess, with examples and non-examples of the behavior. Creating an explicit, simple, and consistent expectation for behavior, along with teaching and posting of those behaviors has been shown to be effective in promoting permanent behavior change and is from the field of School Wide Positive Behavior Support (SWPBS) (Horner & Sugai, 2007). Applied Behavior Analysis was the first and earliest influence of SWPBS, within the SWPBS implementation model, administrators and key stakeholders must identify positively stated behavioral expectations that promote the major social values of the school, define expectations of those values, explicitly teach these expectations, and monitor and encourage expected behaviors (Horner & Sugai, 2007; Biglan, Rusby, & Sprague, 2001). In this case, the administrators and lead evaluator designated engagement as the behavior to increase, and provided a definition of the behavior. During the intervention in Phase 2 of the evaluation, a school wide meeting was held after school, 22 out of 24 teachers attended, there were 2 teachers who missed the meeting due to personal reasons. The director of the school held a small make-up meeting with the 2 teachers who missed the school-wide meeting. School administrators conducted the meeting, during the school-wide meeting; teachers were explicitly informed of the definition of engaged and unengaged teacher behavior during recess. The behaviors were also modeled by school administrators, with time for questions from school staff. Two days after the school-wide meeting, the operational definitions for engaged and unengaged were sent to every teacher via email. After the school-wide intervention meeting, Phase 3 or the “overt” observation phase of recess began, the teachers now knew what the
evaluation team was observing during recess. The “overt” observation phase lasted from mid
September 2009 to June 2010. A total of 11 observations were conducted during the Phase 1
“covert” data collection phase, with 1320 datum points collected. During the Phase 2 “overt”
data collection phase, a total of 27 observations were conducted, with 3240 datum points
collected. Observations over both phases yielded a total of 38 observations, and 4560 datum
points collected.

Data analysis of Phase 1 “covert” observation phase and Phase 2 “overt” observation
phase utilized visual inspection, descriptive statistics, and Allison’s mean plus trend difference
(Allison & Gorman, 1993). The AB single case design of the data set, and the long and rather
stable baseline, justifies the utility of a statistical model that includes mean and trend
differences. The ALLISON-MT statistical model yields good power, and little autocorrelation,
providing that the data set has a stable baseline (Parker & Brossart, 2003).

**Question 6**

“Do teachers understand the importance of treatment integrity, and what are the
barriers to treatment integrity implementation as perceived by teachers?” is a perception and
attitude question utilizing qualitative assessment methods, including surveys, focus groups and
interviews, meant to assess staff’s perceptions and attitudes towards treatment integrity. The
management plan for Question 6 is presented in Table 2-7, along with data analytic plans.

<table>
<thead>
<tr>
<th>Information Required</th>
<th>Information Source</th>
<th>Method for Collecting Information</th>
<th>Information Collection Arrangements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nature and type of school based professional</td>
<td>Teachers and school administrators</td>
<td>Survey, focus group</td>
<td>Whom – Principal evaluator. When – Phase 1 and 2 of</td>
</tr>
</tbody>
</table>
development activities. Knowledge gained from participation in treatment integrity oriented professional development activities.

<table>
<thead>
<tr>
<th>Analysis Procedures</th>
<th>Interpretation Procedures and Criteria</th>
<th>Report of Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive statistics for surveys. Summarize major themes of interviews and focus groups.</td>
<td>Quantitative and Qualitative analyses. Focus group identify themes associated with understanding of treatment integrity, and perceived barriers of treatment integrity</td>
<td>Whom- School administrators How-Feedback Meetings When-After the conclusion of Phase 3 of the evaluation.</td>
</tr>
</tbody>
</table>

*Survey Instrument* The survey was based on Perepletchikova, Hilt, Chereji and Kazdin, (2009) and The Barriers to Treatment Integrity Implementation Survey (BTIIS) which was developed to assess possible impediments to addressing integrity of psychosocial interventions. The BTIIS consists of 30 questions rated on a scale from 1 to 6 (always disagree to always agree). Higher scores indicate more perceived barriers. Items on the survey encompass five domains of possible impediments to Treatment Integrity (TI) implementation, the lack of appreciation of TI barrier (Domain A: 4 items) includes the lack of awareness of the importance of TI and recognition that the validity of the intervention may be rendered questionable without manipulation checks on treatment delivery. Domain B includes 8 items, and assesses the general knowledge about TI barrier including the lack of knowledge about TI procedures for adequately monitoring and documenting treatment delivery. Domain C has 7 items and assess the lack of specific guidelines on establishing, assessing, evaluating, and reporting integrity procedures, the time, cost and labor demands barrier related to TI. Domain D has 5 items and evaluates the time limitations, labor constraints, and funding barriers to addressing integrity procedures.
Domain E has 6 items and address the lack of the requirement for implementing, addressing, evaluating, and reporting TI procedures in order for a study to be published. In the Perepletchikova et al, 2009 survey, item responses with a means of less than 3 were considered “not barriers”; items with mean ratings more than 3 and less than 4 were considered “barriers”; items with mean ratings of more than 4 were considered “strong barriers.”

For the purposes of this dissertation, the following modifications to the BTIIS survey were made, the first being that only domains A through D were utilized. The domains have direct utility within the behaviorally oriented school. In order for the survey to be utilized seamlessly within the school, domain E was dropped. The 6 questions within domain E that was meant to address TI requirements when publishing articles had little utility for the practically oriented school. Thus the questions were dropped from the survey. On distinct set items, specific words were also modified, i.e. taking the question “Treatment fidelity is not imperative in ensuring adequate experimental control” and changing the word experimental into “intervention”. Care was taken to not modify the intent of the question. The survey created by Perepletchikova et al, 2009 was initially created for academics conducting research on TI, thus language was tailored to fit the staff of a behavioral school. However, care was taken to not change the overall meaning of the question.

Survey Procedures The surveys were distributed to staff members at the end of a weekly Friday afternoon staff meeting. The school administrators and researchers stepped out of the room after surveys were distributed. Instructions were presented on the front page of the surveys, and after completion of the anonymous surveys, teachers placed the surveys in communal envelope. Of the 24 teachers present, all 24 completed the survey. Survey data was assessed through descriptive statistics.
Focus Group The questions presented to teachers at the focus group included questions assessing staff attitudes towards the overall importance of treatment integrity in intervention, questions include: 1.) Please provide a definition for treatment integrity? 2.) Do you think treatment integrity is important in the implementation of interventions, please offer reasons as to why or why not you think treatment integrity is important? 3) What percentile of treatment integrity do you think is needed to implement the intervention (behavior plan) effectively? 4.) What do you think your own treatment integrity percentiles are? 5.) What are the barriers that prevent you from implementing the intervention (behavior plan) with integrity? 6.) What makes it easy to implement the intervention (behavior plan) with integrity; 7.) What suggestions do you have to make it easier for you to follow the intervention (behavior plan)? Focus group responses were analyzed and thematic units were identified. Thematic units are defined as “recurring systems of beliefs or explanations”, and were identified and separated based on context, big picture ideas, frequency, and consensus between teachers (Kruger & Casey, 2000).
CHAPTER 3

RESULTS

In this chapter I report results from the data collection methods presented in the previous chapter. The results will be presented for each research question in turn.

Question 1.

*What is the average sample of teachers’ adherence to behavior plans within the school?*

As described in Chapter 2 (refer to the evaluation management plan for Question 1), data on teachers’ adherence to behavior plans were collected in Phase 1 and Phase 3 of the program evaluation. Phase 1 data collection (August 2009 to February 2010) consisted of 3 distinct 1-hour observations per teacher, with 24 teachers in the school, yielding a total of 72 observations. Each teacher was observed implementing the intervention behavior plan in their classroom (1 behavior plan per each of 4 classrooms). These 72 observations provided data for a school-wide average of treatment adherence to behavior plans. Phase 1 data (see Table 3.1) were shared with administrators during a feedback meeting (Phase 2). In particular, administrators learned that the school-wide treatment adherence average was 70.54%, with the range of adherence varying from a low of 30% adherence to a high of 100% adherence. Working with this information, it was decided that continued observations were necessary in order to provide the school with a more robust sample of overall treatment adherence compliance to behavior plans. The resulting Phase 3 data collection (from early March 2010 to June 2010) consisted of 3 distinct 1-hour observations per teacher; with 24 teachers in the school, this again yielded a total of 72 observations. The observations were conducted on 4 different discrete behavior plans than those utilized in the first wave of data collection, with 1 behavior plan per classroom. After completing Phase 3, the average treatment adherence percentage was once again shared with school administrators. The average treatment integrity adherence
percentage obtained from both Phase 1 (72 total observations) and Phase 3 (72 total observations) of data collection was calculated from a total of 144 hours of observations. Over all of the phases, each of the 24 teachers in the school was observed six distinct times, yielding six distinct school wide-observation averages that make up the total treatment adherence data across the school.

The analysis presented in Table 3-1 provides the descriptive statistics for the six distinct observation time points along with the average of overall treatment adherence to behavior plans across the school. The table shows that the average treatment adherence percentage for the school is 70.54%, with a standard deviation of 14.82, and a range of 30 to 100.

Table 3-1 Descriptive Statistics for Treatment Adherence Percentages of All Behavior Plan Observations.

<table>
<thead>
<tr>
<th>Observation (N=24)</th>
<th>Mean Treatment Adherence Percentage</th>
<th>SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation 1</td>
<td>71.50</td>
<td>15.22</td>
<td>33.33 – 91.25</td>
</tr>
<tr>
<td>Observation 2</td>
<td>72.28</td>
<td>18.49</td>
<td>33.00 – 100.00</td>
</tr>
<tr>
<td>Observation 3</td>
<td>68.14</td>
<td>17.19</td>
<td>33.00 – 90.00</td>
</tr>
<tr>
<td>Observation 4</td>
<td>73.98</td>
<td>13.03</td>
<td>48.60 – 95.00</td>
</tr>
<tr>
<td>Observation 5</td>
<td>66.40</td>
<td>16.10</td>
<td>33.33 – 100.00</td>
</tr>
<tr>
<td>Observation 6</td>
<td>70.93</td>
<td>9.05</td>
<td>55.00 – 92.00</td>
</tr>
<tr>
<td>Total (N=144)</td>
<td>70.54</td>
<td>14.82</td>
<td>30.00 – 100.00</td>
</tr>
</tbody>
</table>

Question 2.

*What is the average sample score of the behavioral plan quizzes taken by the teachers?*

The data used for addressing the second evaluation question consisted of permanent products in the form of behavior plan quizzes. The quizzes were administered and scored during November of 2009 and February of 2010. As described in Chapter 2, the data here include two behavior plan quizzes taken by all 24 teachers at the school at the two separate time points,
totaling 48 behavior plan quizzes; each teacher’s two scores were averaged together. Each classroom had a unique behavior plan; teachers working in the same classroom took the same behavior plan quiz.

Table 3-2 Behavior Plan Quiz Average (N=24).

<table>
<thead>
<tr>
<th>Mean behavior plan quiz score</th>
<th>SD</th>
<th>Range of scores</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>75.32</td>
<td>9.165</td>
<td>56-91</td>
<td>75</td>
</tr>
</tbody>
</table>

The mean for behavior plan quizzes is 75.32% out of a possible 100%.

**Question 3.**

*Is there a relationship between treatment adherence percentages and behavior plan quiz scores?*

In order to answer Question 3, correlations between results from Question 1 (treatment adherence percentages) and Question 2 (behavior plan quiz scores) were estimated; these results are presented in Table 3-3.

Table 3-3 Correlations between Average Behavior Plan Quiz Scores and Average Treatment Adherence (N = 24).

<table>
<thead>
<tr>
<th>Correlation between average behavior plan quiz score (75.32) and treatment adherence averages</th>
<th>Significance level</th>
<th>r</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation 1:</td>
<td>0.05/6 = 0.0083</td>
<td>.234</td>
<td>0.271</td>
</tr>
<tr>
<td>Observation 2:</td>
<td>0.05/6 = 0.0083</td>
<td>.429</td>
<td>0.036</td>
</tr>
<tr>
<td>Observation 3:</td>
<td>0.05/6 = 0.0083</td>
<td>.188</td>
<td>0.380</td>
</tr>
<tr>
<td>Observation 4</td>
<td>0.05/6 = 0.0083</td>
<td>.384</td>
<td>0.064</td>
</tr>
<tr>
<td>Observation 5</td>
<td>0.05/6 = 0.0083</td>
<td>.093</td>
<td>0.665</td>
</tr>
<tr>
<td>Observation 6</td>
<td>0.05/6 = 0.0083</td>
<td>.488</td>
<td>0.016</td>
</tr>
<tr>
<td>Median treatment adherence score:</td>
<td>0.05</td>
<td>.477</td>
<td>0.0184*</td>
</tr>
</tbody>
</table>

~ p ≤ .10   *p ≤ .05   **p ≤ .01   ***p ≤ .001

Table 3-3: Correlation between Average Behavior Plan Quiz Scores and Average Treatment Adherence

The table shows first the results of estimating correlations between the average behavior plan quiz score and each of the six average treatment adherence rates. Because one predictor
(average behavior plan quiz score) and six dependent variables (six different school wide averages of treatment adherence) were used in this analysis, the significance level was adjusted for family wise error (i.e., the conditional probability of rejecting one or more absolutely true null hypotheses in a family of several absolutely true null hypotheses) with the Bonferroni correction, which lowers the alpha level when several dependent or independent statistical tests are being performed simultaneously (Shaffer, 1995). Hence for these first six estimations, the significance level was p<.0083 (.05/6). The results indicated that there was no association between the average behavior plan quiz scores and any of the six average adherence scores.

An association was then explored between the average behavior plan quiz score and the median treatment adherence score of all six observations. Because in this analysis, a single predictor (average behavior plan quiz score) and one dependent variable (median treatment adherence score) was used, a significance level of .05 was applicable. The correlation between quiz scores and median treatment adherence observation score was significant and positive, suggesting that higher teacher performance on behavior plan quizzes is associated with a higher level of treatment adherence to behavior plans.

**Question 4:**

*What is the average sample of teacher’s level of engagement during recess?*

**Question 5:**

*Is there a significant difference in the teacher’s level of engagement when utilizing a covert versus overt observation approach*

In order to fully present the scope of the teacher engagement data set; evaluation Questions 4 and 5 will be addressed together. As described in more detail in Chapter 2, Questions 4 and 5 involved collecting recess engagement data in two major phases. Prior to the start of this data collection, the evaluator and school administrators together drafted
operational definitions of engaged and unengaged staff behavior during recess. Phase 1 of recess engagement data were collected according to these definitions, utilizing a covert method: observers were present during recess, but teachers were not informed as to the nature of the observation. Data were collected until an average of overall recess engagement levels was established. These data were analyzed and presented to administrators during a feedback meeting.

At the feedback meeting, it was decided that teacher engagement during recess was an area that required intervening. In order to boost teacher engagement during recess, the evaluation team decided on the following interventions: clearly defining behavioral expectations of teachers during recess, clearly posting behavioral expectations during recess, and informing teachers of the nature of recess observations (overt observations). As described in Chapter 2, teachers were explicitly informed of the operational definition of engaged and unengaged teacher behavior during recess during a routine school-wide meeting. The behaviors were modeled by school administrators, followed by a question and answer session to clarify any confusion. It was also made clear during the staff meeting that recess time was an extension of the school day, and hence teachers were expected to be engaged as exemplified by the operational definition. Within a week of the after school meeting, the operational definitions for engaged and unengaged were sent to every teacher via email. The staff meeting marked the end of Phase 1 (covert) and the start of Phase 2 (overt) recess engagement data collection. After this meeting, Phase 2 observation of recess began, with teachers now aware of what the evaluation team was observing during recess.

Twenty-two out of 24 teachers attended the meeting; two teachers missed the meeting. The director of the school held a small make-up meeting for these teachers to inform them of the new expectations of teacher engagement during recess. However due to scheduling
conflicts, the meeting did not occur until the first datum point of Phase 2 was already collected. Therefore, observation data on these two staff members for the first day of Phase 2 data collection were not included in the overall averages and analysis, both to protect staff member anonymity and to protect the integrity of recess data.

Recess observation protocols (highlighted in Chapter 2) did not change for the intervention phase, except for staff members’ awareness of the nature of the observations, and the behavioral expectations. Table 3-4 below shows the average engagement levels during Phase 1 (covert) and Phase 2 (overt) observation periods.

Table 3-4 Descriptive Statistics for Teacher Engagement Percentages during Recess.

<table>
<thead>
<tr>
<th>Recess Observation Phase</th>
<th>Mean Teacher Engagement Percentage Score</th>
<th>SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 1 (N = 11):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Covert observation,</td>
<td>60.79</td>
<td>13.24</td>
<td>32.64 – 83.88</td>
</tr>
<tr>
<td>1320 datum points</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase 2 (N = 27):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overt observation,</td>
<td>78.75</td>
<td>6.09</td>
<td>66.84 – 89.89</td>
</tr>
<tr>
<td>3240 datum points</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Recess Observations (N = 38):</td>
<td>73.55</td>
<td>11.89</td>
<td>32.64 – 89.89</td>
</tr>
<tr>
<td>(N = 38): 4560 datum points</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As described in Chapter 2, recess observations were conducted with a 10-second partial-interval recording system; each individual datum point represents 10 seconds of partial interval observation of engagement. As indicated in Table 3.4, a total of 11 observations were conducted during the Phase 1 “covert” data collection phase, with 1320 “10-second” datum points collected. During the Phase 2 “overt” data collection phase, a total of 27 observations were conducted, with 3240 datum points collected. Observations over both phases, which
occurred over a period of nine months, yielded a total of 38 observations and 4560 datum points; on average over 190 datum points were observed per teacher. As the summary above indicates, during baseline Phase 1 (covert) observations, teacher engagement averages during recess were lower (M=60.79, SD=13.24) in comparison with intervention Phase 2 (overt) observations (M=78.75, SD=6.09).

The graph in Figure 3-1 is a different view of the analysis presented in Table 3-4, highlighting the teacher engagement averages over time, and each distinct observation.

Figure 3-1: School-wide Teacher Engagement Averages during Recess at Two Time Points.

Assessment of staff engagement during recess was continuous over nine months. As the graph indicates, teacher engagement averages at recess showed a spike after the initial observation in the Phase 1 (covert) data collection, reaching a high of over 80% (Observation 2). This is perhaps due to initial staff reactivity to observation. While there are two important
outliers in the data (the initial spike at Observation 2, and a low point at Observation 7, overall the baseline showed variability (degree to which the performance fluctuate around a mean or slope during a phase) but looked relatively stable. Looking at Phase 2 (overt) data, the graph indicates that the average teacher engagement during recess increased immediately after beginning the intervention, and remained consistently high with some variability in the data. The immediate appearance of change after applying the intervention indicates little to no latency effect between the two phases. Closer inspection indicates a pronounced change in level between Phase 1 (covert) and Phase 2 (overt) observation data, indicating a pronounced change in mean performance between Phase 1 and Phase 2. Introducing the intervention does appear to be associated with a higher average level of teacher engagement at recess, in addition to a different range of teacher engagement behaviors.

To supplement visual inspection of these data, a single subject regression procedure, Allison’s Mean plus Trend difference (Allison MT; Allison & Gorman, 1993), was also conducted on the recess data set in order to further confirm the conclusions drawn from visual inspection. In comparison to other procedures utilized for calculating single-case effect sizes, the Allison-MT procedure has a high level of power, and yields to little autocorrelation (i.e., the correlation of a time series with past and future values; Parker, Brossart, Vannest, Long, De-Alba, Baugh, & Sullivan, 2005). The Allison-MT accounts for both trend and mean differences between phases and is summarized as follows: (1) Based on the baseline data, a line of best fit is calculated; (2) Based on what would be predicted by baseline trend, the residuals for intervention data are calculated by a comparison of observed to predicted; (3) Data are dummy coded by phase to create a new variable; and (4) Residuals are regressed onto the dummy coded variable (Solomon, Klein, Hintze, Cresssey, & Peller, 2011). Hence, after application of this procedure, it yields an effect size to substantiate visual inspection procedures. When using the Allison-MT
procedure, the traditional benchmarks set by Cohen for assessing single case effect sizes are not applicable. Instead, the effect sizes are compared to values recommended by Parker et al. (2005): low effect sizes are equal to or greater than $R^2 = .35$, medium effect sizes equal to or greater than $R^2 = .65$, and high effect sizes are equal to or greater than $R^2 = .90$. The results of the Allison-MT analysis indicated yielded a *medium* effect size (adjusted $R^2 = .70$) indicating that the change in the teacher engagement averages in the baseline Phase 1 (covert) observation phase, and the intervention Phase 2 (overt), as observed through visual inspection is further confirmed by the regression procedure.

The graph in Figure 3-2 highlights two data points, the last recess observation of Phase 1 (covert) and the first recess observation of Phase 2 (overt). These two data points represent the two teachers who were not informed about the expectations for teacher engagement at recess and the observations at recess until after the first two observations of Phase 2. Although the data on these two staff members from the first day of Phase 2 were not included in the overall averages and analysis of data, they are presented here and merit attention.

Figure 3-2
Last Day of Phase 1 (covert) Observations versus First Day of Intervention Phase 2 (overt) Observations.
As in Figure 3-2, the graph shows higher levels of staff engagement during recess on the first day of Phase 2 observations. The highlighted points (in yellow) show the average engagement of the 2 staff members who were absent from the school-wide meeting. Since the two staff members had missed the meeting, they were unaware of the new recess expectations of being engaged. Their average engagement levels (M=30, SD=7.07) are lower when compared to those of their peer teachers observed during recess on the same day (M=85.39, SD=10.43).

The graphs in Figure 3-3 make the comparison explicit. Both graphs depict the last day of baseline Phase 1 (covert) observation and the first day of intervention Phase 2 (overt) observations. The graph on the left (from Figure 3-3) includes the datum points of the absent staff members (again, highlighted in yellow).

Figure 3-3: Comparison of Graphs With and Without Absent Staff Member Datum Points.
The graph on the left has two distinct low datum points, those of the absent staff members, indicating their lower average engagement levels on the playground when compared with their peers. In comparison, the graph on the right without the absent staff members shows a change in level between baseline and intervention. In addition to level change, the graph on the right also displays no latency effect and a higher mean when comparing baseline and intervention phases. This comparison highlights the importance of the lack of information about engaged behavior at recess, and the absence of expectation about engagement at recess.

**Question 6.**

*Do teachers understand the importance of treatment integrity, and what are the barriers to treatment integrity implementation as perceived by teachers?*

Data collected from teacher surveys and focus groups were utilized to address Question 6. Data from each information source is presented below. Surveys were distributed to all teachers in the school. The focus group had a total of 11 participants, and lasted for approximately 1 hour 15 minutes. The focus group was taped and later transcribed; procedures for surveys and focus groups were described in Chapter 2.

**Surveys**

Of a total of 24 teachers present, all 24 self-selected to participate in the survey. The survey, based on The Barriers to Treatment Integrity Implementation Survey (BTIIS) survey
(Perepletchikova et al, 2009), has 24 questions rated on a Likert scale from 1 (always disagree) to 6 (always agree). Items on the survey encompass four domains (A – D) of possible impediments to Treatment Integrity (TI). Domain A, with 4 items, indexes lack of awareness of the importance of TI and recognition that the validity of the intervention may be rendered questionable without manipulation checks on treatment delivery. Domain B, with 8 items, assesses the general knowledge about TI barriers including the lack of knowledge about TI procedures for adequately monitoring and documenting treatment delivery. Domain C has 7 items and assesses the lack of specific guidelines on establishing, assessing, evaluating, and reporting integrity procedures, including the time, cost and labor demands barriers related to TI. Domain D has 5 items and evaluates the time limitations, labor constraints, and funding barriers to addressing integrity procedures.

All respondents answered each of the 24 questions. The survey responses were entered into an SPSS database by evaluators and research assistants for analysis. Averages for the scores of each question were calculated (see Table 3-5), as was an average score for each domain (see Table 3-6). Items with mean rating of ≤ 3 are considered “not barriers,” items with mean rating > 3 and ≤ 4 are considered “barriers,” and items with mean rating of > 4 are considered “strong barriers.” Therefore, higher scores on the survey items indicate more perceived barriers towards the implementation of TI procedures.

Table 3-5 Mean Item Ratings for Barriers to Treatment Integrity Implementation Survey (N=24)

<table>
<thead>
<tr>
<th>Domain</th>
<th>Survey Item (1-24)</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Treatment fidelity is not regarded as imperative in ensuring adequate intervention control</td>
<td>2.38</td>
<td>1.44</td>
</tr>
<tr>
<td>C</td>
<td>There is an inconsistency in the terminology of the aspects of treatment integrity (e.g., behavior plan adherence, teacher competence, intervention differentiation).</td>
<td>3.84</td>
<td>0.99</td>
</tr>
<tr>
<td>C</td>
<td>The field of ABA does not agree as to what is the appropriate method of integrity assessment</td>
<td>2.54</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>Insufficient resources due to constrained funding hinder the adequate implementation of integrity procedures.</td>
<td>4.42</td>
<td>0.90</td>
</tr>
<tr>
<td>A</td>
<td>Report of the treatment integrity procedures is not considered to enhance the credibility of the treatment outcome results.</td>
<td>2.38</td>
<td>1.26</td>
</tr>
<tr>
<td>C</td>
<td>The definition of treatment (intervention) adherence in the literature is ambiguous.</td>
<td>3.08</td>
<td>1.08</td>
</tr>
<tr>
<td>B</td>
<td>The requirements of internal review boards hinder implementation of integrity procedures (e.g., limiting how data are handled and linked to specific teachers, pushing for audio instead of videotaping).</td>
<td>3.23</td>
<td>0.95</td>
</tr>
<tr>
<td>B</td>
<td>Interventions are not sufficiently manualized to permit adequate integrity implementation.</td>
<td>3.38</td>
<td>0.80</td>
</tr>
<tr>
<td>D</td>
<td>Designing and validating treatment integrity measures is labor intensive and time consuming.</td>
<td>3.62</td>
<td>1.20</td>
</tr>
<tr>
<td>B</td>
<td>Once established, adherence and competence (to behavior plan) are believed to be stable and not to fluctuate over time.</td>
<td>2.94</td>
<td>1.15</td>
</tr>
<tr>
<td>C</td>
<td>There are no conventional criteria that specify acceptable levels of treatment integrity.</td>
<td>3.11</td>
<td>1.01</td>
</tr>
<tr>
<td>C</td>
<td>The guidelines for evaluating psychometric properties (validity and reliability) of the treatment integrity measures are unclear.</td>
<td>3.04</td>
<td>1.04</td>
</tr>
<tr>
<td>D</td>
<td>It is expensive and time consuming to provide direct training to teachers (e.g., viewing class time tapes, providing feedback, having regular meetings with staff, role-playing techniques).</td>
<td>3.77</td>
<td>1.53</td>
</tr>
<tr>
<td>B</td>
<td>Teachers may resist close supervision and monitoring of treatment implementation.</td>
<td>3.96</td>
<td>0.82</td>
</tr>
<tr>
<td>D</td>
<td>There is a considerable time requirement in obtaining accurate representation of integrity data (collection of data across teachers, situations, cases and days).</td>
<td>4.31</td>
<td>1.16</td>
</tr>
<tr>
<td>D</td>
<td>Performing manipulation checks on the integrity of behavior plan implementation may be risky, as adherence and competence may be lower than desired and low competence may be lower than desired (e.g., credibility of program data may be compromised by reporting low levels of integrity).</td>
<td>3.46</td>
<td>1.36</td>
</tr>
<tr>
<td>C</td>
<td>There are no established criteria or principles by which treatment integrity may be judged</td>
<td>2.50</td>
<td>0.81</td>
</tr>
<tr>
<td>D</td>
<td>High labor costs may preclude administrators from employing or training integrity raters.</td>
<td>4.50</td>
<td>0.76</td>
</tr>
<tr>
<td>B</td>
<td>Treatments (interventions in behavior plans) are presumed to be effective if significant changes on the dependent measures are obtained regardless of the integrity level of intervention implementation.</td>
<td>3.69</td>
<td>1.29</td>
</tr>
<tr>
<td>B</td>
<td>Intervention manuals are not widely employed because they are thought to limit teacher flexibility in addressing a student’s problems and tailoring of intervention to the individual needs or to address in the moment issues.</td>
<td>3.38</td>
<td>0.98</td>
</tr>
<tr>
<td>A</td>
<td>Once the training of a teacher is completed, supervision and monitoring of treatment implementation does not justify the time</td>
<td>2.27</td>
<td>0.92</td>
</tr>
</tbody>
</table>
and labor costs

A  The cost of implementing integrity procedures outweighs the possible benefits.  2.58  1.10

B  It is generally believed that integrity procedures can be implemented primarily with behavioral interventions but not with other approaches, such as psychodynamic or interpersonal treatments.  3.23  0.91

C  Teacher (interventionist) competence is not clearly defined in the literature.  3.84  1.14

Table 3-6 Survey Results: Domain Averages

<table>
<thead>
<tr>
<th>Domain</th>
<th>Domain M</th>
<th>SD</th>
<th>Item Rating M</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain A: Lack of appreciation of TI</td>
<td>9.62</td>
<td>3.31</td>
<td>2.41</td>
<td>4-19</td>
</tr>
<tr>
<td>Domain B: Lack of knowledge about TI</td>
<td>23.84</td>
<td>3.23</td>
<td>3.41</td>
<td>17-29</td>
</tr>
<tr>
<td>Domain C: Lack of theory and specific guidelines</td>
<td>22.00</td>
<td>5.20</td>
<td>3.14</td>
<td>12-29</td>
</tr>
<tr>
<td>Domain D: Time, cost and labor demands</td>
<td>24.08</td>
<td>4.28</td>
<td>4.01</td>
<td>15-30</td>
</tr>
</tbody>
</table>

Domain A has 4 items, and a score range of 4-24. The average for the entire Domain (all 4 items) is M=9.62, SD=3.31. The average item response for Domain A is 2.41 indicates that teachers do not perceive lack of appreciation for TI procedures as a barrier to TI implementation in the school. Hence, scoring on Domain A indicated that the teachers at the Behavior Day School are aware of the importance of TI and believe that reporting TI enhances credibility of treatment outcome.
Domain B has 7 items, and the total possible score range was 7 to 42. The total average for Domain B (M=23.84, SD=3.23, and Item Ratings M = 3.41) indicates that teacher opinion on average was between sometimes disagree to sometimes agree that lack of general knowledge about TI procedures for adequately monitoring and documenting treatment delivery is a barrier to TI implementation.

Domain C has 7 items and a possible score range of 7 to 42. The total average for Domain C (M=22, SD=5.2, and Item Rating M = 3.14) indicated that, on average, teacher opinion was largely sometimes disagree that lack of specific guidelines on establishing, evaluating and assessing integrity procedures are a barrier to TI implementation within the school.

Domain D has 6 items and a possible score range of 6 to 36. The average total for Domain D (M=24.08, SD=4.28, and Item Rating M = 4.01) indicated that teacher opinion was largely sometimes agree that time limitations, labor constraints, and funding barriers are strong barriers to TI implementation within the school.

Focus Groups

The focus group had a total of 11 participants. The focus group was conducted during a professional development day. The group was voluntary. Every classroom of the school was represented among the 11 participants. The list of questions below was presented to teachers as a guide for the focus groups. However, related topics did come up, and the facilitator allowed teachers to deviate from set questions.

1.) Please provide a definition for treatment integrity.

2.) Do you think treatment integrity is important in the implementation of interventions, please offer reasons as to why or why not you think treatment integrity is important?

3.) What percentage of treatment integrity do you think is needed to implement the intervention (behavior plan) effectively?
4.) *What do you think your own treatment integrity percentages are?*

5.) *What are the barriers that prevent you from implementing the intervention (behavior plan) with integrity?*

6.) *What makes it easy to implement the intervention (behavior plan) with integrity;*

7.) *What suggestions do you have to make it easier for you to follow the intervention (behavior plan)?*

Themes from the focus group were identified and separated based on context, big picture ideas, frequency of themes, consensus amongst the teachers (i.e., whether the teachers agreed with each other) (Kruger & Casey, 2000). Focus group conversation indicated that teachers agreed that multiple methods of assessing treatment integrity are essential for obtaining a clear picture of a staff members overall fidelity. This was discussed further as multiple teachers cited anecdotal examples of certain teachers doing very well on direct observations and reacting appropriately to the target behaviors of kids, but scoring badly on quizzes. Others noted that, while some teachers scored beautifully on behavior plan quizzes, they did not implement the behavior plan as intended during direct observations. All teachers agreed that the anecdotes provided illustrate the need for multiple methods of looking at treatment fidelity.

Another major theme within the focus group was teachers’ lack of consensus as to which method of monitoring treatment integrity was best. All the teachers agreed that multiple methods were important, but did not agree on the exact combination of methods that would yield the most precise overview on an individual’s fidelity. Some teachers suggested direct observations and behavior plan quizzes, while other teachers suggested direct observations and role playing. During this discussion, a theme that emerged was the lack of resources, primarily time, to try all the methods “out there” used to monitor treatment integrity. All teachers
nodded in agreement at the multiple comments expressing the feeling that, despite the desire to observe more and conduct more direct observation, checks are sometimes unrealistic due to time or staffing constraints. Along this line of conversation it was also suggested that the more frequently observations and IOA (Inter Observer Agreement) feedback occur, the better for treatment integrity. All staff members nodded agreement to the suggestions provided.

Another major theme identified in the focus group was staff reactivity to observations. The resounding consensus from participants was “Yes, observations absolutely do cause reactivity.” However, every member of the focus group cited the decrease of reactivity over time and the need for staff observations in ensuring treatment integrity. Teachers in the focus group also highlighted the importance of frequent observations and that “the more frequent the observation the less reactivity”. Another common theme among teachers was the utility of observations: when a staff member said, “There are other staff members who will ask me to do observations to make sure they are doing a procedure correctly and that everyone is on the same page”, others around the table nodded in agreement.

Despite the many points of agreement within the focus group, teachers did not come to consensus on the crucial question of the exact combination of methods needed to assess treatment integrity.
CHAPTER 4

DISCUSSION

Treatment integrity (TI) is a large and highly complex construct. The multiple dimensions of TI across a diverse range of fields not only create a lack of knowledge of TI procedures, but also a highly inconsistent set of guidelines on the practices of evaluating and assessing TI. Along with the complexities of the construct, there are also the logistical challenges of tailoring TI assessment and monitoring recommendations from research to the practical applications of those recommendations. The purpose of this study was to conduct an initial program evaluation of treatment integrity levels, including practices and attitudes, in an alternative education setting within a Participatory Intervention Model (PIM), or partnership model, emphasizing collaboration between the evaluator and primary stakeholders. The following discussion will include the evaluation of environment-specific constructs of TI, and the relationships among the constructs of treatment adherence and interventionist knowledge. The discussion will also look at the Participatory Intervention Model (PIM) which entailed collaboration between the evaluator and school administrators (primary stakeholders of the evaluation), and resulted in the program-specific definition, assessment and intervention to boost teacher engagement during recess. In addition, staff attitudes regarding perceived barriers to TI implementation in the school will also be discussed.

The evaluation was conducted in three phases. In Phase 1, various assessments were utilized to monitor treatment integrity specific to the school using qualitative and quantitative data collection methods. Phase 1 saw a copious amount of data collection for the project due to the important goal of obtaining an initial sampling of different types of treatment fidelity within the school. In Phase 2, data were shared with administrators and it was decided to implement an intervention targeting the staff's unclear expectations with regard to recess fidelity. Finally,
Phase 3 focused on follow-up, in which data were once again collected to assess the efficacy of the intervention, or to create a more robust data set. Measures of data collection were chosen to address these evaluation questions:

(1) What is the average sample of teacher’s adherence to behavior plans within the school?

(2) What is the average sample score of the behavioral plan quizzes taken by the teachers?

(3) Is there a relationship between treatment adherence percentages and behavior plan quiz scores?

(4) What is the average sample of teacher’s level of engagement during recess?

(5) Is there a significant difference in the teacher’s level of engagement when utilizing a covert versus overt observation approach?

(6) Do teachers understand the importance of treatment integrity, and what are the barriers to treatment integrity implementation as perceived by teachers?

The evaluation questions will be discussed briefly, followed by a discussion of this study more broadly.

**Conducting a Program Evaluation within a Participatory Intervention Model (PIM)**

The overall goal of this evaluation was to obtain a sampling of data on the current practices, assessments and attitudes utilized in the school to assess the multifaceted concept that feeds into the umbrella term of treatment integrity. This was done within a Participatory Intervention Model (PIM), integrating theory and research in developing context-specific interventions meant to promote ownership among stakeholders who are responsible for the long term use of the interventions after consultation ceases (Nastasi et al., 2000). Thus, the interventions created within the PIM model (i.e., boosting teacher engagement during recess)
was a collaboration between the researchers and school administrators, or the primary stakeholders, in this evaluation. This study highlights an important point: that this context-specific evaluation of a multifaceted construct of TI, individualized to the school, was successful in integrating evidence-based practice along with identified needs and priorities of primary stakeholders who formed a partnership. Treatment integrity was not only monitored by the evaluator utilizing monomethod assessments; rather, the evaluator and school administrators determined together what was monitored and what interventions were created. The importance of exploring the multiple aspects of TI and their adaptation within the school points to the pressing need for a critical and comprehensive review of treatment integrity utilizing a partnership model, wherein the specific needs of the school are prioritized and the evaluation tailored to fit the school’s needs. The partnership between the evaluator and administrators yielded a program evaluation where previously existing methods of assessing TI (i.e., behavior plan quizzes) were utilized along with the introduction of direct observation of teacher adherence to behavior plans. In addition to collaborating on the priorities of the evaluation, the partnership targeted areas of need within the school and provided an opportunity to intervene on teacher engagement during recess time. This was an area that was especially interesting and worthy of exploration because interventionist (teacher) engagement is not typically a construct associated with TI. However, within the context of recess, a social time for students, interventionist (teacher) engagement and attention was the primary expectation of implementing an intervention. Thus, defining teacher engagement and creating the intervention based on mutually agreed-upon criteria between the school administrators and the evaluators was the key to promoting a sense of ownership among the school administrators and the impetus for systems-level change regarding recess expectations.
Assessing Treatment Integrity

The first three evaluation questions address methods of evaluating TI within the school. Question 1 addressed the school-wide average of teacher adherence to behavior plans. Prior to this study, the Behavior Day School did not assess teacher adherence to behavior plans; since this is an initial program evaluation, the first that had been conducted in this manner within the school, one of the primary goals of this evaluation was to provide the school with an average of teacher adherence. As reviewed in Chapter 1, in the field of school psychology, treatment adherence (i.e., the extent to which the interventionist uses interventions and approaches prescribed by the treatment manual) is typically measured to assess treatment integrity (Schulte, Parker, & Easton, 2009). Treatment adherence is generally measured as the percentage of treatment steps completed within a prescribed treatment or intervention plan. Thus, prior to Phase 1 data collection, the evaluator and administrators had agreed to utilize a direct observation data collection system to assess staff adherence to treatment steps within individualized student behavior plans, thus yielding a percentage of teacher adherence to treatment steps within the behavior plan. Lane and Beebe-Frankenberger (2004) suggested it is best to utilize intervention scripts in conjunction with one or more assessment methods. Since the school already utilizes manualized intervention scripts in the form of behavior plans, observations of teacher adherence to behavior plans incorporate research into practice. By tailoring methods of assessing adherence suggested in research and personalizing it to the school, the evaluator and administrators melded research with practical application, once again embodying the spirit of a partnership model.

Also reviewed in Chapter 1, other methods most commonly used for evaluating treatment integrity include permanent products, including manualized treatments and intervention scripts (Lane & Beebe-Frankenberger, 2004; Lane et.al., 2001). Therefore, Question 2 addressed
teacher performance on behavior plan quizzes, using collated data on the already established TI assessment practice of behavior plan quizzes within the Behavior Day School. The intent of the behavior plan quizzes within the school was to assess teacher knowledge of manualized behavior plans of the students they work with. Thus, permanent products in the form of behavior plan quizzes were gathered, and teacher scores were averaged. Question 3 addressed whether there was a relationship between behavior plan quiz scores and direct observations of treatment adherence percentages. This question had direct practical implications regarding the utility of direct observations and behavior plan quizzes to assess TI within the school. Is there indeed a relationship between knowledge of behavior plans, and adherence to the plans? In research, the constructs of interventionist adherence and competence has been explored. The results were inconclusive regarding a relationship between interventionist adherence and interventionist competence. While competence presumes adherence, adherence does not necessarily presume competence. An interventionist can perform prescribed behaviors exactly, but not well (McGlinchey & Dobson, 2003). As reviewed in Chapter 1, there is a conceptual distinction between adherence and competence related to measurement: adherence is defined as a quantitative construct of treatment integrity (i.e., how frequently the interventionist implements procedures that are prescribed by the manual, and avoids those that are proscribed) while competency is defined as a qualitative construct (i.e., how well prescribed procedures are implemented). However, both adherence and interventionist knowledge in behavior plan quizzes are quantitative constructs, both measureable by percentages; hence Question 3 was framed around the existence of a relationship between interventionist (teacher) adherence and interventionist (teacher) performance on behavior plan quizzes. Data for Question 3 compared each of the six distinct and individual direct observation averages of teacher adherence to behavior plans to average behavior plan quiz scores. The results indicated
that there was not a statistically significant relationship between behavior plan quiz scores and any given individual observation of treatment adherence percentages. It appears that these individual observations of treatment adherence were unstable, perhaps due to inadequate reliability of the individual observations, and the amount of error present in the individual observations. However, by collapsing data across observations and using a median score taken from the six individual observations, there was a statistically significant relationship between the predictor (behavior plan quiz scores) and a more stable variable (median treatment adherence scores). Hence, it may be concluded that while observed teacher behavior varied substantially from observation to observation (among the six individual observations), overall the data demonstrates that teachers who scored higher on the quiz had a tendency to be more effective at behavioral management when measured. This indicates that there is a relationship between treatment adherence to behavior plans and interventionist quiz scores, or knowledge of those plans. It should be noted that there was only significance when the quiz scores were compared to median adherence scores, taken from all six observations.

The copious number of hours of observations (144 hours) that make up the data from which the median adherence score was calculated may have provided the stability required to find a significant relationship between behavior plan quiz scores and treatment adherence. However, the practically of collecting the amount of data required within a typical school day may be a barrier to the utility of behavior observations in practice. As suggested by research, the direct observation of treatment plan implementation is both time and labor intensive, likely resulting in a small sample due to time and labor constraints. The time and labor costs associated with direct observation were also a prominent theme within the focus groups, where teachers directed praise towards the utility of direct observation, but also complained about the time constraints and the practicality of carrying out multiple observations within an already filled
school day. Without the existence of this evaluation, and research assistants aiding in the data collection, the likelihood of direct observations continuing in the Behavior Day School is slight. Thus, when presenting recommendations to school administrators, one challenge was to find staffing and time to continue direct observations within the school.

Teacher Engagement during Recess

Questions 4 and 5 addressed the overall teacher engagement levels during recess. After Phase 1 data collection was complete, an intervention was applied to boost overall teacher engagement in the school. The period of baseline data collected prior to the intervention was dubbed Phase 1 (covert) observation, whereas the period after the intervention was referred to as Phase 2 (overt) observation. Prior to this evaluation, teachers’ behavior and tasks during recess time were not defined. Due to this lack of definition and structure, the behavior of teachers varied. As indicated by the baseline phase data, some teachers were engaged during recess and utilized the time well, prompting students in play and practice learned social skills, while others chatted aimlessly with each other and regarded recess time as a social time for the teachers. It was therefore the goal of the evaluator and school administrators to create consistent systems-level changes, redefining recess time for teachers, and teaching the new behavioral expectations systematically using an evidence-based model. The intervention applied here was based on the SWPBS (School Wide Positive Behavior Support) implementation model, in which a positively stated behavioral expectation that promotes the major social values of the school (in this case, “teacher engagement during recess”), and expectations of those values were defined (Horner & Sugai, 2007). After defining the value and behavior, behavioral expectations should be explicitly taught: this occurred during the staff meetings where the expected teacher engagement during recess was explicitly stated, operationally defined, and modeled for teachers. Lastly, the behavior should be monitored and expected behaviors
encouraged: this was done through continued observation of teacher engagement during recess, and explicitly informing teachers that the observers were looking at their “engagement” (Horner & Sugai, 2007; Biglan et al., 2001). Modeling is an example of systematic teaching. With modeling, a new behavior is acquired through demonstration and shaping strategies that differentially reinforce successive approximation of a target behavior, which in this case is the behavior of engagement. Although designed for students, the SWPBS model, heavily influenced by ABA (Applied Behavior Analysis) was applied to teachers within the school and redesigned to increase teacher engagement behavior.

As a result of this intervention, change occurred and teacher engagement during recess increased. Inspection of baseline and intervention data strongly suggested that the intervention was associated with a change in the overall staff engagement levels at recess. This was further confirmed by the Allison-MT statistical procedure, through which a medium effect size between the baseline and intervention phases was found. The intervention was successful in creating systems-level changes within the school: recess time is no longer an un-defined social time for teachers within the school, but seen instead as an extended work time for teachers and a valuable time for students to play and practice social skills.

**Teacher Attitudes and Perceptions of Treatment Integrity (TI)**

Question 6 assessed whether teachers understood the importance of treatment integrity and the perceived barriers to implementing treatment integrity. In order to address this question, a modified version of the Barriers to Treatment Integrity Implementation Survey (BTTIS; Perepletchikova et al., 2009), which assesses perceived barriers to TI, was distributed to teachers; a focus group was also conducted to incorporate teacher opinions regarding TI. The survey indicated that, on average, teachers at the school did not think that lack of appreciation for the importance of TI was a barrier to TI implementation (questions from Domain A),
suggesting that overall teachers in Behavior Day School understand the value and importance of TI. This was also validated in the focus group, where a major theme that emerged was the overall importance of monitoring TI.

Domain B of the survey assessed whether the lack of general knowledge about TI procedures for adequately monitoring and documenting treatment delivery is a barrier to TI implementation. On average, teachers perceived lack of general knowledge as a barrier to TI implementation. Domain C of the survey assessed whether the lack of specific guidelines on establishing, assessing, evaluating, and reporting integrity procedures are a barrier related to TI implementation. On average, teachers at the school also perceived these issues as a barrier to TI implementation. Similarly in the focus group, staff members also identified the complexity of defining and measuring the construct of TI as an area of confusion and a barrier to TI implementation. These responses highlight common issues about TI implementation not only at Behavior Day School, but in the field of school psychology. The contextual inconsistencies of TI, ranging from the differences in definition, approach, implementation and documentation of treatment, vary greatly between related fields (Jones, Clarke, & Power, 2008; Hagermoser Sanetti & Krachtowill, 2009). These differences are not limited to terminology; each field has a different model of treatment integrity, with inherent differences in service delivery models, settings, and intervention targets. Although this program evaluation was conducted in a single school, utilizing a uniformed ABA approach throughout the school, the complexity of the construct of TI is not absent from a rather insular environment like the Behavior Day School. TI is so prevalent in multiple forms in a diverse range of fields in which the overarching idea of integrity is similar, but the practical application of defining and monitoring TI is so divergent, that it is highly unlikely that teachers will come to the school having been exposed to the exact same model of TI. Hence, it is not surprising that the teachers of the Behavior Day School
perceived the general knowledge and inconsistencies of TI as a barrier. This is consistent with the complexity of the construct of TI and the inconsistencies in the field of school psychology regarding the construct of TI.

Teachers at the Behavior Day School perceived issues in Domain D (time limitations, labor constraints, and funding barriers to addressing integrity procedures) as a strong barrier to TI implementation within the school. This was also consistent within the focus group: teachers agreed that frequent and multiple methods of evaluating TI were best, but time and staffing constraints would make observations impossible. The focus group also highlighted the value of observations, and a need for multiple observations of teacher adherence to plan, noting that multiple observations will decrease reactivity over time and ensure that teachers are carrying out the intervention as planned.

Summary of Program Evaluation Results

As suggested by Posavac and Carey (2007), the primary purpose of a formative program evaluation is to provide data that will guide decisions about the future direction of the program, and provide initial data of existing practices, along with structured feedback to administrators in an effort to produce data-based, sustainable, and systems-level changes. This evaluation provided a school wide average of teacher adherence to behavior plans through observations. In addition to the initial teacher adherence data set, this evaluation also documented school-wide teacher performance on behavior plan quizzes, and found that there was a statistically significant correlation between behavior plan quiz scores and the median treatment adherence score. These results validated the current TI assessment practices of behavior plan quizzes within the school, and identified a relationship between median teacher adherence to treatment plans and behavior plan quiz performance.
Recess, a time during the school day at the Behavior Day School that was traditionally unscripted with inconsistent expectations for teachers was made consistent. Expectations of teacher engagement were operationally defined and baseline data was taken on school wide teacher engagement during recess. After baseline was established, an intervention, decided collaboratively with primary stakeholders following a partnership model, was applied to boost teacher engagement. The intervention was effective, and boosted school-wide teacher engagement during recess; system-wide change occurred due to the newly defined teacher expectations during recess. In summary, this evaluation fulfilled the purpose of a formative evaluation: data about the TI guided future directions of TI, provided structured feedback to administrators, and data-based, sustainable, and systems-level changes were accomplished.

As a result of this evaluation, the Behavior Day School now has a school specific overview of the utility of current TI assessment practices, a concrete definition for recess expectations for staff members, and a sampling of current teacher attitudes towards TI and the associated barriers of TI implementation. Through the collaborative efforts of the evaluator and school administrators within a partnership framework, future practices for TI management and improvement of current practices were also developed.

**Limitations**

The main purpose of evaluation is to make a judgment or decision about the object being evaluated (Fitzpatrick et al., 2004), and as such findings from this evaluation cannot be generalized to similar programs without replicating the evaluation methods used. Also, staff reactivity to observations and overall awareness of participation in this evaluation may have affected their performance. Although care was taken to conduct multiple observations, reactivity to observation is a factor that cannot be fully separated from the data.
Another limitation is the internal nature of this evaluation. The evaluator held the position of research intern within the school. Although she held no direct power over staff members, this evaluation is still considered an internal evaluation. Every precaution was taken to avoid coercion, through consent forms, and care in verbal interactions; however it is still possible that some staff members felt pressure or coercion to participate in this study.

Data were collected in a naturalistic setting; hence the evaluator had no control over variables not related to this evaluation affecting the results. For example, staff absences or staff injuries would often delay observations, or prompt the need for make-up meetings, as was the case when school administrators held a school wide meeting presenting the new expectations of teacher behavior during recess.

This evaluation was lengthy, time consuming, and involved three data collectors. The amount of data collected, and the amount of time spent collecting the data is not feasible in the day-to-day operation of a school. Hence, the detail, time and labor that went into this study are a limitation when providing feedback and recommendations to the school in their continued efforts to monitor TI.

**Future directions and Recommendations**

The results of this evaluation along with known limitations and sensitivity to time and labor constraints led the evaluators to make the following recommendations to school administrators.

Ongoing utilization of behavior plan quizzes and observations of teacher adherence to behavior plans were recommended in the school. In order to make the observations feasible, it was recommended that interns at the school (mostly college seniors from nearby psychology programs) be utilized to conduct the observations. Every semester, interns are prevalent throughout the Behavior Day School; prior to this study, interns would merely observe and
follow the teachers. However, the interns have the capacity to be trained observers.

Designating observation interns may be an option to continue observations. This evaluation observed teacher adherence to entire behavior plans. The observations were time consuming due to extensive details within the plans, since the behavior plans span the entire day the students are at school. It was also recommended to school administrators that instead of conducting observations on entire behavior plans that encompass a six-and-a-half hour school day, observations could be more focused (e.g., math, transition times or lunch times) and therefore less time-consuming and more problem-specific.

It was recommended that monitoring of recess times, and public postings of expectations of teacher behavior during recess be ongoing. Since recess expectations are not formalized, new teachers should also be exposed to the expectations during initial training; the expectations of teacher engagement during recess should be seen as the new norm within the school. Additional school wide staff meetings may be necessary, with role playing and clarifying the operational definition of “engaged” behavior, to make sure the knowledge is current.

Due to the complexity of TI as a large and highly multifaceted construct, it is recommended that school administrators and teachers draft a working school statement of how TI is defined within the school, including teacher expectations of TI along with school administrators’ expectations of TI. It is hoped that this will establish the foundation of consistent expectations and definitions of TI within the Behavior Day School.
APPENDIX

SURVEY

Dear Teachers:

Thank you for agreeing to participate in this survey. This survey is completely anonymous and is in regard to the Treatment Integrity Implementation of interventions in a behavior therapy setting. The survey consists of 24 questions rated on a scale from 1 to 6 (Always Disagree to Always Agree). Please place a checkmark in one block only.

The questions are meant to tap into the field of ABA and the role of Treatment Integrity procedures in the implementation of interventions. Thus, some questions are broad in scope and concern the overarching field of behavior, budgeting, and research.

If you have any questions or require clarification of individual items, please do not hesitate to ask. Once again, thank you for your participation.

Regards,
Lin Tang, M.Ed., Ph.D. (ABD)  William Matthews, PhD
Principal Investigator
Treatment fidelity is not regarded as imperative in ensuring adequate intervention control

There is an inconsistency in the terminology of the aspects of treatment integrity (e.g., behavior plan adherence, teacher competence, intervention differentiation).

The field of ABA does not agree as to what is the appropriate method of integrity assessment.

Insufficient resources due to constrained funding hinder the adequate implementation of integrity procedures.

Report of the treatment integrity procedures is not considered to enhance the credibility of the treatment outcome results.

The definition of treatment (intervention) adherence in the literature is ambiguous.
The requirements of internal review boards hinder implementation of integrity procedures (e.g., limiting how data are handled and linked to specific teachers, pushing for audio instead of videotaping).

<table>
<thead>
<tr>
<th>Always Agree 6</th>
<th>Agree 5</th>
<th>Sometime Agree 4</th>
<th>Sometime Disagree 3</th>
<th>Disagree 2</th>
<th>Always Disagree 1</th>
</tr>
</thead>
</table>

Interventions are not sufficiently manualized to permit adequate integrity implementation.

<table>
<thead>
<tr>
<th>Always Agree 6</th>
<th>Agree 5</th>
<th>Sometime Agree 4</th>
<th>Sometime Disagree 3</th>
<th>Disagree 2</th>
<th>Always Disagree 1</th>
</tr>
</thead>
</table>

Designing and validating treatment integrity measures is labor intensive and time consuming.

<table>
<thead>
<tr>
<th>Always Agree 6</th>
<th>Agree 5</th>
<th>Sometime Agree 4</th>
<th>Sometime Disagree 3</th>
<th>Disagree 2</th>
<th>Always Disagree 1</th>
</tr>
</thead>
</table>

Once established, adherence and competence (to behavior plan) are believed to be stable and not to fluctuate over time.

<table>
<thead>
<tr>
<th>Always Agree 6</th>
<th>Agree 5</th>
<th>Sometime Agree 4</th>
<th>Sometime Disagree 3</th>
<th>Disagree 2</th>
<th>Always Disagree 1</th>
</tr>
</thead>
</table>

There are no conventional criteria that specify acceptable levels of treatment integrity.

<table>
<thead>
<tr>
<th>Always Agree 6</th>
<th>Agree 5</th>
<th>Sometime Agree 4</th>
<th>Sometime Disagree 3</th>
<th>Disagree 2</th>
<th>Always Disagree 1</th>
</tr>
</thead>
</table>

The guidelines for evaluating psychometric properties (validity and reliability) of the treatment integrity measures are unclear.

| Always Agree 6 | Agree 5 | Sometime Agree 4 | Sometime Disagree 3 | Disagree 2 | Always Disagree 1 |
It is expensive and time consuming to provide direct training to teachers (e.g., viewing class time tapes, providing feedback, having regular meetings with staff, role-playing techniques).

<table>
<thead>
<tr>
<th>Agree</th>
<th>Always Agree</th>
<th>Always</th>
<th>Agree</th>
<th>Agree</th>
<th>Always</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disagree 6</td>
<td>Disagree 5</td>
<td>Sometimes Agree 4</td>
<td>Sometimes Agree 3</td>
<td>Disagree 2</td>
<td>Always Disagree 1</td>
<td></td>
</tr>
</tbody>
</table>

Teachers may resist close supervision and monitoring of treatment implementation.

There is a considerable time requirement in obtaining accurate representation of integrity data (collection of data across teachers, situations, cases and days).

Performing manipulation checks on the integrity of behavior plan implementation may be risky, as adherence and competence may be lower than desired and low competence may be lower than desired (e.g., credibility of program data may be compromised by reporting low levels of integrity).

<table>
<thead>
<tr>
<th>Agree</th>
<th>Always Agree</th>
<th>Always</th>
<th>Agree</th>
<th>Agree</th>
<th>Always</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disagree 6</td>
<td>Disagree 5</td>
<td>Sometimes Agree 4</td>
<td>Sometimes Agree 3</td>
<td>Disagree 2</td>
<td>Always Disagree 1</td>
<td></td>
</tr>
</tbody>
</table>

There are no established criteria or principles by which treatment integrity may be judged.

<table>
<thead>
<tr>
<th>Agree</th>
<th>Always Agree</th>
<th>Always</th>
<th>Agree</th>
<th>Agree</th>
<th>Always</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disagree 6</td>
<td>Disagree 5</td>
<td>Sometimes Agree 4</td>
<td>Sometimes Agree 3</td>
<td>Disagree 2</td>
<td>Always Disagree 1</td>
<td></td>
</tr>
</tbody>
</table>
High labor costs may preclude administrators from employing or training integrity raters.

Treatments (interventions in behavior plans) are presumed to be effective if significant changes on the dependent measures are obtained regardless of the integrity level of intervention implementation.

Intervention manuals are not widely employed because they are thought to limit teacher flexibility in addressing a student’s problems and tailoring of intervention to the individual needs or to address in the moment issues.

Once the training of a teacher is completed, supervision and monitoring of treatment implementation does not justify the time and labor costs

The cost of implementing integrity procedures outweighs the possible benefits.
It is generally believed that integrity procedures can be implemented primarily with behavioral interventions but not with other approaches, such as psychodynamic or interpersonal treatments.

Teacher (interventionist) competence is not clearly defined in the literature.
REFERENCES


*Journal of Applied Behavior Analysis, 30*, 693