

2013

An examination of the relationship between measurement methods of treatment integrity using the Good Behavior Game classroom intervention

Kristen Colleen O'Leary

Louisiana State University and Agricultural and Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_theses

 Part of the [Psychology Commons](#)

Recommended Citation

O'Leary, Kristen Colleen, "An examination of the relationship between measurement methods of treatment integrity using the Good Behavior Game classroom intervention" (2013). *LSU Master's Theses*. 2096.

https://digitalcommons.lsu.edu/gradschool_theses/2096

This Thesis is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Master's Theses by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.

AN EXAMINATION OF THE RELATIONSHIP BETWEEN MEASUREMENT
METHODS OF TREATMENT INTEGRITY USING
THE GOOD BEHAVIOR GAME CLASSROOM INTERVENTION

A Thesis

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Master of Arts

in

The Department of Psychology

by
Kristen C. O'Leary
B.A., University of North Texas, 2009
May 2013

Table of Contents

LIST OF TABLES.....	iii
ABSTRACT.....	iv
INTRODUCTION.....	1
METHOD	9
Participants.....	9
Measures	10
Rating of Effective Teaching – Revised	10
Direct Observation Treatment Integrity.....	10
Self-Report Treatment Integrity.....	11
Permanent Product Treatment Integrity.....	11
Intervention Rating Profile – 15	11
Materials	12
Design and Data Analysis.....	12
Procedure	12
Teacher Selection and Screening	12
Teacher and Observer Trainings.....	13
Precautions for Direct Observations	14
Intervention Implementation.....	15
RESULTS	17
DISCUSSION.....	20
REFERENCES	24
APPENDIX A: OBSERVER TREATMENT INTEGRITY	28
APPENDIX B: TEACHER TREATMENT INTEGRITY	29
APPENDIX C: GBG CLASSROOM BEHAVIOR CHART.....	30
APPENDIX D: IRB APPROVAL.....	31
VITA.....	32

List of Tables

1. Teacher Demographic Information.....	9
2. Means and Standard Deviations of Treatment Integrity and Classroom Management Measures	17
3. Fixed Effect Coefficients from Models of Treatment Integrity Measures	18
4. Fixed Effect Coefficients for Items on Observer and Teacher Treatment Integrity Forms.....	18
5. Observer-Teacher Integrity Agreement by Teacher	19

Abstract

School psychologists must often hand over the interventions they plan to be carried out by others. Collecting treatment integrity is a way to ensure that the interventions are being properly implemented according to the treatment plan. Treatment integrity can be measured in a variety of ways: direct observation, self-report, and permanent product, to name a few. However, little research has been done to see if these different methods are capturing the same information about the accuracy with which interventions are being implemented. Therefore, the purpose of this study was to collect all three types of integrity for a single intervention, the Good Behavior Game, and compare the results in order to find out if they are capturing the same information. Participants included four general education elementary school teachers. Hierarchical linear modeling was used to assess for the relationships between methods. Direct observation and teacher report methods had the highest levels of relatedness. There was almost no relationship between the permanent product and the other two methods. Item analyses were then run to see which items were driving the relationship between the observer and teacher integrity measures.

Introduction

The field of school psychology is replete with professionals who entered the occupation in order to work with children. Graves and Wright (2007) found that 92% of students in school psychology graduate programs ranked working with children as one of the major reasons they chose the profession. To someone outside of the field, it may be intuitive to think that this is exactly what school psychologists get to do: spend their days working with children. However, it has become increasingly apparent to school psychologists that their time is spent more on indirect services than in direct contact with children. Conoley and Gutkin (1986) termed this phenomenon the “paradox of school psychology.” We are better able to reach the end goal of helping children by working with adults through consultation. (Conoley & Gutkin, 1986; Gutkin & Conoley, 1990, Meyers, 1995).

Consultation consists of a triadic relationship between a consultant, consultee, and client. The school psychologist functions as the consultant in this relationship and he/she educates and trains the consultee, most likely a teacher or parent, on the intervention or services to be provided to the client, in this case a student. The benefits of consultation are twofold; psychologists can both reach more students and prepare teachers with tools that can be applied to situations in the future (Erchul & Sheridan, 2008). Measurement in consultation can be separated into three major divisions: information pertaining to participants, measurements related to the consultation process itself, and data on student outcomes. For the purpose of this study, we were interested in a particular measurement involved within the consultation process: treatment integrity (Schulte, 2008).

Treatment integrity is most often described as the extent to which an intervention is implemented as planned or intended (Gresham, 1989; McIntyre, Gresham, DiGennaro, & Reed,

2007; Roach & Elliott, 2008; Yeaton & Sechrest, 1981). This concept is known by several names depending on the field in question. In psychotherapy research, the term is treatment fidelity (McHugh, 2009; Moncher & Prinz, 1991; Schoenwald, 2011); in applied behavior analysis, it is procedural reliability (Billingsley, White, & Munson, 1980; Duda, Duchnowski, & Clarke, 2006); in medicine, it is compliance to regimens (Haynes, Taylor, and Sackett, 1979; Luftey, 2004), and in addictions research, it is program implementation (Gresham, 2009). These terms are all just elaborate ways to say that we need to link intervention to outcomes.

As treatment integrity has received increasing amounts of attention, the definition has broadened to include more dimensions. The three that are primarily considered are treatment adherence, interventionist competence, and treatment differentiation. Treatment adherence deals more with the general definition of treatment integrity; that is, it looks at whether treatment steps are followed and the treatment is implemented as planned. Interventionist competence deals with how well the interventionist delivered the treatment; it deals more with quality than number of steps completed. Program differentiation deals with how much the treatment in question differs from other treatments of a similar type (Gresham, 2009; Perepletchikova, Treat, & Kazdin, 2007; Schulte, Easton, & Parker, 2009). Dave and Schneider (1998) expanded the definition to include two more dimensions; the dimensions added were treatment exposure and participant responsiveness. Treatment exposure refers to how much of the treatment the participant received, and participant responsiveness refers to how much the participant was involved with or interested in the treatment. Of the above five dimensions of treatment integrity, it seems that treatment adherence is used most often by researchers. This is because treatment adherence lends itself more easily to use in research and data analysis; it is more easily

quantifiable than the other dimensions. As such, treatment adherence was the facet of treatment integrity measured in this study.

Gresham (1989) put forth several factors that could exert an influence on treatment integrity. Some of these factors are complexity of intervention, time necessary to implement, intervention, materials needed, and motivation of the implementer. These are all things that intuitively seem related to treatment integrity; however, more research is needed to further demonstrate these relationships. The same is true of other factors. The present study examined the relationship between classroom management skills and treatment integrity measures as well as intervention acceptability and treatment integrity.

By measuring treatment integrity, we are able to ensure that the changes seen in the dependent variable (target behavior) are related to the systematic manipulation of the independent variable (intervention) (Baer, Wolf, & Risley, 1968; Peterson, Homer, & Wonderlich, 1982; Sidman, 1960). While this idea has always been present in scientific research, linking treatments to outcomes, not all researchers embraced it in their research. One of the first groups to notice this disconnect and comment on it was Peterson and colleagues.

Peterson et al. (1982) took note of what they termed a “curious double standard,” the fact that researchers always reported information about the dependent variable but rarely reported information on the independent variable. It was assumed that if the desired outcome was reached, then the independent variable must have been put in place correctly. However, that is not the case. Peterson et al. reviewed articles in the *Journal of Applied Behavior Analysis* (JABA) from 1968-1980 in order to see how many included information on the implementation of the independent variable. What they found was not encouraging. A range of 10-50% of experiments

did not operationally define the independent variable when a definition was necessary. In addition, only 16% of articles in that time period reported data on treatment integrity.

This general line of research was continued by other researchers, namely Gresham, Gansle, and Noell (1993) and McIntyre, Gresham, DiGennaro, and Reed (2007). Gresham et al. reviewed the articles involving children in JABA from 1980-1990. This study found that not much had changed in the 10 years since the Peterson et al. (1982) article. Only 34% of the studies reported an operational definition of the independent variable, which was down from the previous study. Treatment integrity was reported in only 15% of the articles. McIntyre et al. picked up where they left off and reviewed JABA articles from 1991-2005. This review of the literature showed that 95% of articles provided operational definitions of the independent variable and 34% of articles provided information on treatment integrity. It also revealed that the average reported treatment integrity from 1968-2005 was only 21%. The results of the above three studies show that more research should be done on treatment integrity and that even though it is at the forefront of the field, awareness still needs to be raised about treatment integrity until the whole field embraces it.

Along with studies that highlighted the lack of treatment integrity being reported in scientific journals, federal legislation recently passed emphasized the importance of treatment integrity. Both the No Child Left Behind Act of 2001 and the Individuals with Disabilities Education Improvement Act of 2004 require evidence-based interventions, and thus the reporting of treatment integrity. The Institute of Education Sciences includes treatment integrity in their 2012 requirements to award research grants. These federal and monetary motivations, in addition to best practices in treatment and research, have helped shed more light on the importance of treatment integrity.

Due to the emphasis on and interest in treatment integrity over the past few decades, it has been the subject of several studies. A particular area of interest is the result of performance feedback on treatment integrity. Several studies have found that training and consultation in and of themselves are not adequate to ensure high levels of treatment integrity. After a few days of intervention, when implementation levels begin to fall, use of performance feedback can bring intervention implementation back to perfect or near perfect levels (Jones, Wickstrom, & Friman, 1997; Noell, Witt, Gilbertson, Ranier, & Freeland, 1997; Noell et al., 2005). Knowledge of when integrity levels begin to drop depends on which method is being used to collect this data. Therefore, it is important to know which method is being used and if it is the most accurate or appropriate measure for the study in question.

Treatment integrity can be measured through several methods, including but not limited to direct observations, self-report, rating scales, interviews, and permanent product (Gresham, 1997; Schulte, 2005). A study by Wickstrom, Jones, LaFleur, and Witt (1998) looked at three different methods of measuring the treatment integrity of consultation cases. A form recording occurrences and nonoccurrences of the behavior, the proximity of intervention materials to child, and direct observation were the three measures collected. Results showed that the form filled out by teachers and proximity of intervention materials yielded higher estimates of integrity than the direct observation. The present study continued in this line of research, however, the three measures being compared were be direct observation, self-report, and permanent product.

Each method used to assess treatment integrity has positive and negative aspects. Direct observation is one of the most commonly used tools by school psychologists in their work. It is considered to be more objective and accurate than other ways in which data can be gathered (Shapiro & Clemens, 2005; Volpe, DiPerna, Hintze, & Shapiro, 2005). However, it is not

without its faults. Direct observation is also costly and results in reactivity and elicits actions based mainly on social desirability factors (Kazdin, 1982). Self-reports are indirect measures of treatment integrity. These indirect methods can be faster and more cost-effective than direct observation, and they provide the school psychologist with other opinions than their own. Along with these positive attributes, self-report measures are subject to bias and socially desirable reporting. Information gathered from them should be interpreted with caution (Shapiro & Clemens, 2005). Permanent products supply the school psychologist with tangible evidence that the intervention has been implemented. It is less subject to bias and therefore may be more reliable (Albers, Elliott, Kettler, & Roach, 2005). However, it could also falsely inflate the level of treatment integrity because it does not ensure that all steps were completed, only that the end result occurred.

In the current study, these three measures of treatment integrity data were gathered on the implementation of the Good Behavior Game (GBG) class-wide intervention. The Good Behavior Game is a universal classroom intervention; it was first discussed by Barrish, Saunders, and Wolf in 1969. They proposed an intervention based on group contingencies and reinforcers natural to the classroom (ex: free time, stickers, extra recess, etc.) that would work to reduce disruptive behavior in students. The original GBG intervention was conducted in a fourth-grade class with several students who had been referred to the office for various disruptive behaviors. The class was broken into two teams, and they were given eleven rules to follow during the game that fell into two categories: out of seat behavior and talking out behavior (Barrish, Saunders, & Wolf, 1969). When a rule was broken, a mark was put on the board corresponding to the team of the offending student. In this way, the misbehavior of one student could affect the outcome for all the students on his/her team. At the end of the game period either the team with the fewest marks

or both of the teams, if they fell below a certain criterion, could earn certain privileges and rewards. Clear reductions in disruptive behavior were demonstrated in both reading and math classes (Barrish, Saunders, & Wolf, 1969).

Replications of the GBG were conducted soon after the initial study was published. In one such replication, the game was run during reading time, when the students were already split into two teams (Medland & Stachnik, 1972). Rather than the teacher acknowledging the students' misbehavior, red and green lights were used to signify rule violations and rule following, respectively. This study found that the GBG reduced disruptive behavior 99% and 97% from baseline levels for the two teams (Medland & Stachnik, 1972). The following year, another replication was done. Harris and Sherman (1973) used the GBG in two classrooms in which the teachers were seeking help with behavior management. One of the authors' primary goals with the study was to demonstrate the effectiveness of the GBG in different classrooms (A fifth and sixth-grade class were used in this study). The results showed that, indeed, the GBG reduced disruptive behaviors in those older grades as well (Medland & Stachnik, 1973).

In the 40 years or so since the initial investigations with the GBG, much more research investigating the game has been produced. It has been shown to be effective with students of all ages, ranging from preschool and kindergarten to high school (Donaldson, Vollmer, Krous, Downs, & Berard, 2011; Kleinman & Saigh, 2011; Salend, Reynolds, & Coyle, 1989; Swiezy, Matson, & Box, 2008). Studies have been conducted not only in typically developing classrooms, but also with students who are emotionally disturbed (Salend et al., 1989), in self-contained classes (Hegerle, Kesecker & Couch, 1979), and in state hospitals (Lutzker & White-Blackburn, 1979). The GBG has been shown to be effective in so many different populations and settings, Embry (2002) wrote an article suggesting it could be considered a "behavioral vaccine,"

an intervention which could be used to prevent a multitude of negative outcomes, not just disruptive behavior.

In sum, the current study compared three different methods of collecting treatment integrity in order to evaluate whether they capture the same information; the data was collected in classrooms implementing the Good Behavior Game intervention. It was hypothesized that the self-report and permanent product integrity measures would relate more highly with one another than either one would relate to the direct observation integrity measure.

Method

Participants

The participants for this study were four general education elementary teachers. Information regarding the teachers can be found in Table 1. Teachers were recruited from elementary schools in East Baton Rouge Parish and Central Community School System in Louisiana, and Prosper and Denton Independent School Districts in Texas. Administration and teachers were consulted in eight schools; however the participants in the study were from only three schools. Administrators at these schools were provided with information and flyers detailing the study and what participation in the study would entail. This information was then distributed to teachers either in their boxes or on an individual basis. Teachers who felt in need of or simply wanted more experience with and exposure to classroom management interventions then volunteered for participation in the study. Since a school psychologist working in a school cannot narrow down the teachers with whom he/she consults, inclusion and exclusion criteria were not used for this study in an attempt to maintain external validity. Teachers were screened before beginning intervention to gain an initial level of classroom management as rated on the *Rating of Effective Teaching – Revised*.

Table 1. Teacher Demographic Information

	District	Yrs. Experience	Grade Taught	Period for GBG	Days Participating in Study
Ms. Gilmore	EBR	4	3 rd	Math	10
Ms. Lane	EBR	6	3 rd	Math	10
Ms. Gellar	EBR	15	2 nd	Reading	10
Ms. Danes	PISD	1	Pre-K	Math	10

Note: EBR denotes East Baton Rouge Parish; PISD denotes Prosper Independent School District

Measures

Rating of Effective Teaching – Revised. The *Rating of Effective Teaching – Revised* (RET-R; Singletary, 2009) is a 14-item rating scale intended to provide a measure of the effectiveness of teachers' classroom management abilities; it can be used after 30 minute classroom observations. Scores on the RET-R range from 0-42, with each item consisting of a 4 point Likert scale. A score of 42 represents perfect classroom management on the part of the teacher, while a score of 0 represents little to no classroom management on the part of the teacher. Scores of 34 (80%) and above indicate desirable levels of classroom management. The RET-R has been shown to have an internal consistency of .81, found by using Cronbach's alpha, and an inter-rater reliability of .95. For use in this study, the RET-R was used to gain an idea of the teacher's classroom management abilities before beginning the intervention. It was also scored during every session of the Good Behavior Game; inter-observer agreement (IOA) data was collected on 20% of the sessions. The average IOA on the RET-R was 93.75% with a range from 89-100%.

Direct Observation Treatment Integrity. In this study, the systematic direct observation consisted of a trained observer from LSU conducting an observation of an entire session of the Good Behavior Game. During this observation, the observer attended to the teacher and her actions as they pertained to the running of the intervention. The observer then filled out the 8-item treatment integrity form after the session reflecting what had been observed. An observer was present for every session in order to complete this form. A sample of this form is attached as Appendix A. IOA data was collected on 20% of the sessions. The average IOA on the observer integrity form was 91.13% with a range from 86-97%.

Self-Report Treatment Integrity. In this study, the self-report measure of treatment integrity consisted of the classroom teacher completing a treatment integrity form at the conclusion of each session of the Good Behavior Game based on the implementation of the intervention that day. The form consisted of the same 8 items present on the direct observation treatment integrity form. This was to allow for more accurate comparison of the two measures of integrity. A sample of this form is attached as Appendix B.

Permanent Product Treatment Integrity. The use of permanent product as a measure of treatment integrity requires that something tangible is collected, to physically show that the intervention was implemented, rather than simply having a form that someone filled out. In this study, two pieces of information were collected to serve as permanent products. These were the paper behavior chart on which the teacher tracked rule violations and a picture of the white board, where the teacher was tracking rule violations for the students to see. A sample of the behavior charts used by teachers in this study is attached as Appendix C.

Intervention Rating Profile – 15. The *Intervention Rating Profile* (IRP-15; Martens, Witt, Elliott, & Darveaux, 1985) is a measure of intervention acceptability to be completed by the implementer of an intervention, in this case, the teacher. It consists of 15 items, each of which is answered on a 6-point Likert scale ranging from 1, indicating *Strongly Disagree*, to 6, indicating *Strongly Agree* (Witt & Elliott, 1985). The IRP-15 was given both pre- and post-intervention to assess the acceptability of the GBG for the teachers and the level of integrity with which it is implemented. Reliability of the IRP-15 is very strong, with a Cronbach's alpha of .98 (Martens, Witt, Elliott, & Darveaux, 1985).

Materials

The materials for this study were provided by the primary researcher. Each teacher was given a poster with the rules for the GBG as it would be conducted in her classroom. Treatment integrity forms were provided to both the teachers and the observers for each session. Cameras or camera phones were necessary to document one aspect of the permanent product, the check marks recorded on the white board. If an observer did not have access to a camera, the researcher provided one to them. Prize boxes were provided in each classroom, out of which students from the winning team to choose. Prizes included things such as pencils, erasers, sharpeners, candy, suckers, small toys, etc. Teachers also had the option of giving the students intangible rewards (e.g.: free time, extra recess).

Design and Data Analysis

The current study was conducted to compare three different measurement methods for the treatment integrity of the Good Behavior Game class-wide intervention. In order to do so, data was analyzed using hierarchical linear models (HLM). Use of HLM allowed for accurate results to be obtained because it can account for the autocorrelations between data points collected on the same teachers (Raudenbush & Bryk, 2002). A total of 40 data points were included in the analyses for each measure, with ten data points each from the four teachers participating. HLM was also run with the treatment integrity measures and the RET-R to see if a significant relationship existed between classroom management skills and level of intervention implementation.

Procedure

Teacher Selection and Screening. In order to find participants for the study, the primary researcher first contacted the administration at school in Southeast Louisiana and North Texas. In

those schools amenable to participating in research, administrators were asked to nominate teachers who would either benefit from or be interested in learning about behavioral interventions and classroom management. Administration in one school chose to nominate teachers. In the other two schools where research was conducted, the administrators preferred that the teachers be contacted via flyers in their boxes; teachers could then choose to contact the primary researcher if they were interested in participating. Recruiting participants in this manner allowed for strengthened relationships with administration and provided information and training to teachers who wanted and needed it. It also added to the external validity and generalizability of the study because these were precisely the types of teachers that school psychologists work with on a daily basis. Teachers that consented to participation in the study were then observed before training. The RET-R was scored for each teacher before the intervention began in order to get an idea of the teacher's current level of classroom management.

Teacher and Observer Trainings. Training sessions were held separately for teachers and observers. Teacher trainings were held in the classroom to both make it more convenient for the participants and to allow for demonstrations in the environment where the intervention was to be run. This also allowed the researcher and teacher to make decisions about where materials such as the rules, prize box, and team points would be best displayed. Teacher trainings lasted about 30-45 minutes, in which the primary researcher explained both the reasons behind the intervention and how it was to be implemented. The researcher explained and modeled aspects of the intervention and allowed time for the participants to ask questions and practice. All steps to the GBG intervention were thoroughly explained and the teacher integrity form (Appendix B) was frequently referenced during training. Also during the training, the teacher was asked to choose reinforcers which they would like to be used as the rewards for the winning team each

day. A reinforcer menu was provided and teachers were allowed to offer other potential rewards they thought would motivate the students. Teachers generally reported adequate understanding of the intervention and had few questions regarding their duties in implementation and filling out required forms.

Observer trainings took place on campus at the university with the primary researcher conducting the trainings. For the observer in PISD, the training took place at the elementary school. Observers consisted of graduate and undergraduate students in a research lab; in PISD the additional observer was an aide from the school. The researcher explained the study and the purpose behind it. The GBG intervention itself was reviewed and then the observers' responsibilities with regards to treatment integrity collection and RET-R ratings. The direct observation treatment integrity form was to be filled out at every session, as was the RET-R form. At the conclusion of the intervention for the day, the observers were to collect the self-report integrity form from the teacher and get the necessary information for permanent product integrity. Time was allowed for questions; these trainings also lasted approximately 30-45 minutes each.

Precautions for Direct Observation. In this study, direct observation by the primary researcher and other observers was used as the gold standard form of treatment integrity. In order to do this, certain measures were taken to address the primary concerns with direct observations. One concern is that there is sampling bias. Direct observations are normally taken on very small segments of intervention implementation. To combat this concern, in this study, direct observations were taken on the entire intervention session, for every session. Another concern with direct observation is reactivity on the part of the person being observed. Although there is no way to completely do away with this issue, steps were taken in this study to protect against

reactivity. The observer was in the classroom with the teacher before the intervention began; this allowed for the teacher to become somewhat acclimated to having observers in the classroom before the treatment integrity ratings and intervention were started. Also, the primary researcher was the one who conducted the teacher trainings, however she was not always the one taking observations. The intent of this precaution was that the teachers would be less likely to associate the observers with the expectations presented by the primary researcher.

Intervention Implementation. Before the intervention began, the teachers divided the class into teams. Ms. Gilmore and Ms. Lane formed 3 teams in their classrooms, while Ms. Gellar and Ms. Danes opted for four teams in their classrooms. At the beginning of the Good Behavior Game intervention, the teacher got the class' attention and indicated to them that it was time for the GBG to begin. The teacher then reviewed the rules for the game, which were as follows: 1) Raise your hand and wait for permission to speak; 2) Follow teacher's instructions within 15 seconds; 3) Stay on task; and 4) Use only kind words. After the first day or two of the intervention, the teachers began asking the students to raise their hands and go over the rules themselves. Next, the teacher reviewed the number of points earned by each team the previous day. These points were left on the board from the previous day for ease of review and to provide the students a visual representation of their previous performance. The points were then erased so the game could start anew for the allotted period. Each session of the GBG lasted for an hour during the class period the teacher deemed the most disruptive. These class periods are listed in Table 1.

After the rules and points were reviewed, the teacher began class instruction as it would normally proceed during that time period. When a child broke one of the rules, the teacher acknowledged the student and the rule which was broken. Then she would state which team got a

check mark for that rule violation. For example, the teacher might say, “Johnny, you didn’t raise your hand before answering that question. Team Two gets a check.” After verbally acknowledging the rule violation, the teacher had to track it in two ways. She would mark it under the corresponding team number on the white board, in order for the children to see how many points their team had at all times. She also had to track the mark on the paper behavior tracker under the corresponding team number and rule that was broken. Throughout the time the intervention was being implemented, the teachers were to give as much positive praise for rule-following as possible.

When the hour allotted for the GBG came to a close, the teacher got the students’ attention and let them know that it was time for the game to end for the day. The teacher reviewed the number of check marks earned by each team for the day. The team with the fewest number of checks earned a prize. If two or more teams were tied for the fewest number of points, all of them received a prize. The teachers were to provide the prizes at the end of the intervention period, not at the end of the school day. After the conclusion of the intervention, the teacher filled out the self-report treatment integrity form.

Throughout the intervention, the observer was present and watched the teacher implement the intervention. During the hour allotted for the intervention, the observer filled out both the direct observation treatment integrity form and the RET-R. At the conclusion of the intervention session, the observer collected the behavior tracker and treatment integrity form from the teacher and took a picture of the white board where team points had been posted. This ensured that at the end of every session, all three measures of treatment integrity had been collected.

Results

The primary analyses were run to assess the relationships between the three measures of treatment integrity with one another and with the RET-R. First, descriptive statistics were run on the main outcome measures; these results can be found in Table 2.

Table 2. Means and Standard Deviations of Treatment Integrity and Classroom Management Measures

	Mean	SD
Observer Integrity	82.65	10.40
Teacher Integrity	90.63	7.56
Permanent Product	96.25	13.34
RET-R	81.80	7.42

Then separate models were run to assess each relationship. Results from these analyses can be found in Table 3. Three of the six possible relationships between measures were found to be significant. The first was the observer integrity and teacher integrity relationship. This model produced a beta coefficient of 0.64, $p < .001$, indicating a strong relationship. The second significant relationship was the observer integrity and permanent product relationship. This model produced a beta coefficient of 0.04, $p = .001$, indicating a weak relationship. The last significant relationship found in the initial analyses was the observer integrity and RET-R relationship. This model produced the largest beta coefficient, 0.89, $p < .001$, indicating a very strong relationship.

Due to the fact that the observer and teacher integrity measures had the strongest relationship (among the treatment integrity measures), further analyses were run. All items on those two measures corresponded with one another, as can be seen in Appendix A and B. The relationship between each of the 8 items was then investigated. Results are summarized in Table 4. Four of the items produced significant relationships; these were items 2, 3, 6, and 7. Item 3,

with a beta coefficient of 0.81, and item 7, with a beta coefficient of 0.78, showed the strongest relationships; their reliability estimates were 0.92 and 0.83 respectively.

Table 3. Fixed Effect Coefficients from Models of Treatment Integrity Measures

	Observer Integrity	Teacher Integrity	Permanent Product	RET-R
Observer Integrity	-	-	-	-
Teacher Integrity	0.64 (0.12)*	-	-	-
Permanent Product	0.04 (0.10)*	-0.06 (0.03)	-	-
RET-R	0.89 (0.11)*	0.03 (0.12)	-0.03 (0.15)	-

Note: Standard errors in parentheses.

* $p < .001$

Table 4. Fixed Effect Coefficients for Items on Observer and Teacher Treatment Integrity Forms

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8
Coefficient	-0.16	0.20*	0.81**	0.24	0.27	-0.16*	0.78**	-0.06
Standard Error	0.50	0.09	0.04	0.13	0.15	0.07	0.21	0.05

* $p < .05$. ** $p < .001$

Other analyses were investigated as well. Agreement was measured between the observer and teacher integrity forms. The average agreement on these measures was 89.13%, with a range from 65-100%. Agreement and ranges varied by teacher. These results can be found in Table 5. The IRP-15 was scored both pre- and post-intervention. The average pre-intervention score on the IRP-15 was 91.5% (range = 83-100%), and the average post-intervention score was 99% (range = 97-100%).

Table 5. Observer-Teacher Integrity Agreement by Teacher

	Average Percent Agreement	Range
Ms. Gilmore	92.9	84 – 97%
Ms. Lane	82.4	65 – 96%
Ms. Gellar	85.5	68 – 100%
Ms. Danes	95.7	85 – 100%
Total	89.13	65 – 100%

Discussion

The aim of the study was to investigate the relationship between different ways of measuring treatment integrity and compare them to direct observation (i.e.; observer integrity), which is being used as the gold standard in this study. Results from HLM analyses indicated that there was a significant relationship between the observer integrity ratings and all other outcome measures (teacher integrity, permanent product, and the RET-R). When these results were examined more closely it appeared that although the observer-permanent product relationship was statistically significant, it did not indicate a very strong relationship. A beta coefficient of 0.04 represents almost no relationship at all. However, the other relationships resulting in statistical significance fared better. The observer-teacher integrity beta coefficient indicated a strong relationship, while the observer-RET-R beta coefficient indicated a very strong relationship. All other models for the relationships between outcome measures were found to be non-significant.

One potential reason for the lack of relationship between the permanent product and the other measures is that it consisted of something fundamentally different. While the other three measures were rating scales and had several items, the permanent product had only two items contributing to it. These items were dichotomous; either the teacher did them and gave them to the observer or she did not. Perhaps if the permanent product could be measured in a different way or used with a different intervention, an academic intervention for example, the outcome would be different.

The relationship between the observer integrity form and the RET-R seems more intuitive. Firstly, these ratings were both completed by the same observer. This observer was making both ratings off of the same intervention session, so it does not come as a surprise that

these ratings should be related. Secondly, the two measures had several items that tapped the same constructs. For example, certain aspects of the GBG, such as praising positive behavior and tracking rule violations, are also just general best practice when it comes to classroom management. These overlapping items also contributed to the strength of the observer integrity – RET-R relationship.

The most interesting relationship found in the primary analyses was that between the observer and teacher integrity measures. The model showed a strong relationship with very high reliability and model fit. This does not go along with previous research, however. Typically, teachers and other raters (parent, researcher, etc.) do not show very strong relationships. Further analyses were done to investigate this finding.

Models were run between the individual items on the observer and teacher integrity forms, since they consisted of the same items. While four of the models resulted in significant relationships, two of these (item 2 and item 6) had very small beta coefficients; as such, they could not help explain much of the strong relationship between the two integrity measures. On the other hand, two of the items did result in very strong relationships; these were items 3 and 7. Item 3 asked about reviewing points from the previous day. In most cases teachers either performed this item perfectly or not at all. Perhaps that made it easier for teachers to rate themselves accurately since it did not have as much to do with rating the degree of performance, just the presence or absence. This explanation can account for the strong relationship found in item 7 as well. Item 7 asked whether points were recorded on the behavior tracking sheet. Teachers also generally performed this step with perfect integrity or not at all. If a teacher had the tracker with her at all times, she remembered to record the points. If she left the tracker on her desk, however, points were not recorded at all until it was time to turn the form in to the

observer. This indicates that the relationship between integrity measures is stronger if items are more black and white in terms of performance.

Agreement was measured between the teacher and observer integrity forms. The average agreement was 89.13% between these raters. IOA data was also taken between observers during some sessions. The agreement between outside observers had an average of 91.13%. Although a direct comparison cannot be made because the sample sizes differ so much (40 vs. 8), there does not seem to be a difference in agreement between two observers and an observer and a teacher. The difference was found only in the range of agreement; this range was much wider with the observer-teacher agreement than it was with two independent observers. This data can be seen in Table 5.

While IRP-15 ratings were taken both pre- and post-intervention, they were not analyzed with regard to level of integrity or classroom management. The acceptability ratings were high across all teachers both before and after the intervention was run, so it would add little information to the study. Perhaps a more involved or difficult intervention with lower acceptability could be used to assess this potential relationship.

This study had several limitations. Although significant relationships were found in the analyses, the sample size and number of teachers included were still fairly small. More differences or relationships may become apparent if there were more data to analyze. Another limitation comes in the form of teacher reactivity. Precautions were taken to guard against this reactivity, but there was something that could not be helped. The sheer presence of an observer in the room was a signal to implement the intervention, regardless of actual level of integrity. If an observer had not been present, there may have been days when the intervention was not run at all. The composition of the integrity measures constitutes another limitation in the study. Rating

scales, observation forms, and permanent products can be designed in endless ways. The composition of the measures was idiosyncratic to this study. Different results may have been found if the measures were designed differently.

Although there were limitations with this study in particular, the topic of integrity measurement merits further research. It has direct implications on how we should measure treatment integrity both most accurately and feasibly in schools. Future studies should investigate other methods of measuring integrity, as only three were used in this study. Future research should also use different interventions, both behavioral and academic, to see if the results are the same in different situations. More teacher variable could be collected as well and included in the models, since HLM has the ability to accurately analyze nested data that we so often come across in school research. If more research is done in this area, it could only prove beneficial for the practice of school psychology.

References

- Albers, C.A., Elliott, S.N., Kettler, R.J., & Roach, A.T. (2005). Evaluating intervention outcomes. In R. Brown-Chidsey (Ed.), *Assessment for intervention* (pp175-199). New York, NY: Guilford Press.
- Barrish, H. H., Saunders, M., & Wolf, M. M. (1969). Good behavior game: Effects of individual contingencies for group consequences on disruptive behavior in a classroom. *Journal of Applied Behavior Analysis*, 2, 119-124.
- Billingsley, F., White, O.R., & Munson, R. (1980). Procedural reliability: A rationale and an example. *Behavioral Assessment*, 2, 229-241.
- Conoley, J.C., & Gutkin, T.B. (1986). School psychology: A reconceptualization of service delivery realities. In S.N. Elliott & J.C. Witt (Eds.), *The delivery of psychological services in schools: Concepts, processes, and issues* (pp. 393-424). Hillsdale, NJ: Erlbaum.
- Dane, A.V., & Schneider, B.H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, 18, 23-4.
- Donaldson, J. M., Vollmer, T. R., Krous, T., Downs, S., & Berard, K.P. (2011). An evaluation of the good behavior game in kindergarten classrooms. *Journal of Applied Behavior Analysis*, 44, 605-609.
- Duda, M., Duchnowski, A., & Clarke, S. (2006). Treatment integrity within applied research settings [PowerPoint slides]. Retrieved from http://rtckids.fmhi.usf.edu/presentations/TreatmentIntegrity_APBS_0306.pdf
- Embry, D. D. (2002). The good behavior game: A best practice candidate as a universal behavioral vaccine. *Clinical Child and Family Psychology Review*, 5, 273-297.
- Erchul, W.E., & Sheridan, S.M. (2008). Overview: The state of scientific research in school consultation. In W.E. Erchul & S.M. Sheridan (Eds.), *Handbook of research in school consultation* (pp. 3-12). New York, NY: Lawrence Erlbaum Associates.
- Graves, S.L. & Wright, L.B. (2007). Comparison of individual factors in school psychology graduate students: Why do students pursue a degree in school psychology? *Psychology in the Schools*, 44(8), 865-872.
- Gresham, F.M. (1989). Assessment of treatment integrity in school consultation and prereferral intervention. *School Psychology Review*, 18(1), 37-50.

- Gresham, F.M. (1997). Treatment integrity in single-subject research. In R. Franklin, D. Allison, & B. Gorman (Eds.), *Design and analysis of single-case research* (pp. 93-117). Mahwah, NJ: Erlbaum.
- Gresham, F.M. (2009) Evolution of the treatment integrity concept: Current status and future directions. *School Psychology Review*, 38(4). 533-540.
- Gresham, F.M., Gansle, K.A., & Noell, G.H. (1993). Treatment integrity in applied behavior analysis with children. *Journal of Applied Behavior Analysis*, 26(2), 257-263.
- Gutkin, T.B., & Conoley, J.C. (1990). Reconceptualizing school psychology from a service delivery perspective: Implications for practice, training, and research. *Journal of School Psychology*, 28, 203-223.
- Hagermoser Sanetti, L., & Kratochwill, T.R. (2005) Treatment integrity assessment within a problem-solving model. In R. Brown-Chidsey (Ed.), *Assessment for intervention* (pp 304-328). New York, NY: Guilford Press.
- Harris, V.W., & Sherman, J.A. (1973). Use and analysis of the “good behavior game” to reduce disruptive classroom behavior. *Journal of Applied Behavior Analysis*, 6, 405-417.
- Haynes, R.B., Taylor, W., Sackett, D.L., Gibson, E.S., Bernholz, C.D., & Mukherjee, J. (1980). Can simple clinical measurements detect patient noncompliance? *Hypertension*, 2, 757-764.
- Hegerle, D.R., Kesecker, M P., & Couch, J.V. (1979). A behavior game for the reduction of inappropriate classroom behaviors. *School Psychology Digest*, 8, 339-343.
- Individuals with Disabilities Education Improvement Act, H.R. 1350, 108th Cong. (2004).
- Institute for Education Sciences (2012). Requests for Applications: Education Research Grants. Retrieved from http://ies.ed.gov/funding/pdf/2012_84305A.pdf.
- Jones, K.M., Wickstrom, K.F., & Friman, P.C. (1997). The effects of observational feedback on treatment integrity in school-based behavioral consultation. *School Psychology Quarterly*, 12(4), 316-326.
- Kazdin, A.E. (1982). Observer effects: Reactivity of direct observation. *New Directions for Methodology of Social and Behavioral Sciences*, 14, 5-19.
- Klienman, K.E., & Saigh, P.A. (2011). The effects of the good behavior game on the conduct of regular education New York City high school students. *Behavior Modification*, 35, 95-105.
- Lutfey, K. (2004). Assessment, objectivity, and interaction: The case of patient compliance with medical treatment regimens. *Social Psychology Quarterly*, 67(4), 343-368.

- Lutzker, J. R., & White-Blackburn, G. (1979). The good productivity game: Increasing work performance in a rehabilitation setting. *Journal of Applied Behavior Analysis*, 12, 488.
- Martens, B.K., Witt, J.C., Elliott, S.N., & Darveaux, D.X. (1985). Teacher judgments concerning the acceptability of school-based interventions. *Professional Psychology*, 16, 191-198.
- McHugh, R.K., Murray, H.W., & Barlow, D.H. (2009). Balancing fidelity and adaptation in the dissemination of empirically-supported treatments: The promise of transdiagnostic interventions. *Behaviour Research and Therapy*, 47(11), 946-953.
- McIntyre, L.L., Gresham, F.M., DiGennaro, F.D., & Reed, D.D. (2007). Treatment integrity of school-based interventions with children in the journal of applied behavior analysis 1991-2005. *Journal of Applied Behavior Analysis*, 40(4), 659-672.
- Medland, M.B., & Stachnik, T.J. (1972). Good behavior game: A replication and systematic analysis. *Journal of Applied Behavior Analysis*, 5, 45-51.
- Meyers, J. (1995). A consultation model for school psychological services. *Journal of Educational and Psychological Consultation*, 6(1), 59-71.
- Moncher, F.J. & Prinz, R.J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review*, 11, 247-266.
- No Child Left Behind (NCLB) Act of 2001, 20 U.S.C.A. § 6301 *et seq.*
- Noell, G.H., Witt, J.C., Gilbertson, D.N., Ranier, D.D., & Freeland, J.T. (1997). Increasing teacher intervention implementation in general education settings through consultation and performance feedback. *School Psychology Quarterly*, 12(1), 77-88.
- Noell, G.H., Witt, J.C., Slider, N.J., Connell, J.E., Gatti, S.L., Williams, K.L., Koenig, J.L., Resetar, J.L., & Duhon, G.J. (2005). Treatment implementation following behavioral consultation in schools: A comparison of three follow-up strategies. *School Psychology Review*, 34(1), 87-106.
- Perepletchikova, F., Treat, T.A., & Kazdin, A.E. (2007). Treatment integrity in psychotherapy research: Analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology*, 75(6), 829-841.
- Peterson, L., Homer, A.L., & Wonderlich, S.A. (1982). The integrity of independent variables in behavior analysis. *Journal of Applied Behavior Analysis*, 15(4), 477-492.
- Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage Publications, Inc.

- Roach, A.T. & Elliott, S.N. (2008). Best practices in facilitating and evaluating intervention integrity. In Thomas, A. & Grimes, J. (Eds.), *Best practices in school psychology V* (pp 195-208). Bethesda, MD: National Association of School Psychologists.
- Salend, S.J., Reynolds, C.J., & Coyle, E.M. (1989). Individualizing the good behavior game across type and frequency of behavior with emotionally disturbed adolescents. *Behavior Modification, 13*, 108-126.
- Schoenwald, S.K. (2011) It's a bird, it's a plane, it's... fidelity measurement in the real world. *Clinical Psychology: Science and Practice, 18*(2), 142-147.
- Schulte, A.C. (2008). Measurement in school consultation research. In W.E. Erchul & S.M. Sheridan (Eds.), *Handbook of research in school consultation* (pp. 33-61). New York, NY: Lawrence Erlbaum Associates.
- Schulte, A.C., Easton, J.E., & Parker, J. (2009). Advances in treatment integrity research: Multidisciplinary perspectives on the conceptualization, measurement, and enhancement of treatment integrity. *School Psychology Review, 38*(4), 460-475.
- Shapiro, E.S., & Clemens, N.H. (2005). Conducting systematic direct classroom observations to define school-related problems. In R. Brown-Chidsey (Ed.), *Assessment for intervention* (pp175-199). New York, NY: Guilford Press.
- Sidman, M. (1960). *Tactics of scientific research*. New York: Basic Books.
- Singletary, L. (2009). *Rating of Effective Teaching – Revised*. Baton Rouge, LA.
- Swiezy, N.B., Matson, J.L., & Box, P. (2008). The good behavior game: A token reinforcement system for preschoolers. *Child & Family Behavior Therapy, 14*, 21-32.
- Volpe, R.J., DiPerna, J.C., Hintze, J.M., & Shapiro, E.S. (2005). Observing students in classroom settings: A review of seven coding schemes. *School Psychology Review, 34*(4), 454-474.
- Wickstrom, K.F., Jones, K.M., LaFleur, L.H., & Witt, J.C. (1998). An analysis of treatment integrity in school-based behavioral consultation. *School Psychology Quarterly, 13*(2), 141-154.
- Witt, J. C., & Elliott, S. N. (1985). Acceptability of classroom management procedures. In T. R. Kratochwill (Ed.), *Advances in school psychology*. Hillsdale, NJ: Erlbaum.
- Yeaton, W.H. & Sechrest, L. (1981). Critical dimensions in choice and maintenance of successful treatments: Strength, integrity, and effectiveness. *Journal of Consulting and Clinical Psychology, 49*(2), 156-167.

Appendix A: Observer Treatment Integrity

Observer Treatment Integrity

Teacher _____

Date _____

	(Not at all)			(Perfectly)	
1. Told the children that GBG was going to begin.	0	1	2	3	4
2. Reviewed the rules with the class.	0	1	2	3	4
3. Reviewed the points from the previous day.	0	1	2	3	4
4. Praised positive behavior.	0	1	2	3	4
5. Pointed out and corrected rule infractions.	0	1	2	3	4
6. Tracked marks on board.	0	1	2	3	4
7. Tracked marks on behavior chart.	0	1	2	3	4
8. Provided winning team with their reward.	0	1	2	3	4

(Number of Points / Total Possible (32))

Total: _____ / 32

Appendix B: Teacher Treatment Integrity

Teacher Treatment Integrity

Name _____

Date _____

	(Not at all)			(Perfectly)	
1. I told the children that GBG was going to begin.	0	1	2	3	4
2. I reviewed the rules with the class.	0	1	2	3	4
3. I reviewed the points from the previous day.	0	1	2	3	4
4. I praised positive behavior.	0	1	2	3	4
5. I pointed out and corrected rule infractions	0	1	2	3	4
6. I tracked marks on the board.	0	1	2	3	4
7. I tracked marks on the behavior chart.	0	1	2	3	4
8. I provided the winning team with their reward.	0	1	2	3	4

(Number of Points / Total Possible (32))

Total: _____ / 32

Appendix C: GBG Classroom Behavior Chart

GBG Classroom Behavior Chart

Date _____

Behavior	Team 1	Team 2	Team 3	Team 4
Talking Out				
Not Following Instructions				
Off-Task				
No Unkind Comments				

Appendix D: IRB Approval

Application for Exemption from Institutional Oversight

Unless qualified as meeting the specific criteria for exemption from Institutional Review Board (IRB) oversight, ALL LSU research projects using living humans as subjects, or samples, or data obtained from humans, directly or indirectly, with or without their consent, must be approved or exempted in advance by the LSU IRB. This Form helps the PI determine if a project may be exempted, and is used to request an exemption.

Institutional Review Board
 Dr. Robert Mathews, Chair
 131 David Boyd Hall
 Baton Rouge, LA 70803
 P: 225.578.8692
 F: 225.578.6792
 irb@lsu.edu
 lsu.edu/irb

– Applicant, Please fill out the application in its entirety and include the completed application as well as parts A-E, listed below, when submitting to the IRB. Once the application is completed, please submit two copies of the completed application to the IRB Office or to a member of the Human Subjects Screening Committee. Members of this committee can be found at

- A Complete Application Includes All of the Following:
 - (A) Two copies of this completed form and two copies of part B thru E.
 - (B) A brief project description (adequate to evaluate risks to subjects and to explain your responses to Parts 1&2)
 - (C) Copies of all instruments to be used.
 - *If this proposal is part of a grant proposal, include a copy of the proposal and all recruitment material.
 - (D) The consent form that you will use in the study (see part 3 for more information.)
 - (E) Certificate of Completion of Human Subjects Protection Training for all personnel involved in the project, including students who are involved with testing or handling data, unless already on file with the IRB. Training link: ()
 - (F) IRB Security of Data Agreement: ()

1) Principal Investigator: Rank:
 Dept: Phi: E-mail:

2) Co Investigator(s): please include department, rank, phone, and e-mail for each

IRB # ES867 LSU Proposal # _____
 Complete Application
 Human Subjects Training

3) Project Title:

Study Exempted By:
 Dr. Robert C. Mathews, Chairman
 Institutional Review Board
 Louisiana State University
 203 B-1 David Boyd Hall
 225-578-8692 | www.lsu.edu/irb
 Exemption Expires: 2/21/2015

4) Proposal? (yes or no) If Yes, LSU Proposal Number

Also, if YES, either This application completely matches the scope of work in the grant
 OR More IRB Applications will be filed later

5) Subject pool (e.g. Psychology students)
 *Circle any "vulnerable populations" to be used: (children <18; the mentally impaired, pregnant women, the aged, etc.). Projects with incarcerated persons cannot be exempted.
 6) PI Signature Date (no per signatures)

** I certify my responses are accurate and complete. If the project scope or design is later changes, I will resubmit for review. I will obtain written approval from the Authorized Representative of all non-LSU institutions in which the study is conducted. I also understand that it is my responsibility to maintain copies of all consent forms at LSU for three years after completion of the study. If I leave LSU before that time the consent forms should be preserved in the Departmental Office.

Screening Committee Action: Exempted Not Exempted Category/Paragraph 1
 Reviewer Mathews Signature [Signature] Date 2/22/12

Vita

Kristen Colleen O’Leary was born in July of 1986 in Denton, Texas. She was raised in Denton, Texas, and subsequently attended college there at the University of North Texas. She received her Bachelor of Arts in psychology with a minor in English in 2009. In 2010, Kristen moved to Baton Rouge, Louisiana where she began her graduate career in the school psychology doctoral program at Louisiana State University. Kristen’s research interests include the use of empirically based classroom management strategies to reduce disruptive behavior in children.