6-2018

The Effects of Peer Observation on Teacher Intervention Integrity

Garrett D. Warrilow

Western Michigan University, garrett.d.warrilow@gmail.com

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations

Part of the Experimental Analysis of Behavior Commons, and the Teacher Education and Professional Development Commons

Recommended Citation

https://scholarworks.wmich.edu/dissertations/3304

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact wmu-scholarworks@wmich.edu.
The observer effect is a term used to describe a process during which an observer's behavior changes as they measure the behavior of another individual. The observer effect has been widely studied in organizational behavior management but has limited research with teachers in schools. The current study sought to use the observer effect as part of a teacher training package by evaluating the effects of peer observations on an observing teacher's implementation integrity of components of a Positive Behavioral Interventions and Supports (PBIS) system. Participants were four middle school teachers of students in general education classrooms. The primary dependent variables were the number of praise and corrective statements made by the teachers to their students, and how closely the teacher approximated a ratio of three praise statements to every one corrective statement. Secondary dependent variables included the number of behavioral expectations set by the teacher for the students, and the number of office referrals written by the teacher. During baseline, teachers were asked to teach as usual while being observed three to four times per week. During the intervention, teachers completed weekly observations of their peers and recorded the number of praise and corrective statements as well as the number and type of behavioral expectations set by their peer teacher. Following the intervention, in a return to baseline, teachers did not conduct peer observations but taught in their usual manner while the experimenter conducted weekly observations. Results suggest that peer observations increased
intervention integrity of target classroom management behaviors for three of four participants. That is, three of the four participants increased their ratio of praise to corrective statements and expectation setting during the peer observation phase of the study. Additionally, based on social validity data, participants were highly satisfied with the observation process. Implications of these findings for teacher training in schools are discussed as well as how the findings of this study relate to existing research on the observer effect.
ACKNOWLEDGMENTS

Thanks must be given to the environment which shaped me into the person I am today. My family (especially Dennis Warrilow and Brynn Warrilow) created the foundation upon which my success has been built. Special thanks must be given to Brandon Schools, The Oakland Science Mathematics and Technology Academy (OSMTech), Western Michigan University, Dave Green, Kelly Sawyer, Alyce Dickinson, Doug Johnson, and Denise Ross for their influence on both my academic and professional growth. Additionally, I could not have completed this project without my committee members Amy Damashek, Ron VanHouten, and Ronnie Detrich. Ronnie deserves additional credit for providing incredible mentorship and countless hours of advising beginning before the topic of this dissertation was decided, and continuing well after it was completed. My friends who have supported me in many ways over many years also deserve thanks. In particular, Kelley Howard, Mike Thiel, Luke Butler, Evan Maltas, Nick Smith, and Ian Wagner, deserve mention. Finally, my significant other, Brianna Forbis, must be credited for her incredible patience and support through the completion of my Ph.D. and this dissertation.

Garrett D. Warrilow
TABLE OF CONTENTS

ACKNOWLEDGMENTS........................................................................................................... ii

LIST OF TABLES..................................................................................................................... vi

LIST OF FIGURES................................................................................................................ vii

INTRODUCTION.................................................................................................................... 1

The Effects of Varied Integrity on Outcomes of Interest......................................................... 2

Variables Affecting Teacher’s Intervention Integrity............................................................. 7

Intervention Acceptability and Training’s Effects on Intervention Integrity.......................... 7

Feedback’s Effects on Intervention Integrity........................................................................ 9

Behavioral Pinpoints and Peer Observation For Performance Management......................... 12

Peer Observation and the Observer Effect............................................................................ 14

Related Research................................................................................................................... 20

Purpose of The Current Study............................................................................................... 22

METHODS.............................................................................................................................. 24

Participants, Participant Selection and Assignment............................................................... 24

Setting..................................................................................................................................... 25

Dependent Variables............................................................................................................ 26

Independent Variables......................................................................................................... 29

Data Collection..................................................................................................................... 30

Experimental Design............................................................................................................ 32

Procedures........................................................................................................................... 33
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention Integrity</td>
<td>37</td>
</tr>
<tr>
<td>RESULTS</td>
<td>38</td>
</tr>
<tr>
<td>Praise to Correction Ratio</td>
<td>38</td>
</tr>
<tr>
<td>Expectation Setting</td>
<td>45</td>
</tr>
<tr>
<td>Referrals</td>
<td>48</td>
</tr>
<tr>
<td>Interobserver Agreement</td>
<td>51</td>
</tr>
<tr>
<td>Teacher Observation Accuracy</td>
<td>51</td>
</tr>
<tr>
<td>Intervention Integrity</td>
<td>51</td>
</tr>
<tr>
<td>Intervention Acceptability</td>
<td>52</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>53</td>
</tr>
<tr>
<td>Major Findings</td>
<td>53</td>
</tr>
<tr>
<td>The Relation Between the Current Findings and Previous Research</td>
<td>55</td>
</tr>
<tr>
<td>The Implications of this Study on Schools and Teacher Training</td>
<td>57</td>
</tr>
<tr>
<td>Limitations</td>
<td>60</td>
</tr>
<tr>
<td>Future Research</td>
<td>62</td>
</tr>
<tr>
<td>CONCLUSION</td>
<td>65</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>66</td>
</tr>
<tr>
<td>APPENDICES</td>
<td></td>
</tr>
<tr>
<td>A. Informed Consent Document</td>
<td>72</td>
</tr>
<tr>
<td>B. Behaviors Warranting a Referral</td>
<td>75</td>
</tr>
<tr>
<td>C. Teacher Observation Data Sheet</td>
<td>77</td>
</tr>
<tr>
<td>D. Experimenter Observation Data Sheet</td>
<td>79</td>
</tr>
</tbody>
</table>
Table of Contents - Continued

E. Post-Experiment Survey................................................................. 81
F. HSIRB Approval Letter................................................................. 84
LIST OF TABLES

1. Mean Number of Praise Statements Per Observation by Phase .............. 42
2. Mean Number of Corrective Statements Per Observation by Phase ........ 42
3. Mean Number of Praise Statements Per Corrective Statement by Phase. 43
4. Mean Percentage of Integrity for Approval Ratio by Phase ................. 43
5. Mean Percentage of Integrity for Expectation Setting by Phase .......... 47
# LIST OF FIGURES

1. Multiple-Baseline of Percentage of Integrity for Approval Ratio ........ 39
2. Multiple-Baseline of Praise Statements ........................................ 40
3. Multiple-Baseline of Corrective Statements ................................. 41
4. Multiple-Baseline of Percentage of Integrity for Expectation Setting ...... 46
5. Referrals Written by Teacher 1 .................................................... 49
6. Referrals Written by Teacher 2 .................................................... 49
7. Referrals Written by Teacher 3 .................................................... 50
8. Referrals Written by Teacher 4 .................................................... 50
INTRODUCTION

In the landmark report *A Nation at Risk: The Imperative for Educational Reform*, Gardner (1983) wrote: “The educational foundations of our society are presently being eroded by a rising tide of mediocrity that threatens our very future as a Nation and a people…” (p. 5). President Ronald Reagan’s National Commission on Excellence in Education published these findings 34 years ago, and yet American students’ achievement has shown little growth since the 1970’s (National Center for Educational Statistics, 2011). In 2012, students from the U.S. were outscored by 22 nations in Science and 19 nations in reading literacy on the Program for International Student Assessment (PISA). Additionally, 15-year-old students from the U.S. scored below the international average in mathematical literacy, with 29 education systems averaging higher scores on the PISA (Kena et al., 2016).

While student achievement has changed little since the 1970s, a number of interventions have been developed, shown effectiveness in the research setting, and have been implemented in the schools; yet broad scale change has yet to be realized (Detrich, 2013). One explanation for this lack of success may lie in the integrity with which academic interventions are developed and delivered. Determining a causal relationship between student outcomes and the interventions intended to affect them relies on having a well-specified, controlled, and consistently delivered independent variable (i.e., intervention). Failing to provide such calls into question if the effects demonstrated in the dependent variable (i.e., the student outcomes) are the result of the independent variable’s influence or other extraneous factors.
While it is a critical component of good science to assure the independent variable is the true cause of the change(s) observed in the dependent variable, amongst the intervention literature it is uncommon to measure and report the integrity with which interventions are delivered (Gresham, 2009). For example, McIntyre, Gresham, DiGennaro, and Reed (2007) reviewed the frequency in which studies published in the Journal of Applied Behavior Analysis (JABA) reported integrity and found that only 30% (46 out of 144) did so. This is an improvement in the results reported by Peterson, Homer, and Wonderlich (1982), and Gresham, Gansle, Noell, Cohen, and Roseblum (1993), who found 20% and 16% of articles in JABA included data on integrity, respectively. However, over the 25-year span these reviews represent this improvement seems marginal (Gresham, 2009).

The Effects of Varied Integrity Levels on Outcomes of Interest

Failing to measure the integrity of an independent variable does not mean that an intervention is not the cause of a change. However, a number of studies have shown that the level of integrity at which an intervention is delivered influences the degree to which desired outcomes are observed. For example, in an observational study, Arkoosh, Derby, Wacker, Berg, McLaughlin, and Barrett (2007) provided a descriptive analysis of the effects varied levels of integrity have on children when delivering functional communication training (FCT). Of five children receiving FCT, three successfully had their aberrant behaviors reduced (the impetus for the intervention) while results were not sustained for two children. Two of the three who maintained behavior change received two of the three interventions components at 80% integrity or greater, while the two with poor outcomes saw integrity rates below this threshold. Similarly, DiGennaro, Martens,
and Kleinmann (2007) found correlations between the accuracy of implementation of a student’s behavior plan and the student’s rate of the problem behavior across four teacher-student dyads. For three of the four dyads, significant correlations were found with the largest yielding $r = -.78$.

A handful of studies have taken an experimental approach to examining the effects of varied levels of treatment integrity on student outcomes. Wilder, Atwell and Wine (2006) examined the effects of a three-step prompting procedure at 100%, 50% or 0% integrity (i.e., the three-step procedure was used 100% of the time the demand was placed, 50% of the time, or not at all) on compliance to three common demands (‘bring me the item’, ‘come here’, and ‘put away the item’) with two preschool-aged children. The researchers used a multielement design preceded by a baseline phase in which no prompting procedures were utilized, and compliance rates were under 20% for all three demands. When the prompting procedure was delivered at 100% integrity, compliance increased to a mean of 91% and 79% for each child, respectively. When receiving 50% integrity, mean compliance rates were 54% and 41%; at 0% integrity, one student averaged 6% compliance while another student never complied.

St. Peter Pipkin, Vollmer, and Sloman (2010) utilized a translation model across several experiments to evaluate the effects of varied levels of integrity when implementing a differential reinforcement of alternative behavior (DRA) procedure. Noting that all integrity failures (i.e., failure to provide the intervention/procedure in accordance with the protocol) are not equal, the experimenters varied levels of integrity in regards to errors of omission (i.e., failing to deliver an earned reinforcer) and in terms of errors of commission (i.e., providing reinforcement following a problem behavior). In
their first experiment, undergraduate students could either click a red circle, the “appropriate” behavior, or a black circle, the “problem” behavior, which moved around a computer screen. During baseline clicking the black circle was reinforced by earning points on a fixed ratio one (FR 1) schedule, while the red circle was put on extinction. During full-integrity DRA the roles were reversed, with the black circle going on extinction and the red circle receiving points on a FR 1 schedule.

To simulate integrity failures responses on each circle were put on a random-ratio (RR) schedule. For example, to simulate 80% integrity with both omission and commission errors (hereafter referred to as combined errors), the appropriate behavior would have a 0.8 probability of reinforcement (i.e., an RR 1.25 schedule) and the problem behavior would have a 0.2 probability of receiving reinforcement (i.e., an RR 5 schedule).

The undergraduate students were assigned to a commission errors only condition, omission errors only condition, or a combined errors condition in which the effects of 100%, 80%, 60%, 40%, and 20% integrity were evaluated. For the omission errors only group, participants earned points less frequently when engaging in the appropriate behavior as the percent of integrity decreased (i.e., they moved from a FR 1 to a RR 1.25, to a RR 1.66, and so on). For the commission group, as integrity levels fell the probability of a problem behavior being reinforced increased (i.e., from extinction to a RR 5, to a RR 2.5, etc.), while the appropriate behavior remained on a FR 1. For the combined errors group, the probability of reinforcement for appropriate behaviors decreased as the integrity level fell, while the probability of reinforcement became available for the problem behavior and increased proportionally.
Ultimately, failures of integrity in each condition lead to decreases in appropriate behavior, although this was not observed until integrity fell to 40%. Additionally, differential effects were found based on the type of integrity failures, with errors of commission having a greater impact on problem behavior than those of omission. In the omission only group, problem behaviors were rarely if ever evoked, while the commission group and combined error group saw increases in the problem behavior often approximating a ratio of responses that served to maximize reinforcement.

In a second experiment St. Peter Pipkin et al. (2010) replicated the combined error condition with a fourth-grade student who had been diagnosed with autism. The results mirrored the findings of the first experiment. The student showed higher rates of appropriate behavior compared to problem behavior when integrity was at 60% or above, and higher rates of problem behaviors than appropriate behaviors when the integrity was at 40% and below.

DiGennaro-Reed, Reed, Baez, and Maguire (2011) analyzed the effects of commission errors during discrete-trial training for identification of nonsense shapes. Three 8-year-old boys with developmental delays were presented three different nonsense shapes, each associated with a different level of integrity in regards to the rate of commission errors. That is when the shape was selected incorrectly the student was reinforced (i.e., given a token and verbal praise) either 100%, 50% or 0% of the time. Across all three participants the highest rates of performance (i.e., correct identification of the shapes) were demonstrated when 0% of incorrect responses were reinforced. While one student showed clear differentiation in performance between the 50% and 100% commission error rates (the shape reinforced incorrectly 50% of the time was identified
correctly more often than the shape reinforced incorrectly 100% of the time), the remaining two students demonstrated similarly low accuracy under both conditions, suggesting 50% integrity can be just as detrimental as complete failures of integrity in some circumstances.

Stephenson and Hanley (2010) had a teacher implement a six-component antecedent intervention and three-step prompting procedure to attain compliance from two preschool children. After implementing the intervention with success (compliance averaging between 60% and 80%), the experimenters had the teacher systematically remove the antecedent and prompting procedures from 1 of 10 trials in which they presented instructions. In other words, during sessions a student in one of the ten instructions/trials would not receive the antecedent intervention, and in the same or another would not receive the three-step prompting procedure (the trial from which they were removed was randomly selected each session). If compliance remained high, i.e., greater than 40%, in the next sessions only eight of the trials would have the antecedent intervention, and eight would include the consequence-based intervention. This procedure continued until three successive sessions yielded compliance rates at or below 40%. In the end, compliance remained above this threshold until integrity was at 20%, i.e., only the antecedent and three-step prompting interventions were presented in two of the ten trials. This suggests that if an intervention can be implemented with high integrity to begin with the desired behavior change may maintain even as implementation integrity decreases.

Across these studies a consistent theme is illustrated, as implementation integrity decreases so too does the level of the desired behavior change. While there are a number
of factors that might influence the strength of this relationship (i.e., at how high did implementation integrity begin, or what type of integrity failures are occurring) this finding has been demonstrated across a number of different behaviors and populations.

**Variables Affecting Teachers’ Intervention Integrity**

While there is a body of evidence that underscores the value of implementation integrity, if teachers are unable to achieve or maintain the necessary levels to obtain desired outcomes, then the research is of little worth. As such, a number of researchers have sought to evaluate the variables which act to impede or facilitate practitioner integrity.

**Intervention Acceptability and Training’s Effects on Intervention Integrity**

Sterling-Turner and Watson (2002) trained 64 undergraduate students to implement a treatment protocol that they then delivered to a confederate undergraduate who was demonstrating facial tics. Before and after implementing the protocol, each participant rated how acceptable they found the treatment plan to be. These ratings were then correlated with the level of treatment integrity at which the participants implemented the intervention. Ultimately, no significant correlation between acceptability ratings and level of integrity was found, suggesting the intervention does not have to be well-liked to be implemented with fidelity.

Sterling-Turner, Watson, Wildmon, and Watkins (2001) conducted a study comparing the effects of three types of training on integrity. Sixty-four undergraduates were assigned to one of three conditions: didactic training, modeling training, or rehearsal/feedback training. The didactic training group received only a verbal explanation of the treatment protocol, while those in the modeling group watched a 5-
minute video of the protocol being implemented accompanied by a verbal description. The rehearsal/feedback training group implemented the protocol for five minutes with the experimenter and a confederate demonstrating facial motor tics. During this five-minute session, the experimenter provided prompts for the first two times each component of the protocol was implemented, then provided corrective feedback or praise for incorrectly and correctly implemented components respectively.

Following the training sessions, participants implemented the intervention on a confederate and had their integrity measured based on seven behavioral pinpoints. An ANOVA comparing the integrity levels of these three groups yielded a significant result with $p < .0001$. Those receiving the rehearsal/feedback training were found to have the highest levels of integrity, followed by the modeling training, while the didactic training group demonstrated the lowest levels of integrity of the three. This study lends evidence to the importance of the provision of feedback, and the effects observing another’s performance can have on implementation integrity. Although these training procedures require more effort from the experimenter or coach they have been shown to increase implementation integrity and as such may be worth investing resources into.

DiGennaro-Reed, Codd, Catania, and Maguire (2010) examined the effects of using video modeling for training, as well as video modeling coupled with performance feedback. Three teachers began in a baseline condition in which they were given only a written description of a treatment plan for their student. This was followed by a phase in which a four- to six-minute video of the student’s treatment plan being implemented was shown to the teacher before each observation. The experimenters then introduced performance feedback with the video-modeling. This performance feedback included the
experimenters providing verbal feedback about the teacher’s prior session performance, and prompting the teachers while watching the video to note upcoming segments because they represented errors that had previously been exhibited.

During baseline integrity levels averaged 41%, which improved to 84% when the video-modeling was implemented. While two teachers still demonstrated variability in their performance (one had their integrity measured at as low as 20%) this disappeared when the performance feedback was added to the video modeling. All three teachers demonstrated 100% integrity in every session under this condition and maintained this level of performance when measured one week later. These findings give additional support as to the effectiveness of video-modeling (integrity more than doubled upon its implementation) and also elucidate how including performance feedback may further assist in increasing the fidelity of implementation.

**Feedback’s Effects on Intervention Integrity**

DiGennaro-Reed et al. (2010) are not the only ones to find the use of feedback to be an important variable in increasing, achieving, and maintaining the desired levels of implementation integrity from teachers. Witt, Noell, LaFleur, and Mortenson (1997) conducted a nonconcurrent baseline with four teachers in which each teacher received training until they could implement an academic intervention with 100% integrity. Upon achieving this level of integrity, all feedback to the teachers was removed (i.e., a post-training baseline was implemented), at which point no teacher demonstrated integrity above 80% for more than two days. The post-training baseline was followed by a daily feedback condition in which teachers received graphic feedback on their implementation integrity and their student’s performance on an academic task. With the introduction of
performance feedback, immediate increases in integrity were demonstrated by all four teachers, with each eventually achieving 100% integrity and two doing so after receiving feedback only once.

Noell, Witt, Gilbertson, Ranier, and Freeland (1997) extended this work by removing the intensive training that Witt et al. (1997) provided at the beginning of their study. Instead, the three teachers in the Noell et al. study received what the authors believed more closely resembled the typical model for introducing a new intervention. The teachers in the study met with the experimenters and were provided only the rationale for the intervention, data demonstrating its effectiveness, and a verbal explanation of how it was to be implemented. During this phase teachers were able to achieve high levels of integrity for the first two to four sessions, however, by the end of the phase integrity was averaging near or below 40%. Once daily performance feedback was introduced (5-min morning meetings which included praise for success and corrective feedback for errors) integrity immediately improved to averages above 80%, and all teachers attained perfect integrity on multiple occasions.

DiGennaro, Martens, and Kleinmann (2007) utilized a system of feedback to achieve and maintain four special education teachers’ integrity when implementing function-based treatment packages for students’ problem behaviors. During one phase of the study teachers set goals for a student’s performance and received written feedback in regards to the student’s approximation to the goal. In the next phase the experimenters added feedback on the teacher’s accuracy in implementing the intervention, and a negative reinforcement contingency in which the meeting with the experimenter would be canceled if the teacher demonstrated perfect integrity. While goal setting on student
performance was only effective in increasing integrity for two of the four teachers, all four teachers saw dramatic improvements when also receiving feedback on their performance with the meeting cancellation contingency (integrity increases above baseline were between 79% and 100%).

Additionally, at the end of their study, DiGennaro et al. included a phase in which a teacher demonstrating three consecutive sessions of 100% integrity began to have their feedback systematically removed. Their performance feedback went from every day to every other day, then to once per week, and finally to once every other week. All four teachers were able to move to feedback every other day, three moved to once a week, and one was able to move to receive feedback every other week (the authors note constraints with the school’s calendar was a factor in moving all teachers to receiving feedback every other week), suggesting daily feedback is not necessary to maintain integrity over time.

A few studies also support the effectiveness of feedback being delivered on a relatively infrequent basis. Mortenson and Witt (1998) trained four teachers to implement interventions from one of their student’s behavior plans. After the teachers received training (i.e., prompts by the teacher consultant/experimenter as the teacher implemented the intervention) and could implement the intervention at 100% integrity, all feedback was removed. Once integrity levels dropped below 70%, weekly performance feedback was delivered. Three of the teachers received the weekly feedback and saw immediate increases in their demonstrated levels of treatment integrity (10% to 46% from the post-training baseline), while the fourth never dropped below the 70% threshold and therefore never received feedback.
Codding, Feinberg, Dunn, and Pace (2005) trained five teachers to implement an antecedent- and consequence-based intervention for one of their students. The experimenters then observed the teachers every one to three weeks to collect data on their rates of integrity. Once stable or decreasing performance was observed for a teacher feedback was delivered. This feedback included praise for the antecedents and consequences delivered correctly and corrective feedback on components that were omitted or implemented incorrectly. Although the teachers received feedback every 2.1 weeks (on average), all five demonstrated improved integrity for the consequence components, and four of the five teachers saw improvements in the delivery of the antecedent components.

While there is evidence that integrity is essential to achieve the outcomes that teachers and researchers desire in their students, and work has been done to identify variables that factor into achieving it, systems which are feasible for implementing new interventions on a broad scale have yet to be evaluated. To date, the research on increasing and maintaining teacher integrity relies on the teacher consultant role, which is often played by experimenters in reported studies. Unfortunately, it seems unlikely that a school could afford to hire the number of consultants that would be necessary to implement a school-wide intervention. Ultimately, it seems necessary to develop systems which utilize the most plentiful resource a school has and already invests in, its teachers.

**Behavioral Pinpoints and Peer Observations for Performance Management**

One solution to broad scale change may come from the safety literature. The foundation of behavior-based safety (BBS) is built upon pinpointing the behaviors which constitute the safest means to complete a work task (e.g., bending one’s knees when
lifting an item), then delivering feedback to highlight what was safe and what was not when an employee engages in the task. This parallels the structure of the integrity work with teachers in that both have utilized pinpointed behaviors and a system to increase engagement with each identified.

A seminal study of behavior-based safety was done by Komaki, Barwick, and Scott (1978) who intervened in two departments at a wholesale bakery. After reviewing the previous three years of injuries, and seeking input from the managers at the bakery, a list of 15 and 20 pinpointed behaviors were identified for the wrapping and makeup departments respectively. In a multiple baseline format employees were then trained utilizing example and non-examples of the behavior in a 30-minute slideshow and were additionally given feedback via a publicly posted chart of their work group’s percent safe (i.e., their level of integrity) which was updated daily based on that day’s observation. During baseline conditions, the wrapping and makeup departments averaged 70.0% and 77.6% safe respectively. During the intervention phase these percentages rose to a mean of 95.8% for the wrapping department, and 99.3% for the makeup department. To demonstrate experimental control, a reversal phase was introduced and the feedback was removed, at which point the integrity levels fell back to near baseline levels with 70.8% and 72.3% of the pinpointed behaviors being completed safely in the wrapping and makeup department respectively. This study is part of a larger body of literature showing similar outcomes (e.g., Lebbon, Sigurdsson, and Austin, 2012; Cooper, 2006; Medina, McSween, Rost, and Alvero, 2009).
Peer Observation and the Observer Effect

While the efficacy of pinpointing critical behaviors and giving feedback on them has been demonstrated to be effective, Komaki et al. (1978) still relied on an independent observer to collect data. This is reasonable in a plant where many workers can be observed at once, however, in a school where each teacher works independently and can only be observed individually, the practicality of this system is lost. Additionally, using an independent observer (e.g., a behavior coach, consultant, or experimenter) decreases the likelihood of maintenance and institutionalization of the observation process as the position may not be maintained, or may disappear when the experiment ends, leaving no structure to maintain the systems. Such being the case, researchers have been examining the impacts peer observation systems can have not only on integrity levels of those observed and receiving feedback but additionally on those who observe and are delivering the feedback.

Alvero and Austin (2004) and Sasson and Austin (2005) examined the impact observing others had on the behavior of the observers (i.e., the observer effect). In Alvero and Austin’s study eight participants were asked to complete three tasks (a) typing a few paragraphs using a word processor on a computer; (b) dialing a phone number and leaving a brief message on an answering machine; and (c) picking up a cardboard box containing five pieces of paper, placing it on a chair, removing a specific piece of paper, and placing the box back down on the floor. Each participant was assigned to one of two experimental groups and had their performance videotaped and scored based on a list of eight target behaviors. Both groups underwent a baseline condition in which they were asked to perform each of the three tasks at minimum four times. This was followed by an
information phase in which they were given a handout that listed four of the eight target behaviors (four of eight were assigned to one group, the other four were assigned to the other group) and how to perform them safely. Following the information phase, the first observation phase took place in which participants observed a five-minute video of a confederate performing the experimental tasks (i.e., tasks the participant had been engaging in for the experiment). During this observation phase, the participants were asked to collect data on the confederate’s safety performance using a checklist containing the same four target behaviors given to them during the information phase. Immediately after scoring one of the videos, participants were then asked to perform the tasks as they had before. This was followed by an additional observation phase in which the remaining four target behaviors (i.e., the four the participants had not received information about, nor completed an observation focused on) were introduced to replicate the findings of the first observation phase and demonstrate experimental control.

One behavior (sitting with feet on the floor) moved from near zero to between 70% and 100% engagement in participants during the information phase. Additionally, temporary increases in lifting with knees and using the proper neck position were observed for three of the participants, but these were followed by gradual decreases towards baseline performance. However, of the 32 opportunities for the replication of target behavior (eight participants that could replicate four target behaviors each), 25 of 32 of these opportunities saw no change in performance (i.e., 78% of the time there was no behavior change).

Since all eight behavior targets were eventually introduced to all eight participants, there were 64 opportunities for the replication of the target behaviors during
the observation phase. However, unlike the information phase which saw 78% of the behaviors unchanged, only 6% (i.e., four of the 64) did not change during the observation phases, two of which were due to a procedural flaw in the experiment (participants were never instructed to stand up when they lifted and as such would not meet the criteria for lifting properly). For 18 of the 64 opportunities (28% of the targets), gradual increases in safe behavior were observed, moving from 0% levels to eventually achieve between 83% and 100% integrity. The remaining 66% of behaviors (42 of 64) demonstrated dramatic and immediate increases with two moving from 0% safe during baseline to 100% safe during the observation phase.

Sasson and Austin (2005) extended the work of Alvero and Austin (2004) with 11 employees at a hospital. Ultimately, they found that those who engaged in observing others demonstrated significant increases in the target behaviors regardless of the accuracy of their observations (i.e., there was no significant correlation between accuracy of feedback and the performance of the observers). This may be related to the process of self-monitoring which nearly every participant in the Alvero and Austin (2004) study reported they engaged in on a post-experiment survey. These studies give strong evidence as to the effects observing others can have on one’s own behavior. Additionally, both demonstrated that the provision of the pinpoints alone was largely ineffective in causing a change in behavior, while dramatic improvements can be yielded by observing and evaluating others’ performance.

Lebbon, Sigurdsson, and Austin (2012) explored how the frequency of observations and feedback in a peer-observation system at a food and drink industry site affected the number of Occupation Safety and Health Administration (OSHA) recordable
incidents (i.e., workplace accidents resulting in treatment which goes beyond first aid), lost days (i.e., days in which an employee cannot work due to a work-related injury), and restricted days (i.e., days in which an employee can only perform portions of their normal job roles due to a work-related injury).

After a baseline condition in which all employees were provided with a binder that included written materials on safety procedures (the standard practice at the site), all employees were provided with a flyer which outlined the behavior pinpoints being examined in the study. This was followed by a phase in which a subsample of employees was selected and trained to conduct peer observations and deliver feedback to those observed. Observers were trained to praise safe behaviors and also provide some corrective feedback if at-risk behaviors were observed. Following implementation with the subsample, all employees were trained to be observers and the system of peer observation was implemented site-wide.

Before the implementation of the peer observation system, the average number of OSHA incidents per semester was 28.6, restricted days was 489.0, and lost days was 18.0. After the subsample of employees began these observations OSHA incidents and restricted days reduced to an average of 18.0 and 229.7 per semester respectively, while lost days increased to 112.3. When all employees were participating in the observation system reductions in OSHA incidents and restricted days in comparison to baseline were demonstrated, with OSHA incidents averaging 21.0 per semester and restricted days averaging 86.4 per semester. However, lost days were still elevated averaging 45.6 per semester. When correlating the number of observations conducted with the number of OSHA incidents a significant negative correlation was demonstrated with $r = -0.68$ ($p <$
lending strong evidence as to the efficacy of peer observations in increasing safe behaviors and in turn reducing injurious incidents. While the average number of lost days did increase during both observation phases of the study, the authors note that the nature of injury severity measures makes it difficult to interpret this outcome. For example, one injury may result in 100 lost days, or 100 injuries may lead to one lost day each, with both representing 100 lost days. The decrease in the total number of OSHA incidents points to the idea that a handful of more debilitating injuries took place during these phases of the study and thus inflated the lost day figures.

Cooper (2006) developed a peer-observation system targeting safe behaviors at a papermill. After introducing all personnel to the procedural steps involved in the behavioral safety process (e.g., helping to develop behavior checklists, observing and being observed, setting improvement targets, and receiving feedback about safety performance), participants were recruited to assist in performing observations of their peers. The observers conducted 15-30 minute observations focused on a tailored list of behavior pinpoints for the department, and were encouraged to provide verbal feedback in regards to safety performance after conducting each observation. Additional feedback was provided to all employees in the form of weekly publicly posted charts displaying the percent safe observed by the experimenters along with a unit-determined goal line for percent safe. This process was iterative in that baseline conditions occurred (i.e., no feedback or observations took place), which were then followed by the recruitment of new observers for the observation process, and the reinstitution of feedback. This occurred on three separate occasions. During the first baseline, employees demonstrated a percent safe of 49.3%, which improved to exceed the unit’s goal (70% safe) and achieve
70.6% safe. The baseline for the second phase ranged between 66-69% safe, which improved to an average of 77.9% safe with the reintroduction of the intervention. During the third and final baseline, percent-safe was found to be 75.75% (a 25.7% increase from the first baseline) and improved to an average of 83.7% when the intervention was introduced for the final time.

In the 12 months prior to Cooper’s intervention 202 injuries were recorded at the papermill. This was reduced to 150 injuries in the first 12 months of the peer-observation systems implementation (a 24.75% reduction) and further reduced to 112 injuries in the following 12 months (an additional 25% reduction, and an overall reduction in the total number of injuries of about 45%).

Medina, McSween, Rost, and Alvero (2009) also examined the use of peer observations on safe behavior at a petroleum refinery. As part of a pilot study, 189 employees were trained to use a checklist of critical safe behaviors and given monthly feedback on the percentage of employees completing at least two observations per month, and the number of safety concerns raised. During the 14 months preceding the intervention, one or more accidents occurred every month. Upon implementing the peer-observation system, no incidents were recorded for 8 of the following 11 months, which was then followed by 24 consecutive months in which no recordable injuries were recorded. The success of the pilot study led to the peer observation system being implemented across the refinery, which resulted in the number of lost-time incidents being reduced by more than half.

These studies give strong evidence as to the positive outcomes that can be produced from the use of pinpointed behaviors in a peer-observation system. The use of
such a system can be effective in not only increasing the integrity in which a task is engaged in by the person being observed and receiving feedback, but additionally for the observer who evaluates their performance.

**Related Research**

The current researcher conducted a literature search for “teacher peer observations” which yielded the following results: a) 212,000 books, articles, and book chapters; b) of the 212,000+ publications, 47,854 articles were peer-reviewed journal articles. A cursory review of the first 200 of the peer-reviewed articles found only 61 publications discussed teachers observing or coaching other teachers (i.e., 30.5%). Of these 61 articles, only two measured behavior change brought about by the observation process. The remaining 59 articles instead focused on and measured responses to surveys, and the content of post-observation interviews based on perceptions of topics such as community support and well-being (e.g., Adshead, White, & Stephenson, 2009; Harper & Nicolson, 2012; Karagiorgi, 2012; Amrein-Beardsley & Osborn Popp, 2012).

Bruce and Ross’s 2008 article was one of the two studies which measured changes in teachers’ performance as a result of engaging in peer observation. In their study, 12 teachers were paired into dyads and asked to observe their peer-partner and give feedback, be observed and receive feedback, help their partner set mathematics teaching goals, and reciprocally set their own goals with the assistance of their partner. Feedback was provided on three dimensions of teaching mathematics which had been identified in previous in-service training including a) selection of appropriate mathematics tasks, b) increasing student construction of mathematical knowledge, and c) providing opportunities for student-student interaction. To ensure consistency across
teachers observation training was undertaken, and during each observation the teachers recorded examples that illustrated the reason for their ranking. The researchers found that teachers did increase their use of open-ended questions, and facilitated more student-student interactions; however, no changes were observed for student construction of knowledge.

Englert and Sugai (1983) examined the effects of observation and feedback for 20 preservice special education teachers. They compared the changes in performance when receiving feedback from a peer using an open-ended observation process against changes in performance for those who received feedback from a peer using a specific, well-defined observation instrument. Those who were randomly assigned to utilize the developed observation tools recorded interactions between teachers and students in 15-second intervals, noting whether the student was engaged in inappropriate or appropriate behavior and additionally noting whether the teacher provided approval or disapproval for the behavior. They also recorded data on whether a teacher provided an attention signal, received a correct or incorrect response from their student, provided reinforcement of correct responses, provided correction or prompting of incorrect responses, and if the teacher told the correct answer following an incorrect response. Those assigned to the control group were asked to create their own observation system following a discussion of desirable teacher strategies and methods for behavior management and direct instruction (e.g., the pinpoints identified from the structure observation tool).

Each participant in the study observed a peer five times and was observed by their peer five times. After each observation, the pair met to discuss the performance of the observed peer. Following this meeting, the observed peer was given the data sheet the
observer created and was tasked with writing a brief report that described their
instructional methods and the findings of the peer observer (the experimental group was
carefully to summarize the data they obtained by graphing the teacher-pupil data).

In the end, those who received feedback from the well-defined observation tool
maintained a significantly higher level of pupil accuracy, provided more opportunities for
student responses, prompted correct responses less often than control teachers, and were
less likely to tell correct answers following incorrect responses than those receiving
feedback from the open-ended observations. The authors note that these differences
cannot be explained by peer observation alone nor by the categories of focus as these
were held constant across both groups. Instead, they explain that it was the nature of the
observation instrument that made the significant impact on behavior, and having
pinpoints of the behaviors you are seeking to change is one mechanism to achieve it.

**Purpose of Current Study**

This experiment looked to expand and extend the findings of Englert and Sugai
(1983), and Alvero and Austin (2004), and additionally sought to evaluate if the success
demonstrated in the safety literature might be realized in the schools.

Since the use of an independent observer (e.g., a behavior coach, consultant, or
experimenter) decreases the likelihood of maintenance and institutionalization of an
intervention, a peer observation systems will be established to see if the positive
outcomes of Lebbon, Sigurdsson, and Austin (2012), Cooper (2006), and Medina,
McSween, Rost, and Alvero (2009), could be replicated.

A structured observation tool, like the ones used by Englert and Sugai (1983), was
utilized to target rates of praise and corrective feedback, and student expectations set by
the teachers. Unlike Englert and Sugai, this experiment utilized full-time teachers in a public school instead of pre-service teachers. Additionally, the participants were not asked to meet and discuss the outcomes of the observations. It was designed as such to aid in the evaluation of the observer effect (i.e., the effect observing another has on one’s own behavior) demonstrated by Alvero and Austin (2004), and Sasson and Austin (2005).

Before teachers began conducting observations participants were first introduced to the interventions evaluated in this study via a half-day professional development run by the experimenter acting in the role of the school’s Behavior Coach. The training was similar to what Noell et al. (1997) described as the typical procedure for introducing an intervention, and included only the rationale for the intervention, and a verbal explanation of how it was to be implemented.

Mortenson and Witt (1998), and Codding, Feinberg, Dunn, and Pace (2005) demonstrated that feedback on performance can be delivered relatively infrequently and still be effective. A goal of this study was to evaluate a practical system that could be implemented in any average public school, and many teachers will attest that time is one of their most precious resources. It is possible that observing another’s performance serves a similar function to performance feedback. As such this experiment had teachers observe one another only one time per week.

There are seven primary components of a Positive Behavioral Interventions and Supports (PBIS) system. They include: (a) an agreed upon and common approach to discipline, (b) a positive statement of purpose, (c) a small number of positively stated expectations for all students and staff, (d) procedures for teaching these expectations to students, (e) a continuum of procedures for encouraging displays and maintenance of
these expectations, (f) a continuum of procedures for discouraging displays of rule-violating behavior, and (g) procedures for monitoring and evaluation the effectiveness of the discipline system on a regular and frequent basis (“Tier 1 FAQs”, 2018). This study looked to evaluate if peer observations can serve to increase the implementation integrity of the expectation setting component, and procedures for encouraging displays and maintenance of these expectations.

Observations were completed by teachers in a multiple-baseline format. This design was utilized to assist in maintaining experimental control and diminish explanations of behavior change due to confounding and/or extraneous variables introduced during the experiment (Baer, Wolf, & Risley, 1968).

METHODS

Participants, Participant Selection, and Assignment

Participants in this study were a team of four seventh grade teachers that worked in a public middle school and provided informed consent to participate in this study. The team of seventh grade teachers was selected because of the high number of office referrals that they submitted for students in their classes. That is, the team of teachers with the most referrals in the middle school were selected to receive the intervention in the current study.

Teacher 1. Teacher 1 was a seventh grade science teacher with two years of prior teaching experience. This was his first year as a teacher in the middle school where the study occurred. Teacher 1 had an undergraduate degree in biology and was completing a graduate degree in biology. He was certified to teach science at the secondary school level. Teacher 1 had received no prior training on the dependent variables in this study.
Teacher 2. Teacher 2 taught seventh grade reading with six years of prior experience. This was his first year as a teacher in the middle school where the study occurred. Teacher 2 was certified to teach English and Language Arts, Speech Communication, and Theatre, and was also working on his Master's in Special Education, specifically focused on reading. Teacher 2 had never previously been trained on the dependent variables in this study.

Teacher 3. Teacher 3 taught seventh grade writing with three years of prior experience. This was Teacher 3’s first year as a teacher in the middle school where the study occurred. Teacher 3 was certified to teach English in grades 5-12 and had not started her master's degree. Teacher 3 had not previously been trained on the dependent variables in this study.

Teacher 4. Teacher 4 taught seventh grade social studies with 28 years of prior experience. Teacher 4 taught for 25 years in the middle school where the study occurred. She had her Bachelor's Degree in Anthropology and was certified to teach Sociology, Psychology, Anthropology, and History. Teacher 4 had earned a master's degree and approximately 30 additional graduate credits. She was previously trained on the dependent variables in this study on multiple occasions.

Setting

This study took place in a public middle school in a small city in the Midwestern region of the United States. The school employed 22 teachers including four fifth grade, four sixth grade, five seventh grade, five eighth grade, and four special education teachers. The middle school contained approximately 425 students from grades five through eight. In the 2016-2017 academic year, 59% of students were White, 18.5% were
African-American, 11.6% were Hispanic, 8.7% were two or more races, 1.7% were Asian, 0.4% were Native American, and 0.2% were Pacific Islander. Low-income students were 69.7% of the population and 23% of the student body was labeled as learning disabled. The current study was conducted with the seventh grade teaching team and its students. This team taught approximately 99 students with each teacher instructing five 57-minute class periods each day. The mean number of students in each class that was observed was 21 (range, 17 to 22).

Teachers’ classrooms were equipped with individual student desks for each student. Additionally, in every classroom the teacher’s desk was equipped with a desktop computer connected to a projector and speaker system. The projector and speaker system could also be connected to a tablet, which was often used as a camera to display documents with which the teacher worked. Classrooms were connected to a small closet or cabinets that could be utilized for storage by the teacher. During the experiment, observers stood at the back of the room or sat at an open student desk/table to minimize disruption to the class.

**Dependent Variables**

There were three primary dependent variables utilized in this study: 1) the number of praise/positive statements made by the teacher, 2) the number of corrective statements made by the teacher and, 3) the ratio of teacher praise/positive statements to corrective feedback (i.e., approval ratio). These dependent variables represent procedures for encouraging displays and maintenance of the classroom expectations, a critical component of a PBIS system.
Praise/positive interactions were defined as feedback to students following correct behaviors exhibited by one or more students. For instance, a teacher telling a student, “Thank you for remaining in your seat” or “I like how everyone is raising their hands to ask for help” are examples of praise/positive statements.

Corrective feedback was defined as an interaction in which a teacher tells or asks one or more students to engage in a different, or the appropriate, behavior following a misbehavior. Corrective feedback could have occurred if a student was walking around a room and the teacher told or asked the student to take their seat, or when all students were speaking too loudly, and the teacher told the students they ‘need to quiet down’ or “Shhh.” Statements such as, “Please be quiet, thank you.” were scored as a corrective statement unless there was a 2-second pause between the what could be considered the corrective phrase (“Please be quiet…”) and the praise statement (“thank you”).

The approval ratio was determined by evaluating the teachers’ approximation to a 3:1 positive feedback to corrective feedback interactions during an observation. For example, if a teacher demonstrated a 2:1 ratio, their integrity was evaluated as 66.7% as \( \frac{2}{1}/\frac{3}{1} = 0.66 \). If a teacher engages in a rate of three praise statements for every two corrective feedback statements, their integrity score would be 50% as \( \frac{3}{2}/\frac{3}{1} = 0.50 \). If a teacher’s praise to corrective feedback ratio exceeded the three to one standard (e.g., 4:1 or 7:2) their integrity was scored as 100%. Only Teacher 4 claimed to have ever been trained on the approval ratio previously.

Two secondary dependent variables were also evaluated. The first was the percentage of expectations set by the teacher before every classroom activity transition. The second was the number of office discipline referrals written by the teacher.
Expectations were presented to address four specific student behaviors: (a) the types of conversations students may have and how loud these conversations may be, (b) how students can ask the teacher for help, (c) the type of activities in which the students will be engaged, (d) how students may move around the classroom during an activity, and (e) what participation in the activity looks like for the student.

Teachers’ expectation setting was measured using the CHAMPs paradigm. CHAMPs is an acronym in which the “C” represents the conversation and conversation sound level students are allowed while engaging in the activity (e.g., you may speak with the person next to you and you must whisper only). “H” is utilized to define how students can ask for help when they run into problems while completing the activity (e.g., if you need help you may come up to my desk). “A” specifies what the activity students will be engaging in is (e.g., you will be completing a multiplication worksheet). “M” is used to define the type of movement students may engage in around the classroom (e.g., you should remain in your seat for the duration of the activity unless you need to sharpen your pencil). “P” explains to students what it will look like if a student is engaging in the activity or participating in an activity (e.g., you will have your eyes on the paper and I will see your pencil moving if you are participating in the activity).

Teachers in the school were familiar with CHAMPs as it was used in the school during the 2012-2013 school year. Since that time, it has not been formally taught to staff, however, books on its use were available to any teacher who had an interest. Only Teacher 4 in this study stated they had any formal training on CHAMPs previously.

According to the CHAMPS protocol, expectations for each component of CHAMPs should be established before each new classroom activity (i.e., after a
transition). A transition was marked by a change in the materials needed for participation (e.g., going from taking notes and using a notebook to completing a worksheet), or any activity in which new directions were given by the teacher to complete the task (e.g., when going from writing a free write in a notebook to writing a persuasive essay).

To calculate the percentage of expectations set by the teacher, the number of activities was multiplied by five (the number of components of CHAMPS) and divided by the number of expectations vocally set. For example, if during an observation a teacher has students engage in two activities and specifies only the conversation (“C”) and activity (“A”) in both instances their percentage would be 40% since 5*2 = 10 and they set two expectations each time, so 4/10 = .40 or 40%.

Office referrals served as the second secondary dependent variable. Office referrals were documented instances in which students engaged in problem behaviors that warranted the school’s office staff to intervene and result in the student receiving a lunch detention, in-school suspension, and/or out of school suspension. Referrals were logged by the teachers in the school district’s online learning management system. Referrals included behaviors such as fighting, swearing at a teacher, destruction of property and many other misbehaviors. A complete list of behaviors that warranted referrals is available in Appendix B.

Independent Variable

The independent variable for this study was a participant's observation and scoring of a peer teacher's performance on the dependent variables.

Peer observations and performance scoring. During the study, a teacher observed each of the other teachers in their grade group (i.e., one seventh grade teacher
observed each of the other seventh grade teachers) and scored their peer’s performance for the approval ratio and expectations set. The teachers were asked to complete observations from 15- to 20-minutes in length. During the observation, the observer tallied the number of positive and corrective interactions completed by their peer teacher. They also noted the number of activities for which the peer teacher should have delivered the CHAMPs expectations and for which CHAMPs expectations were delivered. The observing teacher filled out a data sheet which prompted them to record the data noted above as well as to calculate the approval ratio (see Appendix C). Each teacher observed one other teacher every week until they had completed one observation of every teacher in their grade-level teaching team. The data sheets they filled out were collected weekly by the experimenter to confirm the level of integrity with which the observation was conducted (i.e., the independent variable was being carried out with integrity) and to address any questions that the participants had about the process.

Data Collection

Before the beginning of the academic year began, data regarding the number of office referrals submitted by each teacher in the school were collected to begin identifying possible candidates for the intervention. These data included the total number of referrals submitted by the teacher, the number of students for which they were submitted, and a categorization of each referral into specific referral categories (e.g., 29 referrals across 14 students for Absenteeism/Tardy). Throughout the study, data on the number of referrals, the number of students for which referrals were written, and the frequency of use for each category of referral were collected for all teachers in the school.
To collect data on teachers’ integrity, an observation data sheet (see Appendix D) was utilized during 15- to 20-minute observation sessions conducted by the lead experimenter and his research assistant. Teachers identified to receive the intervention were observed three to four times per week for the duration of the study to evaluate performance changes.

**Post-experiment survey.** At the conclusion of the experiment, each teacher participant was asked to complete a post-experimenter survey which can be seen in Appendix E. This survey was utilized to determine the social acceptability of each phase of the intervention and to obtain feedback about additional barriers and catalysts to a teacher’s success.

**Interobserver agreement.** One research assistant was trained to conduct observations of the teachers in the study. For the research assistant to observe independently, they had to demonstrate interobserver agreement (IOA) of 90% or greater with the lead experimenter for two consecutive observations. IOA was calculated by dividing the total number of agreements for the measures (positive interactions and corrective interactions by one-minute intervals, classroom transitions/activities, and CHAMPs expectations observed) by the number of agreements plus disagreements.

The research assistant and each teacher completed one simultaneous observation to evaluate the observing teachers’ accuracy in measuring the dependent variables. Since the teacher observation forms were not separated into intervals for positive and correct statements the measure of accuracy lacked point-to-point correspondence. Instead, accuracy was determined based on the final count in number of praise statements, corrective statements, classroom activities/translations, and CHAMP expectations set, and
the assumption that the difference in these frequency counts represented disagreements. For example, if Teacher 1 scored five praise statements, five corrective statements, one transition, and five expectations, and the researcher scored four praise statements, four corrective statements, one transition, and three expectations set, accuracy would be measured at 75% (i.e., 12 agreements (4 for praise statements, 4 for corrective statements, 1 transition, and 3 expectations), and 16 agreements plus disagreements).

**Experimental Design**

This study employed a multiple baseline across participants design (Baer, Wolf, & Risley, 1968; Komaki, Barwick, & Scott, 1978). Each of the four teachers was exposed to baseline conditions and an intervention phase in which they observed and scored a peer’s performance. Two of the four participants also entered a second baseline condition after the intervention, and one received a weekly feedback phase after the intervention. A fourth participant did not complete any post-intervention phases because of participant attrition.

One week after the first teacher began conducting observations, the second teacher began conducting weekly observations. This pattern continued with the remaining participants in a multiple baseline format. To determine the sequence of teachers the participants would observe, each teacher in the seventh grade team was assigned a numerical value one through five. Beginning with Teacher 1, the random number generator was used to select a value one through five, until each value (two through five, in this case) had been drawn at least one time. The sequence in which these values was presented determined the sequence of observation. Any repeated values (e.g., two being
presented twice before four was ever generated) were ignored to assure each teacher observed each of their peers just one time.

**Procedures**

**Pre-baseline.** Before the beginning of the school year, all teachers in the middle school participated in a half-day professional development conducted by the experimenter. During the professional development, the teachers engaged in an interactive lecture (e.g., the experimenter asked questions periodically and had teachers respond) in which the CHAMPS system for setting behavior expectations, praise statements, corrective feedback, and what constitutes an office referral were defined. Along with these definitions, a rationale for the use of each of these components was provided, as well as how data would be collected on each teacher (e.g., the system of observations by the experimenter and his research assistant, and how teachers may observe each other). The participant selection process began the third week of the school year. Participants from the grade-level teaching team with the highest number of referrals were enrolled in the study when they had provided signed informed consent forms to participate.

When the experiment was introduced to the teachers, they were assured all data would remain confidential and were encouraged not to discuss what they recorded when observing their peers. This was done for two reasons: 1) to help teachers feel comfortable that their data were protected, and 2) to reduce the possibility that peer feedback would play a role in the outcomes observed.

The teacher participants were given the informed consent approximately two days before the study’s procedures were outlined to allow them to review the document before
meeting with the experimenter. The experimenter met with the teachers during their common lunch time. During this meeting the experimenter again explained the observation process, asked if there were any questions in regards to the what was being measured, asked if there were any questions in regards to the informed consent document, and concluded by presenting the schedule of observations (i.e., the schedule for when each teacher would observe each of their peers). This meeting took approximately 15 minutes. Observations by the experimenter and his research assistant began the week following this pre-baseline introduction to the experiment.

**Baseline.** During baseline, participants were each observed approximately three days a week, once a day at the same time, for 15 to 20 minutes. During the observations the experimenter or research assistant entered the room and found a seat near the back of the class where they remained for the duration of the observation. Teachers continued to teach as if nothing had changed, and the observer made every effort to avoid interactions with the students and the teacher. During the observation, the observer completed the observation form, tallying the number of praise and corrective statements they heard, and the opportunities to set expectations along with which expectations were established.

The experimenter and/or research assistant remained in the room for 15 minutes, or until five minutes after the last activity transition occurred. For example, if an activity transition was observed during the 13th minute of the observation, the observer remained in the class until the 18th minute. This was done to reduce the possibility that the data on expectation setting were made inaccurate because the observer left the class before the teacher stated all of their expectations.
In the event an observation was interrupted (e.g., a fire drill began in the middle of the observation, or the class was called down for an assembly) the session was immediately terminated and rescheduled for the next day in which an observation was not already scheduled. Any data collected from an incomplete observation were excluded from consideration.

At the end of the observation, the lead experimenter and/or research assistant stood up (if seated) and walked quietly out of the room, making every effort to minimize the disruption to the class. Minimal eye contact was made with the teacher and students and silent waves were given in response to students and/or the teacher saying goodbye.

**Intervention - peer observation and performance scoring.** After the baseline phase, participants began completing observations of the teachers in their grade group. They conducted one observation per week of one teacher in the grade group. The teachers were instructed to complete observations in the same manner as the experimenter. That is, they were asked to conduct observations 15 to 20 minutes in length, enter the rooms and make every effort not to interact with the class or teacher, record data on their observation data sheet, and leave the class making every effort to minimize the disruption to the learning environment. Participants were also encouraged to complete the observations on the same day every week. The experimenter attempted to collect the teacher’s observation data sheet on that day. In the event the participant had not completed the observation, the experimenter asked them to complete it the next day and followed up then in the same manner. During this phase observations by the experimenter and research assistant continued in the same manner as during the baseline phase.
**Return to Baseline.** After two of the teachers had observed each of the teachers in their group, they began a return to baseline phase. That is, they returned to the same conditions as the original baseline in which teaching as usual took place (i.e., no observations were completed by the teacher). Data were still collected by the experimenter and his research assistant in the same manner as the previous phases of the experiment.

**Weekly feedback.** Teacher 3’s performance during the peer observation phase (lowered praise statements, increase corrective statements, and less expectation setting) warranted additional intervention to increase integrity levels on the approval ratio and expectation setting. This came in the form of a weekly feedback condition. During these weekly meetings, the lead experimenter met with the teacher and showed him the data collection sheets from observation(s) earlier in the week. The teacher’s approval ratio, opportunities to set expectations, expectations set, and additional notable events that took place during the observation related to these measures were discussed. For example, a feedback session might include the following, “Here is the data I collected when I observed you yesterday. You made three praise statements and one correction, a three to one ratio. You did a really nice job ignoring that student who was whispering loudly while they were supposed to be silent reading, and instead found Sally and praised her for doing what you asked. You also had one transition when you went from silent reading to your lecture. Before you started the lecture you set the ‘conversation’, ‘activity’ and ‘help’ expectations, but I did not hear anything for ‘movement’ or ‘participation’.” The experimenter frequently paused while presenting this information to allow the teacher to respond and ask questions. Ultimately, no questions were asked, but instead the teacher
often provided evaluative statements about his performance (e.g., “Oh, I did great.”), and stated goals for future performance (e.g., “It is my goal to increase my praise statements.”). These meetings were approximately five to ten minutes in length and took place during the teacher’s plan period.

**Intervention Integrity**

Intervention integrity was determined by evaluating the observation forms completed by the teachers. Three components were considered when determining intervention integrity. The first was if the observation was completed during the week it was scheduled to occur. The second and third components were whether the two major sections of the observations forms were completed (i.e., if the approval ratio and expectations were scored). If any tallies were marked in either the praise or corrective statement boxes the approval ratio was scored as having been completed with integrity. If at least one transition was marked as having occurred, the expectation setting section was scored as having been completed with integrity. In the event either section did not meet these criteria the “notes” area of the observation form was referenced for a possible explanation as to why these may not have been completed (e.g., “There was no activity transition.” or “I did not hear any praise or corrective statements.”). If an explanation was present the sections were scored as having been completed with integrity. To determine the percentage of integrity for any given observation the number of sections completed were divided by three (i.e., by the total number of components). For example, if one component was scored as not being completed with integrity but the other two were, the percent integrity would 66.7% (2/3 = .667*100 = 66.7%).
RESULTS

Praise to Correction Ratio

Figure 1 displays each participant's percentage of integrity for the approval ratio by session. Figure 2 displays the participants’ number of praise statements per session, while Figure 3 displays the participants’ number of corrective statements per session. Teacher 1 accepted another position in a different school after completing the peer-observation phase of the study; as such, he did not participate in, nor did he have data collected for, the return to baseline phase.

Table 1 presents each teacher’s mean number of praise statements made for each phase. Table 2 presents each teacher’s mean number of corrective statements made for each phase. Table 3 presents each teacher’s mean number of praise statements made per corrective statement made for each phase. Table 4 presents each teacher’s mean percentage of integrity for the approval ratio for each phase.
Figure 1. Multiple Baseline of Percentage of Integrity for Approval Ratio. The figure presents the percentage of integrity for the approval ratio by session across teachers. The dotted lines mark the mean percentage of integrity for the phase. The numbered arrows specify when and which teachers the participant observed. Arrows 1-4 refer to Teachers 1 through 4. Arrow 5 marks when the seventh grade math teacher was observed. Arrow 6 refers to a sixth grade general education teacher being observed, and arrow 7 refers to an eighth grade reading teacher being observed.
Figure 2. Multiple Baseline of Praise Statements. The figure presents the number of praise statements by session across teachers. The dotted lines mark the mean number of praise statements for the phase. The numbered arrows specify when and which teachers the participant observed. Arrows 1-4 refer to teachers 1 through 4. Arrow 5 marks when the seventh grade math teacher was observed. Arrow 6 refers to a sixth grade general education teacher being observed, and arrow 7 refers to an eighth grade reading teacher being observed.
Figure 3. Multiple Baseline of Corrective Statements. This figure presents the number of corrective statements by session across the teachers. The dotted lines mark the mean number of praise statements for the phase. The numbered arrows specify when and which teachers the participant observed. Arrows 1-4 refer to teachers 1 through 4. Arrow 5 marks when the seventh grade math teacher was observed. Arrow 6 refers to a sixth grade general education teacher being observed, and arrow 7 refers to an eighth grade reading teacher being observed. The 35 is the value of a data point off the graph for Teacher 3.
Table 1

*Mean Number of Praise Statements Per Observation by Phase*

<table>
<thead>
<tr>
<th>Participant</th>
<th>Baseline</th>
<th>Peer-Observation</th>
<th>2nd Baseline</th>
<th>Weekly Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher 1</td>
<td>1.9</td>
<td>5.6</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Teacher 2</td>
<td>4.9</td>
<td>3.1</td>
<td>-</td>
<td>7.0</td>
</tr>
<tr>
<td>Teacher 3</td>
<td>2.9</td>
<td>9.6</td>
<td>9.0</td>
<td>-</td>
</tr>
<tr>
<td>Teacher 4</td>
<td>8.7</td>
<td>8.7</td>
<td>6.5</td>
<td>-</td>
</tr>
<tr>
<td>Average</td>
<td>4.6</td>
<td>6.7</td>
<td>7.8</td>
<td>7.0</td>
</tr>
</tbody>
</table>

Table 2

*Mean Number of Corrective Statements Per Observation by Phase*

<table>
<thead>
<tr>
<th>Participant</th>
<th>Baseline</th>
<th>Peer-Observation</th>
<th>2nd Baseline</th>
<th>Weekly Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher 1</td>
<td>8.4</td>
<td>7.3</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Teacher 2</td>
<td>3.9</td>
<td>4.6</td>
<td>-</td>
<td>3.5</td>
</tr>
<tr>
<td>Teacher 3</td>
<td>10.4</td>
<td>10.9</td>
<td>9.1</td>
<td>-</td>
</tr>
<tr>
<td>Teacher 4</td>
<td>7.2</td>
<td>2.0</td>
<td>1.8</td>
<td>-</td>
</tr>
<tr>
<td>Average</td>
<td>7.5</td>
<td>6.1</td>
<td>5.5</td>
<td>3.5</td>
</tr>
</tbody>
</table>
Table 3

*Mean Number of Praise Statements per Corrective Statement by Phase*

<table>
<thead>
<tr>
<th>Participant</th>
<th>Baseline</th>
<th>Peer-Observation</th>
<th>2nd Baseline</th>
<th>Weekly Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher 1</td>
<td>0.2</td>
<td>0.8</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Teacher 2</td>
<td>1.3</td>
<td>0.7</td>
<td>-</td>
<td>2.0</td>
</tr>
<tr>
<td>Teacher 3</td>
<td>0.3</td>
<td>0.9</td>
<td>1.0</td>
<td>-</td>
</tr>
<tr>
<td>Teacher 4</td>
<td>1.2</td>
<td>4.4</td>
<td>3.5</td>
<td>-</td>
</tr>
<tr>
<td>Average</td>
<td>0.7</td>
<td>1.7</td>
<td>2.3</td>
<td>2.0</td>
</tr>
</tbody>
</table>

Table 4

*Mean Percentage of Integrity for the Approval Ratio by Phase*

<table>
<thead>
<tr>
<th>Participant</th>
<th>Baseline</th>
<th>Peer-Observation</th>
<th>2nd Baseline</th>
<th>Weekly Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher 1</td>
<td>7.6%</td>
<td>35.7%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Teacher 2</td>
<td>47.1%</td>
<td>20.8%</td>
<td>-</td>
<td>54.6%</td>
</tr>
<tr>
<td>Teacher 3</td>
<td>16.9%</td>
<td>35.6%</td>
<td>47.0%</td>
<td>-</td>
</tr>
<tr>
<td>Teacher 4</td>
<td>46.7%</td>
<td>88.1%</td>
<td>75.6%</td>
<td>-</td>
</tr>
<tr>
<td>Average</td>
<td>29.6%</td>
<td>45.1%</td>
<td>61.3%</td>
<td>54.6%</td>
</tr>
</tbody>
</table>

**Teacher 1.** During the initial baseline, Teacher 1 averaged 1.9 praise statements per observation (range, 0 to 3) and 8.4 corrective statements (range, 4 to 13). During the peer observation phase, Teacher 1 averaged 5.6 praise statements (range, 0 to 12) and 7.3 corrective statements (range, 1 to 21). During the second baseline, Teacher 1’s approval ratio integrity averaged 7.6% (range, 0% to 19.5%), and during the peer observation phase it was 35.7% (range, 0.0% to 100%).
**Teacher 2.** During the initial baseline, Teacher 2 averaged 4.9 praise statements (range, 0 to 13) per observation and 3.9 corrective statements (range, 0 to 13). During the peer observation phase, Teacher 2 averaged 3.1 praise statements (range, 0 to 14) and 4.6 corrective statements (range, 0 to 11). Due to the decrease in praise statements and increase in corrective statements Teacher 2 entered a weekly feedback condition. During the weekly feedback condition, Teacher 2 averaged 7.0 praise statements (range, 1 to 19) per observation and 3.5 corrective statements (range, 1 to 7). During the second baseline, Teacher 2’s approval ratio integrity averaged 47.1% (range, 0% to 100%), during the peer observation phase it was 20.8% (range, 0% to 100%), and during the weekly feedback condition his integrity averaged 54.6% (range, 27.78% to 100%).

**Teacher 3.** During the initial baseline, Teacher 3 averaged 2.9 praise statements (range, 0 to 5) per observation and 10.4 corrective statements (range, 1 to 35). During the peer observation phase, Teacher 3 averaged 9.6 praise statements (range, 2 to 19) and 10.9 corrective statements (range, 6 to 20) during this phase. During the return to baseline phase, Teacher 3 averaged 9.0 praise statements (range, 2 to 19) per observation and 9.1 corrective statements (range, 3 to 19). During the second baseline, Teacher 3’s approval ratio integrity averaged 16.9% (range, 0% to 100%), during the peer observation phase it was 35.6% (range, 7.4% to 77.8%), and during the withdrawal phase her integrity averaged 47.0% (range, 4.2% to 100%).

**Teacher 4.** During the initial baseline, Teacher 4 averaged 8.7 praise statements (range, 0 to 19) per observation and 7.2 corrective statements (range, 1 to 19). During the peer observation phase, Teacher 4 averaged 8.7 praise statements (range, 4 to 14) and 2.0 corrective statements (range, 0 to 5) during this phase. During the withdrawal phase,
Teacher 4 averaged 6.5 praise statements (range, 3 to 13) per observation and 1.8 corrective statements (range, 0 to 5). During the second baseline Teacher 4’s approval ratio integrity averaged 46.7% (range, 0.0% to 100%), during the peer observation phase it was 88.1% (range, 33.3% to 100%), and during the withdrawal phase her integrity averaged 75.6% (range, 20.0% to 100%).

**Expectation Setting**

Figure 4 displays each participant's percentage of integrity for expectation setting. Table 5 presents each teacher’s mean percentage of integrity for expectation setting for each phase.
Figure 4. Multiple Baseline of Percentage of Integrity for Expectation Setting. The figure presents the percentage of integrity for expectation setting by session across teachers. The dotted lines mark the mean percentage of integrity for the phase. The numbered arrows specify when and which teachers the participant observed. Arrows 1-4 refer to teachers 1 through 4. Arrow 5 marks when the 7th grade math teacher was observed. Arrow 6 refers to a 6th grade general education teacher being observed, and arrow 7 refers to an 8th grade reading teacher being observed.
Table 5

Mean Percentage of Integrity for Expectation Setting by Phase

<table>
<thead>
<tr>
<th>Participant</th>
<th>Baseline</th>
<th>Peer-Observation</th>
<th>2nd Baseline</th>
<th>Weekly Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher 1</td>
<td>66.7%</td>
<td>80.0%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Teacher 2</td>
<td>52.5%</td>
<td>50.0%</td>
<td>-</td>
<td>62.2%</td>
</tr>
<tr>
<td>Teacher 3</td>
<td>61.7%</td>
<td>83.0%</td>
<td>78.6%</td>
<td>-</td>
</tr>
<tr>
<td>Teacher 4</td>
<td>52.5%</td>
<td>68.8%</td>
<td>68.0%</td>
<td>-</td>
</tr>
<tr>
<td>Average</td>
<td>58.3%</td>
<td>70.4%</td>
<td>73.3%</td>
<td>62.2%</td>
</tr>
</tbody>
</table>

**Teacher 1.** During the baseline phase, Teacher 1 averaged setting 66.7% (range, 0% to 100%) of the expectations possible. During the peer observation phase, he averaged setting 80% of expectations possible (range, 60.0% to 100%).

**Teacher 2.** During the baseline phase, Teacher 2 averaged setting 52.5% (range, 20% to 100%) of the expectations possible. During the peer observation phase, he averaged setting 50.0% of expectations possible (range, 20% to 80%). During the weekly feedback condition, Teacher 2 averaged setting 62.2% of expectations possible (range, 40% to 100%).

**Teacher 3.** During the baseline phase, Teacher 3 averaged setting 61.7% of the expectations possible (range, 20% to 80%). She averaged setting 83.0% of expectations possible during the peer observation phase (range, 40% to 100%). During the second baseline, Teacher 3 averaged setting 78.6% of expectations possible (range, 40% to 100%).
**Teacher 4.** During the baseline phase, Teacher 4 averaged setting 52.5% of the expectations possible (range, 20% to 100%). During the peer observation phase, she averaged setting 68.8% of expectations possible (range, 40% to 100%). During the second baseline, Teacher 4 averaged setting 68.0% of expectations possible (range, 30% to 100%).

**Referrals**

Figures 5, 6, 7, and 8 display the referrals written by day for Teacher 1, 2, 3, and 4 respectively. Across the baseline phase, Teacher 1 averaged 2.4 referrals per day (range, 0 to 5), while during the peer observation phase he averaged 1.9 (range, 0 to 5). Teacher 2 averaged 2.6 referrals per during baseline (range, 0 to 11), 3.0 during the peer observation phase (range, 0 to 10), and 1.9 during his weekly feedback condition (range, 0 to 8). Teacher 3 averaged 0.7 referrals per day during the baseline phase (range, 0 to 3), 1.94 during the peer observation phase (range, 0 to 6), and 1.4 during the withdrawal phase (range 0 to 4). Teacher 4 averaged 1.6 referrals per day during baseline (range, 0 to 5), 1.8 during the withdrawal phase (range, 0 to 5), and 1.2 during the withdrawal phase (range 0 to 4).
Figure 5. Referrals Written by Teacher 1. This figure demonstrates the number of referrals written by Teacher 1 by day. The dotted lines represent the mean number of referrals written per day for each phase.

Figure 6. Referrals Written by Teacher 2. This figure demonstrates the number of referrals written by Teacher 2 by day. The dotted lines represent the mean number of referrals written per day for each phase.
Figure 7. Referrals Written by Teacher 3. This figure demonstrates the number of referrals written by Teacher 3 by day. The dotted lines represent the mean number of referrals written per day for each phase.

Figure 8. Referrals Written by Teacher 4. This figure demonstrates the number of referrals written by Teacher 4 by day. The dotted lines represent the mean number of referrals written per day for each phase.
Interobserver Agreement

28.3% of observations conducted by the research assistant were also observed by the lead experimenter to assure there is no observer drift (34/120 sessions). The average IOA was 95%. For the session in which IOA was 66%, the two observers were too far apart from each other while observing. Due to this distance when quiet praise or corrective statements were made, only one observer was able to hear and record them. To rectify this issue during all future observations, the experimenter and research assistant stayed within two meters of each other and held their clipboards close to their bodies to mask what and when they were recording.

Teacher Observation Accuracy

One observation by each teacher participant was conducted simultaneously with a researcher to evaluate the levels of accuracy in their scoring. Accuracy was determined based on the final frequency count in the number of praise statements, corrective statements, activity transitions, and CHAMP expectations set, and the assumption that the difference in these frequency counts represented disagreements. Based on these calculations, accuracy of the teachers’ scoring ranged from 50% (Teacher 2) to 75% (Teacher 4).

Intervention Integrity

Teachers were asked to complete observations 15 to 20 minutes in length once per week during the intervention phase. Based on the start and end times recorded on the observation sheets the length of observations ranged from 9 to 16 minutes, with the average nearing 12 minutes (11.75 minutes). Every teacher completed their observation every week during the peer observation phase, and every major section was completed on
each observation form except one, in which no praise nor corrective statements were recorded. However, while no tallies were recorded, the teacher noted "No real positive or corrective - kids just did what asked!" As such, this section and the observation was scored as having been completed with integrity. Overall, the intervention integrity of the independent variable for this study was 100%.

**Intervention Acceptability**

The participants of the study were asked to complete an anonymous, online post-experiment survey to determine their view of the intervention's acceptability. All questions utilized a Likert scale from one to five followed by free-response answer where they were prompted (but not required) to explain their answers. For the Likert responses the "1" represented "not at all" or "never", and "5" represented "very much" or "every time”.

When asked how much the participants enjoyed being observed by another teacher, scores averaged 4.0 out of 5 (range, 3/neutral, to 5/very much). When asked how much they liked observing another teacher scores averaged 4.0 (range, 3 to 5). When asked how much observing another teacher helped their teaching scores averaged 4.33 (range, 4 to 5). When asked whether the benefits of observing another teacher outweighed the amount of time it took, the scores averaged 4.0 (range, 2 to 5). When asked how often the participants talked with the teacher they observed about their performance scores averaged 3.0 (range, 1 to 4). When asked how often those who observed spoke to the participant about their performance scores averaged 2.33 (range, 1 to 3).
DISCUSSION

This study sought to evaluate the impact of peer observations on an observing teachers’ implementation integrity for two interventions representing critical components of a Positive Behavioral Interventions and Supports (PBIS) system. Four teachers completed weekly observations of peer teachers during which they recorded the number of praise and corrective statements made by their peer teachers, and recorded the number and type of expectations set by their peer teacher for the students. Results suggest that peer observations increased intervention integrity for three of four participants. That is, three of the four participants increased their ratio of praise to corrective statements and expectation setting. Additionally, based on the post-experiment survey, the participants enjoyed the observation process, believed it helped their teaching, and believed its benefits outweighed the time it required.

Major Findings

The improvement in the approval ratio from baseline to the intervention phase ranged from 18.7% to 41.4% and averaged an integrity increase of 29.4%. For two of the three teachers who increased their approval ratio integrity, this was accomplished by increasing the number of praise statements made. During baseline, Teacher 1 averaged less than two praise statements per session, which increased to approximately six praise statements during the peer observation phase. Teacher 3 averaged less than three praise statements per sessions during baseline, which increased to close to ten during the observation phase. Additionally, for both teachers, the first session after their first observation of a peer-teacher resulted in a greater number of praise statements than had ever been demonstrated during any session while in baseline. The third teacher who saw
an increased approval ratio integrity did so by decreasing their rate of corrective statements. During the peer-observation phase Teacher 4 decreased from just over seven corrective statements per session to two corrective statements per session, a 72% reduction. Additionally, while the mean number of corrective statements only decreased by one for Teacher 1, and slightly increase for teacher 3 during the peer-observation phase (an increase of 0.5 corrective statements), both saw their trend of increasing use of corrective statements reverse once they began conducting peer observations.

While Teacher 2 saw a decrease in his intervention integrity on average for the approval ratio, a noticeable downward trend in his number of corrective statements, and an upward trend in their number of praise statements, was observed across the intervention phase. When including his performance decrease (a 26% decrease from baseline to intervention on average) the average change in integrity for the approval ratio across all four participants was an increase of 15.5%.

For expectation setting integrity improvements of 13.3% to 21.3% were observed, and averaged 17.0% between the baseline and peer-observation phase. The most notable performance change was demonstrated by Teacher 3. During baseline Teacher 3 never achieved 100% integrity and averaged about 62% integrity. During the peer-observation phase, this average increased to 83%, and she achieved 100% integrity in four out of ten sessions.

Again, Teacher 2 demonstrated decrease in integrity on average from baseline to the peer-observation phase, although it was less than a 3% decrease. When including his performance, the average increase from baseline to the peer-observation phase for expectation setting across the teachers was 12.1%.
Anecdotally, Teacher 3 appeared to acquire a strategy Teacher 2 utilized as a result of the peer observation process. Teacher 2 integrated the CHAMPs expectation into the slides he presented to his class. Teacher 3 began to do the same thing after observing Teacher 2. These slides seemed to act as a prompt for Teacher 3 to vocally explain the expectations, which assisted her in achieving the impressive increases in expectation setting. This result aligns with VanHouten and Sullivan’s (1975) study in which it was demonstrated that the provision of environmental prompts markedly improved teachers’ rate of praise. While Teacher 2 posted his expectations via the slides he did not consistently vocally explain them to his students (a requirement outlined in the professional development at the beginning of the school year).

The secondary dependent variable used to evaluate whether increasing integrity in these behavior management interventions had an impact on student behavior was the number of office referrals written by the participants. Across the study there was relatively little change in these rates; each teacher’s average fluctuated by about one referral written per day across each phase. However, it is worth noting that three out of four of the teachers did see an increase in the average number of referrals they wrote per day during the peer-observation phase. It is possible the intervention made the teacher’s attend to students’ misbehaviors at a higher rate (which might also explain why some teacher’s number of corrective statements increased) thus increasing the number of documented incidents of misbehavior.

The Relation Between the Current Findings and Previous Research

In Alvero and Austin’s (2004) study of the observer effect, only 22% of behaviors improved when the participants were given information about the behaviors of interest
(i.e., written behavior pinpoints), while 94% improved after they scored a 5-minute video of a confederate’s performance on these behaviors. In this study, after undergoing training in accordance with Noell et al.’s (1997) study (i.e., participants were given a verbal description of the intervention, and a rationale for its use) participants averaged implementation integrity of 29.6% (range, 7.6% to 47.1%). While engaging in the peer observation process, this improved to average 45.1% (range, 20.8% to 88.1%). This increase in performance provides additional evidence of the effects observing another has on the observer, and expands the literature by including evidence of the observer effect in the school setting with teachers as participants.

Additionally, these findings provide further support as to the efficacy of using structured observation tools, and adds general education teachers to the populations with which they are effective. Previously, Englert and Sugai (1983) demonstrated that greater performance improvements were yielded when pre-service teachers used structured observations tool when compared with those who used an open-ended observation process (i.e., when they were told what to attend to, but not given a specific form to complete when conducting their observations).

Mortenson and Witt (1998), and Codding, Feinberg, Dunn, and Pace (2006) found that teachers who received relatively infrequent performance feedback (weekly and biweekly, respectively) improved their implementation integrity. The findings of this study demonstrate that engaging in brief (9 to 15 minute), relatively infrequent (weekly) observations also improved intervention implementation integrity.

The previous success demonstrated in the research on weekly feedback is what led the current experimenter to utilize it when Teacher 2 saw a decrease in his
intervention implementation integrity. This study does provide some additional support for the use of weekly performance feedback. That is, weekly feedback did increase Teacher 2’s integrity from the peer-observation phase (the phase immediately preceding it), however this was only an 8% increase in comparison to his average integrity during the initial baseline phase. Ultimately, Teacher 2’s performance was highly variable throughout the study which make these outcomes difficult to interpret.

The Implications of this Study on Schools and Teacher Training

For a school beginning to implement a PBIS system, this intervention may offer a cost-effective means to increase teachers’ implementation integrity for two of its primary components: setting expectations for students, and utilizing procedures to increase and maintain behaviors which meet these expectations. As an added benefit, teachers enjoyed the intervention and thought it was worth the time commitment it necessitated.

This intervention required each teacher to attend a half-day professional development (3 hours in length); one fifteen-minute meeting to reiterate the observation process and have the observation schedule explained; and approximately 48 minutes of their time (on average) to complete the observations. On the post-experiment survey teachers rated this time commitment as being worthwhile in relation to the benefits it yielded, and reported they enjoyed observing their peers, being observed, and felt the observations were beneficial.

Anecdotally, all but one teacher had no problems and made no complaints about completing the observations. Each teacher utilized their plan hour (an hour in which they have no students they use to create and plan lessons) to complete the observations, except for Teacher 4. Instead, Teacher 4 asked on multiple occasions whether she needed to
complete the observation that week. The experimenter always remind her that for experimental reasons she would need to complete it within the week, but that she could withdraw at any time. She said she would not withdraw and would complete the observation if someone could cover her class while she went and observed. The experimenter acted as a substitute on three occasions (approximately 20 minutes on each occasion) in order for her to complete each observation. Interestingly, Teacher 4 had the greatest performance improvements across the participants.

The experimenter used two days to develop the half-day professional development, one half-day to deliver the professional development, two hours to create the teacher observation form, and substituted in classes on six occasions (three times for Teacher 4 and once for Teacher 1, 2, and 3 when they were scheduled to observe a peer with a coinciding plan hour) for approximately 20 minutes so that teachers could complete their observations. In total this represents a time commitment of three days. To bring this intervention to scale in the school, the experimenter would need to substitute in a classroom approximately 22 times (assuming once for every teacher). Since each occasion of substituting would be approximately 20 minutes in length, the total time commitment for this intervention would increase to approximately four days of work (substituting would represent about one full day of time). Given that not every teacher in the school would need the intervention, and it may not be necessary to observe peers with coincidental plan hours to yield the benefits demonstrated in this study, it is likely the overall time commitment would be less.

For the geographic area in which the school in this study is located, hiring a substitute teacher to cover every teacher represents an additional cost of less than $150.
This is far less than the cost to hire a consultant for one day, which is especially significant when considering a consultant might only work with four or five teachers in that time and hiring the substitute for one day would be to help 20+ teachers.

While the current system did increase integrity across the teachers, only two of the participants consistently implemented at least one of the behavior management interventions above 80% integrity as a result. If one considered implementing a multi-tiered support system for teachers as suggested by Sanetti and Collier-Meek (2015) (that is, a system of interventions which increase in intensity until the teacher achieves the desired behavior change), the observation system used in this study may represent an effective first-tier intervention because of its ease of implementation. While not all teachers will achieve desired performance as the result of a first-tier intervention, these interventions still serve to reduce the financial burden a school would normally incur to hire a consultant or coach, especially when considering only teachers in need of more time-intensive, intrusive interventions would receive this help.

It is also possible that an observation system could be created to follow a multi-tiered support system structure to achieve the desired outcomes without ever needing additional external resources. Again, the current intervention might represent an initial intervention put into place in a multi-tiered support system. This could then be followed by more intensive interventions for teachers who do not show adequate progress towards the desired outcomes. That is, more intensive observation systems may be used, including more frequent observations, observations longer in duration, meeting with the peer to discuss the observation, coupling the observation process with additional training, or a combination of these until performance has met the mastery criterion.
Limitations

Using an approximation for the desired 3:1 approval ratio represents a more nebulous measure of intervention integrity in comparison to the measures utilized in other integrity research. In the related literature, a discrete list of behaviors or behavior pinpoints was utilized, often accompanied by specific discrete antecedents (e.g., when a student refuses to comply). In the current study, praise and corrective statements could take many forms and might be occasioned by many different student behaviors. While examples of when to use praise statements and what praise statements should look like were included in the pre-baseline professional development sessions, it would be impossible to cover every instance in which these might be utilized. Due to the limited training that the teachers received, it is possible that stimulus control was not properly established and, as such, the praise behaviors were not evoked at the desired rate.

An additional limitation of this study was the secondary dependent variable used to measure student behavior change. It seems that office referrals did not act as a sensitive enough measure to demonstrate these changes, as referrals represent misbehaviors which are so great that they warrant office staff intervention. Most student misbehaviors are not this severe (e.g., interrupting a teacher/peer, refusing to comply, getting out of one’s seat, etc.). As such, in the future it would be worthwhile to examine in-class behaviors of students, and evaluate how the most common misbehaviors are affected by increasing the integrity of a teacher’s approval ratio and expectation setting.

Another limitation of the present study was the collection of expectation setting data. Expectations were recorded any time a change in activities took place (e.g., moving from the warm-up to lecture, or from taking notes to a group activity). Unfortunately,
because observations were 15 to 20 minutes in length, it was common that there was no change in activity, and as such there were no data collected on expectation setting. For this reason, the amount of data collected on expectation setting was greatly limited.

There was no formal assessment to determine if teachers understood what a corrective or praise statement was, nor if they knew the expectations they should be delivering and when they should be delivering them. However, interobserver agreement data were collected with each teacher which might serve as a rough proxy measure of their understanding. Agreement ranged from as low as 50% to as high as 75%. While Sasson and Austin (2005) demonstrated that the accuracy of the observer does not necessarily affect the behavior change in the observer, it is worth noting the teacher with the lowest accuracy (Teacher 2) saw declines in performance on average while the teacher with the highest accuracy (Teacher 4) saw some of the greatest increases in integrity. It should be acknowledged that there was no point-to-point correspondence when calculating these accuracy scores, as the teachers’ observation forms did not separate their scoring of praise and corrective statements by intervals. Thus it is possible the researcher and teacher never agreed, and their final frequency count regarding the number of praise and corrective statements was only coincidental.

Several environmental variables also may have affected the impact of the peer observation system on teacher behavior. For example, Teacher 2 experienced illness during the study, which affected his performance. Leadership and principal attrition during the study (the school had three different principals between August and December) affected teacher attrition and student behavior schoolwide. To address the issue of teacher attrition in the seventh grade team (that is, the science and math teachers
leaving their positions after the study began), participants observed a sixth-grade general education teacher and an eighth-grade reading teacher instead. These two teachers were selected for observation because of the high integrity with which they delivered the approval ratio and CHAMPs expectations, and because they expressed comfort with having others come and observe them. It is possible that observing teachers of non-seventh grade students may have had a differential impact on participants' teaching performances.

**Future Research**

Future studies might train teachers to identify a specific problem-behavior that they would like to reduce for their students. This could then be used as a targeted antecedent for providing praise. Having just one antecedent for teachers to focus on would help make training on how/when to praise more specific, which would likely aid in establishing stimulus control. Targeting a specific student problem-behavior would also help with creating an effective secondary dependent variable to measure student behavior changes as a result of increases in the teacher’s intervention integrity. Using this approach would lend itself to a multiple-baseline study in which the effects of intervention could be tested across behaviors. That is you could examine whether training to praise the correct behaviors in the presence of a specific misbehavior (e.g., praising the silent student while another is talking out of turn) results in the generalization of praising appropriate behaviors in the presence of other misbehaviors (i.e., beginning to praise students remaining in seats when their peers are inappropriately moving around the classroom).
It would also be interesting to study what the effect on the observer is when observing peers delivering the targeted intervention with high versus low integrity levels. For example, Teacher 1 saw praise rates immediately increase after observing the seventh grade math teacher who regularly demonstrated 100% integrity (17 of 19 sessions observed) of the approval ratio. Additionally, Teacher 2’s praise rates did not begin increasing until after observing the sixth- and eighth grade teachers who regularly demonstrated integrity at or near 100%. His first two observations were of Teacher 3 and Teacher 4 who were still in baseline and averaging integrity rates below 20% and 50% respectively.

Another variable warranting future study would be the amount of time spent observing (i.e., the frequency and/or the duration of observations). In this study, weekly ten to fifteen-minute observations appeared to be effective in increasing intervention integrity, but might biweekly be as effective? Perhaps two observations per week would achieve more rapid acquisition and greater change more quickly, or completing the four observations that the participants did in this study in one day would have the greatest impact. There are many different observation schedules that should be examined so the most effective system can be identified.

Since this study looked to examine the impact observing has on the observer, participants were encouraged to keep the data they collected confidential. The hope was that this would eliminate/reduce peer feedback and the effects it might have on intervention integrity (based on the result of the post experiment survey, it appears they spoke with each other very little about the observations they completed). While observing alone was effective, perhaps the outcomes demonstrated in this study could be enhanced
if peer-feedback was introduced. DiGennaro, Martens, and Kleinmann (2007) showed dramatic improvements in integrity can be realized when adding a negative reinforcement contingency in which a feedback meeting is canceled if the teacher demonstrates perfect integrity. A similar system utilizing peer-meetings might also appeal to teachers who would then receive feedback as needed, but could also reduce the time commitment to the intervention by achieving high levels of integrity.

Future studies should look to examine the long-term effects engaging in a peer observation system has on integrity once the peer-observations have concluded. Unfortunately, in the present study, only two teachers entered the return to baseline/withdrawal condition. Teacher 3 saw her approval ratio integrity increase by about 16% during the return to baseline, while Teacher 4 saw her performance drop off approximately 19% from the peer observation phase. These outcomes were reversed in regards to expectation setting. Teacher 3 demonstrated a decrease of about 2% during the withdrawal phase for their expectation setting integrity, while Teacher 4 saw an increase of 2%.

Finally, it would also be worthwhile to examine how coupling high-quality training with this peer-observation process affects levels of integrity. Training participants until they delivered the interventions at 100%, removing the training, and then implementing the peer-observation process for teachers whose integrity drops below a threshold may offer an alternative path to achieving and/or maintaining the desired performance.
CONCLUSION

The present study provides evidence that a peer-observation system can be utilized to increase teacher’s intervention integrity. It supplies additional evidence for the effects of observing another on the observer’s performance (i.e., the observer effect), and lends support to the use of structured observation tools.

The peer observation system utilized in this study represents a practical approach to increasing intervention integrity that may easily be implemented on a broad scale in nearly any school system and structure. It also presents an opportunity for future research, including how a peer model, the schedule of observation, and peer feedback might impact a teacher’s implementation integrity.
REFERENCES


Codding, R. S., Feinberg, A. B., Dunn, E. K. and Pace, G. M. (2005), Effects of
immediate performance feedback on implementation of behavior support plans.


doi:http://dx.doi.org.libproxy.library.wmich.edu/10.1300/J075v26n03_01


through peer observation and observation system technology. *Teacher Education and Special Education*, 6(1), 7-17. doi:10.1177/088840648300600102


Komaki, J., Barwick, K. D., & Scott, L. R. (1978). A behavioral approach to
doi:http://dx.doi.org.libproxy.library.wmich.edu/10.1037/0021-9010.63.4.434


doi:10.1300/J075v24n04_01


http://doi.org/10.1901/jaba.2010.43-47


Van Houten, R., Sullivan, K. (1975). Effects of an audio cueing system on the rate of

doi:10.1901/jaba.1975.8-197


doi:10.1901/jaba.1997.30-693
Appendix A

Informed Consent Document
You have been invited to participate in a research project titled "The Effects of Work Conditions on Teacher Performance." This consent document will explain the purpose of this research project and will go over all of the time commitments, the procedures used in the study, and the risks and benefits of participating in this research project. Please read this consent form carefully and completely and please ask any questions if you need more clarification.

What are we trying to find out in this study?
The purpose of this study is to investigate the effects of work conditions on teacher behaviors.

Where will this study take place?
The study will take place in Comstock Northeast Middle School.

What is the time commitment for participating in this study?
Time commitment will vary depending on the number of peer observations one conducts during the 2017-2018 academic year. This may be as little as 0 and as many as 12, with each being 15-20 minutes in length.

What will you be asked to do if you choose to participate in this study?
Any participant will be asked to complete their normal job roles, as well as complete feedback forms that will be delivered to the teacher observed (if you are asked to complete a peer observation).

What information is being measured during the study?
There will be two primary measures in this study, the ratio of teacher praise/positive statements to corrective feedback, and the percentage of expectations set by the teacher before every classroom activity in regards to the conversations students may have, how students can ask for help, what the activity is the students will be engaging in, how students may move around the class during the activity, and what it should look like for a student to show they are participating in the activity. Information regarding number of student disruptions and referrals will also be collected.

What are the risks of participating in this study and how will these risks be minimized?
Having one’s own work performance evaluated by others may cause slight discomfort. Meetings to discuss these evaluations will take place frequently in order to help reduce this discomfort.
As in all research, there may be unforeseen risks to the participant. If an accidental injury occurs, appropriate emergency measures will be taken; however, no compensation or additional treatment will be made available to you except as otherwise stated in this consent form.

What are the benefits of participating in this study?
It is possible that receiving this intervention will result in higher rates praise for students, and that with increased praise and expectation setting that students will have improved behavior and thus improve a teachers’ quality of life. You may also benefit from knowing that you are helping to advance science to create a better work environment for other teachers in the future. The results from this study may also help schools to choose techniques that have been demonstrated to be effective at increasing teacher rates of praise and expectation setting.

Are there any costs or compensation associated with participating in this study?
Participants will be compensated $20 for completion of the study. There are no costs associated with participating in this study.

Who will have access to the information collected during this study?
All information obtained in this study will remain strictly confidential. Your name will not appear on any papers on which this information is recorded, nor will you be identified in public presentations on the study.

What if you want to stop participating in this study?
You can choose to stop participating in the study at any time for any reason. You will not suffer any prejudice or penalty by your decision to stop your participation. You will experience NO consequences either professionally or personally if you choose to withdraw from this study. The investigator can also decide to stop your participation in the study without your consent.

Should you have any questions prior to or during the study, you can contact the primary investigator, Denise Ross at (269) 387-4925 or denise.ross@wmich.edu. You may also contact the Chair, Human Subjects Institutional Review Board at 269-387-8293 or the Vice President for Research at 269-387-8298 if questions arise during the course of the study. This consent document has been approved for use for one year by the Human Subjects Institutional Review Board (HSIRB) as indicated by the stamped date and signature of the board chair in the upper right corner. Do not participate in this study if the stamped date is older than one year.

I have read this informed consent document. The risks and benefits have been explained to me. I agree to take part in this study.

______________________________  ________________________________
Participant’s Name and Signature  Date
Appendix B

Behaviors Warranting a Referral
<table>
<thead>
<tr>
<th>Category of Behavior</th>
<th>Notes</th>
<th>Teacher Managed</th>
<th>Community Service</th>
<th>Lunch Detention</th>
<th>ISS</th>
<th>OSS</th>
<th>Administrator's Involvement/Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Donut</td>
<td>Out-of-seat behavior</td>
<td>Teacher redirect; failure to comply is in violation of behavior contract</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Donut</td>
<td>Out-of-seat behavior (getting into teacher's personal space)</td>
<td>Teacher redirect; failure to comply is in violation of behavior contract</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Donut</td>
<td>Entering the wrong classroom intentionally</td>
<td>Teacher redirect; failure to comply is in violation of behavior contract</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Donut</td>
<td>Intentionally opening the door to the wrong gender bathroom</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Documentation Only</td>
<td>Incidents: writing/drawing drawings, using phone in class, leaving 2nd floor, false code</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Firearms Possession/Handgun</td>
<td>Firearm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Firearms Possession/Other than Handgun</td>
<td>Firearm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fost/Drug</td>
<td>The manufacture, cultivation, use, or distribution of any drug, narcotic, controlled substance</td>
<td>Acknowledgement and written language contract must be signed</td>
<td></td>
<td></td>
<td>5-10+</td>
<td>Contact Police, Long-term Suspension</td>
<td></td>
</tr>
<tr>
<td>Presentation/Prose/Politics</td>
<td>Cultural bias in test administration</td>
<td>Teacher redirect; written and verbal language contract must be signed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presentational language/perturbations</td>
<td>Threats of violence between students</td>
<td>Teacher redirect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presentational language/perturbations</td>
<td>Threats of violence between students (including swatting)</td>
<td>Teacher redirect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presentational language/perturbations</td>
<td>Intimidating or insulting in test administration</td>
<td>Teacher redirect; written and verbal language contract must be signed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presentational language/perturbations</td>
<td>Inappropriate behavior in test administration</td>
<td>Teacher redirect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Documentation Only</td>
<td>Incidents: writing/drawing drawings, using phone in class, leaving 2nd floor, false code</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other weapons possession</td>
<td>Non-gun turned into office without prompting cancellation brought to school</td>
<td>Teacher immediately contacts office of student with item</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other weapons possession</td>
<td>Knives (blade shorter than 3&quot;)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other weapons possession</td>
<td>Knives (blade longer than 3&quot;)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical violence w/o injury</td>
<td>Low intensity, inappropriate physical contact/shove moves</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical violence w/o injury</td>
<td>High intensity, inappropriate physical contact (fighting, shoving, hitting, fighting with or without object)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical violence w/ injury</td>
<td>Low intensity, inappropriate physical contact (fighting, shoving, hitting, with or without object)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical violence w/ injury</td>
<td>High intensity, inappropriate physical contact (fighting, shoving, hitting, with or without object)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Racioethnic intimidation</td>
<td>Threats or intimidation of an individual or group based on race, religion, color, national origin, or other personal characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Racial slurs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sexual Harassment</td>
<td>Unwanted sexual behavior in violation of school rules; soliciting of an individual</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sexual Harassment</td>
<td>Unwanted sexual behavior in violation of school rules; soliciting of an individual</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Theft/property damage</td>
<td>Thievery against a personal item or person with value less than $10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Theft/property damage</td>
<td>Thievery against a personal item or person with value more than $10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Theft/property damage</td>
<td>School/property items with value less than $10 (tissue, paper clips)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Theft/property damage</td>
<td>School/property items with value more than $10 (tissue, paper clips)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Theft/property damage</td>
<td>Weapons</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Theft/property damage</td>
<td>Property damage to school property (souvenirs, furniture, electronics, bathroom stall, mirrors, computers)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Threat of violence (gun, knives, etc.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Threat of violence (blunt object)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tobacco possession, etc.</td>
<td>Possession of</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tobacco possession, etc.</td>
<td>Use/possession of</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*ISS = In-School Suspension, OSS = Out-of-School Suspension*
Appendix C

Teacher Observation Data Sheet
Observation Form

<table>
<thead>
<tr>
<th>Date:</th>
<th>Start Time:</th>
<th>End Time:</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Praise/Positive Interactions</th>
<th>Corrective Feedback</th>
</tr>
</thead>
</table>

Ratio of Positive to Corrective (e.g., 3:1, 2.5:1, 5:1):

<table>
<thead>
<tr>
<th>Number of Transitions (e.g., warm up to free write, etc.):</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
</table>

Place an X in the box when an expectation is set for an activity, leave it blank otherwise.

<table>
<thead>
<tr>
<th>C = Conversation</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>H = Help</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A = Activity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M = Movement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P = Participation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Total**

Notes:________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
Appendix D

Experimenter Observation Data Sheet
# Observation Form

**Date:**

**Start Time:**

**End Time:**

<table>
<thead>
<tr>
<th>Praise/Positive Interactions</th>
<th>Corrective Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td><strong>Sum</strong></td>
<td><strong>Sum</strong></td>
</tr>
</tbody>
</table>

**Ratio of Positive to Corrective (e.g., 3:1, 2.5:1, 5:1):**

---

**Number of Class Transitions (e.g., warm up to free write):**

1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐

Place an X in the box when an expectation is set for an activity, leave it blank otherwise.

<table>
<thead>
<tr>
<th>C = Conversation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>H = Help</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A = Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>M = Movement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>P = Participation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Total</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Notes (record on the back if needed):

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________
Appendix E

Post-Experiment Survey
Post-Experiment Survey

Participant #_______________ Date_______

1. How much did you like **being observed** by another teacher?
   
   1 2 3 4 5

   Not at all       Very much

   **Explain your ranking:**

2. How much did you like **observing** another teacher?
   
   1 2 3 4 5

   Not at all       Very much

   **Explain your ranking:**

3. How much do you believe **observing** another teacher helped your teaching?
   
   1 2 3 4 5

   Not at all       Very much

   **Explain your ranking:**

4. Do you believe the benefits outweigh the amount of time you spent observing?
   
   1 2 3 4 5

   Not at all       Very much

   **Explain your ranking:**

4. How often did you talk to the person you observed about their performance?
   
   1 2 3 4 5

   Never       Every Observation
Explain your ranking:

4. How often did the person who observed you talk to you about your performance?

1  2  3  4  5

Never  Every Observation

What barriers to successful teaching did you face during the Fall semester?

What facilitated successful teaching during the Fall semester?

What do you think could be improved in this peer observation process for the future?

Thank you!
Appendix F

HSIRB Approval Letter
Date: November 13, 2017

To: Denice Ross, Principal Investigator
    Garrett Warrillow, Student Investigator for dissertation
    Mya Hernandez, Student Investigator

From: Amy Naugle, Ph.D., Chair

Re: HSIRB Project Number 17-08-15

This letter will serve as confirmation that the changes to your research project titled “Peer Observations, Feedback, and Teacher Integrity” requested in your memo received November 13, 2017 (to add compensation of $20 cash for participation; to revise consent document to reflect this change) have been approved by the Human Subjects Institutional Review Board.

The conditions and the duration of this approval are specified in the Policies of Western Michigan University.

Please note that you may only conduct this research exactly in the form it was approved. You must seek specific board approval for any changes in this project. You must also seek reapproval if the project extends beyond the termination date noted below. In addition if there are any unanticipated adverse reactions or unanticipated events associated with the conduct of this research, you should immediately suspend the project and contact the Chair of the HSIRB for consultation.

The Board wishes you success in the pursuit of your research goals.

Approval Termination: August 30, 2018