Fall 2019

Is the Written Component of Tootling Effective? A Comparison of the Group Contingency to a Comparison Writing Procedure

James Derieux
University of Southern Mississippi

IS THE WRITTEN COMPONENT OF TOOTLING EFFECTIVE? A COMPARISON OF THE GROUP CONTINGENCY TO A COMPARISON WRITING PROCEDURE

by

James R. Derieux

A Thesis
Submitted to the Graduate School,
the College of Education and Human Sciences
and the School of Psychology
at The University of Southern Mississippi
in Partial Fulfillment of the Requirements
for the Degree of Master of Arts

Approved by:

Dr. Evan Dart, Committee Chair
Dr. Daniel Tingstrom
Dr. Keith Radley

_____	_____	_____
Dr. Evan Dart	Dr. Sara Jordan	Dr. Karen S. Coats
Committee Chair	Director of School	Dean of the Graduate School

December 2019

Published by the Graduate School

ABSTRACT

Previous tootling literature has demonstrated positive effects of the intervention when implemented in settings from elementary school (e.g., Skinner, Cashwell, & Skinner, 2000) to high school (e.g., Lum et al., 2019). Several studies have shown meaningful effects for the tootling intervention on increasing class-wide academically engaged behavior (AEB) and decreasing disruptive behaviors (DB) (e.g., Cihak et al., 2009). However, no studies in the current literature have examined the effects of the individual components of the tootling intervention on class-wide behaviors. The current study sought to examine the effects of the written component of the tootling intervention on class-wide levels of AEB and DB by comparing traditional tootling to a comparison writing procedure and a no-treatment control condition. Although this study demonstrated variable results, it is the first study to shed light on the importance of looking at the individual components of the tootling intervention. The results and outcomes of the study are discussed.

ACKNOWLEDGMENTS

I would like to thank my chair, Dr. Evan Dart, for his time and encouragement throughout this project. I would also like to thank Dr. Daniel Tingstrom and Dr. Keith Radley for being on my committee. I am forever grateful for all three professor's guidance during this project and throughout my time thus far in the School Psychology program at USM. Secondly, I would like to thank the high school, its teachers, and its students who participated in my project. I appreciate them dedicating their time and energy in order to make this project a reality. Additionally, I would like to thank my colleagues who helped with data collection throughout the duration of this project. Thank you to Stefanie Schreiber, Taylor Ben, Kenneth Ross, Caitlyn Weaver, and Halley Blanchard for helping to make this project run smoothly and successfully. I could not have done it without them.

DEDICATION

I would like to dedicate this project to my family with whom I would not have made it to this point in my life without. First, I would like to thank my parents, Robert Derieux and Katie Maldonado, for their love and support throughout all of my endeavors. I would like to thank my grandparents for setting such a strong example of how far hard work, determination, faith, and happiness can get you in life. I would like to thank my sisters for always believing in me. I also want to thank all of my extended family for being a constant source of encouragement. In addition, I would like to thank all of friends, especially, Elizabeth Lown, Lynda Hayes, and my cohort members for their friendship and support throughout our time in the School Psychology program at USM.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER I- INTRODUCTION

Disruptive behaviors (DB) can be a major distraction to teachers and students in the classroom setting. Lane (2007) found that students exhibiting disruptive behaviors in the classroom inhibit other students from learning and prevent the classroom teacher from being able to teach. Teachers also report great difficulty in implementing classroom management procedures that could be used to reduce or mitigate students' disruptive behavior (Hoglund, Klingle, & Hosan, 2015). Thus, it is imperative to identify effective and efficient methods for managing the behavior of students who exhibit behavioral difficulties, so they do not impede their academic success or the learning of their peers.

Researchers have conducted a large number of studies aimed at determining which classroom management procedures facilitate an enriched learning environment in the classroom (Emmer & Sabornie, 2015). In order to meet the needs of students at-risk for behavioral problems, both the psychology and education communities have made it their mission to shift toward making the school an agent for behavior change by way of three-tiered models of positive behavioral interventions and supports (PBIS; Horner and Sugai, 2000; Lewis and Sugai 1999). Within a multi- tiered system of support (MTSS), academic and behavioral instruction and intervention are combined in order to provide all students with the proper supports needed to be successful. The MTSS framework is a three- tiered framework in which each tier represents a different level of intensity of interventions and supports. Tier 1 represents universal academic and behavioral supports available to all students across all classroom settings. Tier 2 represents more intensive supports, in which few students in need of more intensive supports in addition to the support received in tier 1, receive more individualized instruction and intervention. In

Tier 3, select individuals still in need of additional support to the instruction and interventions provided in tier 1 and tier 2 receive the most intensive academic and behavioral supports. To qualify for services beyond that Tier 1 level of support there are three criteria that must be met by schools to assess their student's needs for additional supports. These criteria include, scores on a statewide reading assessment, student's post intervention progress toward meeting state-wide educational standards, and proper documentation indicating a student's underachievement is not due to the lack of the school's attempt to provide adequate instruction or intervention (Gamm, Casserly, Uro, Walston, Hall, et al., 2012). Thus, the MTSS framework is a preventative measure that fits well within the intended PBIS model of support that schools are moving toward to aid at- risk students.

Lane (2007) furthered the research conducted by Horner and Sugai (2000), by using an existing PBS model to research the relationship between academic success and behavioral problems in middle school students. Lane (2007) evaluated this on each level of the three-tier model. Sixty-eight students at an inclusionary middle school in Tennessee participated in the study in which they were assessed to determine the effects of three-tiered interventions to improve academic success for those at risk for emotional and behavioral disorder (EBD). The findings concluded that the tier 2 and 3 students' academic achievement increased significantly, in the way of their knowledge of effective study skills as compared to regular school practices (Lane, 2007). Based on these findings, although it has been proven that targeted intervention of this kind works with all age groups, it is important that targeted interventions for students with and at risk for EBD be implemented while children are in elementary school. According to Kazdin

(1993), the elementary level setting is the best window of opportunity for targeted interventions to be successful in working with and preventing the development of EBD (Kazdin, 1993). These interventions have opened the door for more research into highly efficient and effective methods in which to decrease problem behavior in the classroom setting.

Group Contingencies

Group contingencies are an efficient approach to help combat problem behaviors in the classroom setting. Over the past four decades, research has proven group contingencies to be effective interventions for reducing student's problem behaviors in the classroom setting (Litoy & Pumroy 1975; Theodore, Bray, Kehle, & Dioguardi, 2004). There are three group contingencies that have been shown to be effective at decreasing disruptive behaviors in the classroom setting. These group-contingency systems are called dependent, independent, and interdependent. Dependent group contingencies consist of reinforcement being contingent upon the performance of an individual or select group of students. If the individual or select group of students engage in the target behavior then access to reinforcement is provided, if the target behavior is not emitted then access to reinforcement is not provided. Independent group contingencies consist of reinforcement being contingent upon the performance of each individual student's performance. Therefore, rather than access to reinforcement being provided to an entire group of students, each student that performs the target behavior will receive access to reinforcement. Interdependent group contingencies consist of reinforcement being contingent on the performance of an entire group of students. Access to reinforcement for the entire group of students is contingent upon them meeting a pre-

determined criterion. Once the criterion is met the entire group of students will receive access to reinforcement.

Gresham and Gresham (1982) examined the efficacy of the three group-contingency systems on disruptive behavior in the classroom. Disruptive behavior was defined as verbal and physical aggression (hitting; pushing; using imperatives), out of seat, talking out of turn, inappropriate laughing, and property destruction (e.g., throwing objects). In this study, a withdrawal design consisting of eight experimental conditions, two baselines, two dependent, two independent, and two interdependent group contingencies were implemented. In baseline, observers coded the frequency of student's disruptive behaviors and no contingencies were in place. Following baseline, the interdependent group contingency was implemented in which the class was divided in two teams and were instructed that the team with the fewest instances of disruptive behavior would receive reinforcement, as long as disruptive behaviors did not exceed five instances. In the next phase, the dependent group contingency was implemented in which the two students who were the most disruptive during baseline were assigned as team captains to the two teams previously established. Reinforcement in this condition was contingent upon the disruptive behaviors of the team captains. The team captain with the fewest instances of disruptive behavior, as long as they did not exceed five instances, would earn reinforcement for their team. Lastly, the independent group contingency was implemented in which the students were competing with one another for access to reinforcement. The student or students with the fewest instances of disruptive behavior, that did not exceed five instances, received access to reinforcement. All four phases were then repeated in the same manner for verification of effects.

Results of this study indicated that the interdependent and dependent group contingency systems were the most effective at decreasing disruptive behaviors. It is probable that interdependent and dependent group contingencies were more effective than the independent group contingency because they incorporated group cooperation in which the students were observed praising and reprimanding their peer's performance. Whereas, during the independent group contingency there was a thinner schedule of reinforcement available to the class due to the same students repeatedly engaging in low enough levels of disruptive behaviors to access reinforcement (Gresham and Gresham, 1982).

In 2017, Maggin, Pustejovsky, and Johnson published a meta-analysis on the use of group contingencies in the classroom targeting students with problem behaviors to examine their efficacy and determine whether group contingencies are an evidence-based intervention to address student behavior in the classroom. The results of the analysis indicated that there is support for group contingency research, citing 40 studies with 148 cases meeting What Works Clearinghouse standards, including the criteria needed for treatments to be considered evidence based (Kratochwill, Hitchcock, Horner, Levin, Odom, Rindskopf, & Shadish, 2013). The authors pointed out that group contingencies are effective at targeting both individual and group behaviors. Group contingency interventions targeting individual performance accounted for approximately 63.6% of interventions, which most commonly occurred in general education classrooms. The median grade level for these interventions was fourth with 75.4% of students being male. Further demographics were under-reported with less than 60% of studies reporting the demographics of the target individuals including such items as, ethnicity, academic

achievement, disability ruling, and primary language. The majority of the studies included in the analysis, approximately 85%, used group contingencies based upon interdependent procedures.

Consistent with the data reported for interventions targeting individual behaviors, 55.8% of students in the interventions targeting group behaviors were male and less than 60% of studies reported the demographics of the target individuals. In addition, third grade was the median grade level for classroom- based studies. The primary dependent variables included in these studies were more frequently associated with disruptive behaviors rather than academic engagement, 55% (k = 22) and 15% (k = 5) of studies respectively, with the remaining 33% (k = 13) of studies including results for both variables. Furthermore, the analysis analyzed effect sizes for both disruptive behaviors and academically engaged behavior, finding that intervention outcomes generally led to improvements in student behaviors around two standard deviations above baseline levels. In group contingency studies targeting both individual and group behaviors, 88% utilized teachers as the primary implementer. In addition, the majority of the studies reported data for procedural fidelity and social validity, with outcomes commonly demonstrating high fidelity, as well as, high implementer ratings indicating the procedures were both useful and easy to implement (Maggin, Pustejovsky, & Johnson, 2017). This literature is important to consider when determining which type of group contingency might be most effective for evaluating the effects on a given independent or dependent variable, as well as, a particular population.

The previous studies support the use of group contingency interventions in the classroom setting and provide several key factors beneficial to the format of the current

study. First, this literature indicated that the use of group contingencies led to large effect sizes for both disruptive behaviors and academically engaged behavior signifying improvements in student behaviors in the classroom. Secondly, the literature indicated that the majority of studies utilizing group contingencies had high rates of fidelity and teacher social validity with teachers implementing the intervention. Lastly, this literature supports the authors choice in using an interdependent group contingency in that the majority of group contingency interventions are reportedly based on interdependent procedures.

One such interdependent group contingency that has proven effective is the Good Behavior Game (GBG) (Barrish, Saunders, & Wolf, 1969). The GBG is an intervention in which teachers divide their students into two or more teams and record the amount of disruptive behaviors for each team during a period of time. The team with the fewest amount of disruptive behaviors have the opportunity to receive reinforcement. A prime example of the effectiveness of the GBG is a study completed by Lannie and McCurdy (2007). During implementation of the GBG, an unknown criterion was set that had to be met in order for participants to receive reinforcement. Whichever team met the criterion at the end of the game, meaning they had the least amount of disruptive behaviors, would receive reinforcement, while the other team received nothing (Lannie and McCurdy, 2007). The results of the study indicated that after implementation of the game, students' on-task behavior increased substantially while disruptive behaviors decreased significantly. According to Tingstrom, Sterling-Turner, and Wilczynski (2006), other studies have shown that interdependent group contingencies, such as the GBG, are effective in managing students' behavior in the classroom because the inclusion of such

activities can be easily incorporated as reinforcers themselves versus providing tangibles alone. In addition, problem behavior was shown to maintain or increase when the students engaging in the behavior were receiving attention from their peers. Teachers have reported liking these type of activities as they tend to keep them from singling out students with excessive problem behaviors (Tingstrom, Sterling-Turner, & Wilczynski 2006); however, one major limitation of the GBG is peer influence. Students may exhibit unnecessary peer pressure on the students who are not engaging in the desired prosocial behaviors. In addition, there is potential for students who are already exhibiting problem behaviors to get frustrated with the procedures and variability in reinforcement delivery (Tingstrom, Sterling-Turner, & Wilczynski 2006). Thus, interest in the general effectiveness of interdependent group contingencies led researchers to investigate alternative arrangements.

Tootling

Although tootling is a relatively new interdependent group contingency intervention, so far, all of the research conducted indicates that it is an effective method for addressing disruptive behavior in the classroom setting, primarily at the elementary school level. Tootling is an approach in which students are taught to report their peers' prosocial behaviors, rather than tattling on their peers. During the tootling intervention, students are provided with notecards on which to record their prosocial behaviors or "tootles" and have the option to place their notecards into a container on the teacher's desk. An example of a "tootle" is "John completed his independent classwork quietly", whereas, a non- example is "Sarah would not stop talking to me while I did my independent classwork." Students have the opportunity to work toward a common goal

which is predetermined by the teacher and researcher. The common goal is selected by evaluating trends in the baseline data of either academically engaged behavior (AEB), disruptive behaviors (DB), or both. At the end of class, the teacher counts the number of appropriate "tootles" turned in to the container and records the number of "tootles" in a public place where the students can see it. If the goal is met within the time-frame laid out by the teacher (e.g., a week), then the class gains access to reinforcement. Results of implementing tootling in elementary school settings has shown that when students gain access to reinforcement for reporting peer's prosocial behaviors, disruptive behaviors decrease. In addition, the implementation of the tootling intervention increased the amount of opportunities for positive reinforcement to be delivered for appropriate behaviors (Cihak, Kirk, & Boon, 2009).

A study by Skinner, Cashwell, and Skinner (2000) trained 28, fourth-grade students in a general education classroom to identify and report their peer's prosocial behaviors through the use of the tootling intervention with a public posting component. An ABAB withdrawal design was used to assess the effects of the interdependent group contingency. A public posting component was implemented to remind students to report their peer's prosocial behaviors as well as to indicate the class' progress toward the goal. In baseline, students were told to report their peer's prosocial behaviors; however, no goal was in place and no feedback was given in the form of public posting. During intervention sessions, notecards were taped to the student's desks and students were told to record their peer's prosocial behavior on the notecards and place them in a box located on the teacher's desk. The group goal in the first implementation of tootling was 100 "tootles." Each morning the researcher told the class the number of "tootles" they

appropriately reported the day before and then hung a poster board on the wall with a ladder drawn on it and an icon to indicate where the class was with their goal. The top of the ladder represented the class goal. Once the goal was met, the students received a day off from tootling in order to gain extra recess time. After the initial implementation of the tooting intervention, a withdrawal phase was implemented in which procedures were the same as in baseline. After withdrawal, the intervention was re-implemented, during which time the goal and public posting component were reinstated. This time the goal was increased to 150 “tootles”.

Results of the study indicated low rates of tootling during the initial baseline phase, followed by variable rates of tootling throughout intervention ending with an upward trend. Upon implementation of the withdrawal phase, there was an immediate decrease in rates of “tootles” with stable, near zero rates continuing throughout the phase. In the reimplementation of the tootling phase, an immediate increase in rates of “tootles” occurred followed by variable, high rates throughout the remainder of the study. Although this study was successful at increasing students’ rates of tootling, there were several limitations pointed out to be addressed in future research. First, there was a lack of experimental control due to the variable rates of responding during baseline and intervention. Internal validity was threatened due to the school principal implementing a group- oriented punishment procedure that could have affected student’s performance during the study. Secondly, the scope of external validity was narrowed due to its small sample size of teachers as well as its use of only one dependent variable. Therefore, suggested future research included expanding on the populations with which the intervention is implemented, particularly the number and diversity of teachers utilized.

Additionally, the literature should be expanded by evaluating the effects of the tootling intervention on more dependent variables (Skinner, Cashwell, & Skinner, 2000).

Cashwell, Skinner, and Smith (2001), sought to replicate and extend the study by Skinner and colleagues (2000) by implementing tootling with a public posting feedback component in a second- grade classroom to increase and maintain tootling rates. Findings of this study were consistent with the previous findings in that students in elementary school could be taught to identify and report their peers' prosocial behaviors. Results of this study showed that the tootling intervention with a public posting component lead to an increase in the number of prosocial behaviors (i.e., "tootles") reported; however, results did indicate that following the withdrawal phase, there was an immediate decrease in "tootles" reported, signifying that the initial high rates of tootling could have been caused by novelty effects. Therefore, students may not have maintained their behaviors unless other methods were implemented to reinforce the tootling intervention. Thus, it was recommended for future research to evaluate methods by which to maintain student's attentiveness to peers' prosocial behaviors after interventions, such as tootling, are withdrawn. Additionally, evidence for social validity was found as the participating classrooms continued implementation of the tootling procedures for the remaining two and a half weeks of the school year. It was presumed that this was due to the procedural variables of the tootling intervention being specific. For example, the intervention specifically targeted positive behaviors. This maintenance could also have occurred due to process variables, such as soliciting teacher and student participation in developing specific procedures (Cashwell, Skinner, & Smith, 2001).

Cihak, Kirk, and Boon (2009) found that tootling meets the criteria for PBS and was the first study to look at the effects on student's behaviors. These criteria include; "engaged all students in a common effort to reach a goal, [tootling] recognized and identified positive behaviors, students were taught to identify positive behaviors and encouraged to perform such behaviors, and [tootling] decreased inappropriate classroom behaviors" (Cihak, Kirk, & Boone, 2009, pg. 276). In this study, Cihak, Kirk, and Boone (2009) examined the effects of tootling on a third-grade inclusion classroom using an ABAB design. Findings indicated decreases in disruptive behaviors for students both with and without disabilities across both intervention and reimplementation. Within the last three days of implementation of tootling, there were no occurrences of problem behavior observed. Results also indicated that tootling increased positive peer pressure by increasing student's encouragement of their peers to perform prosocial behaviors in order to gain access to reinforcement, thus decreasing disruptive behavior. Several limitations were noted in this study. First, a more comparative-treatment design, such as an alternating treatment design, could have been used instead of an ABAB design to parse out the individual effects of the tootling components and allowed them to be assessed further. Secondly, a small sample size consisting of only one third-grade classroom was used. In addition, social validity was limited due to only one teacher completing the rating scale. Lastly, the effects of the interdependent group contingency could not be differentiated from the tootling intervention. Therefore, the decrease in disruptive behaviors cannot be specifically attributed to one or the other. The authors noted that future studies should look at fading group contingency procedures to differentiate the

effects of the group contingency and the tootling intervention (Cihak, Kirk, & Boon 2009).

Due to the lack of literature on interventions encouraging students to acknowledge and report their peers' prosocial behaviors, Sherman (2012) sought to compare two such interventions. The interventions compared in this study were positive peer reporting (PPR) and positive peer reporting combined with tootling. Although both of these interventions encourage students' reporting of peer prosocial behaviors, they differ in several of their components. PPR involves students reporting prosocial behaviors publicly, whereas, tootling involves students reporting prosocial behaviors privately. During implementation of PPR, students can report prosocial behaviors only during a particular time of the day. On the other hand, in tootling, students can report prosocial behaviors throughout the day. Lastly, PPR is typically used to increase the prosocial behaviors of an individual whereas tootling is typically used to increase the prosocial behaviors of an entire group.

In this study, two multiple baseline designs were used to assess PPR and PPR plus tootling interventions effects on decreasing inappropriate behaviors and increasing appropriate behaviors of four participants across four different general education classrooms in an elementary school setting. The researchers hypothesized that adding in the tootling component to the PPR would decrease inappropriate behaviors more than either PPR or tootling would alone. In addition, it was hypothesized that students may feel more comfortable being able to report prosocial behaviors privately through the tootling component rather than publicly. Also, students had the option to report prosocial behaviors either at a designated time specified for PPR or throughout the day. The

opportunity for students to report prosocial behaviors throughout the day gave them time to identify more substantial and appropriate behaviors. The authors concluded that both the PPR intervention and the PPR plus tootling intervention were equally effective at decreasing inappropriate behaviors and increasing appropriate behaviors across all four participants. Two limitations were found specifically regarding the tootling component. First, the number of "tootles" was not calculated from day to day meaning the response effort for the PPR plus tootling may have been higher than for the PPR intervention alone. The authors noted that students may have found it easier to make verbal statements during the PPR intervention rather than writing tootles down when the tootling component was added. Thus, resulting in fewer tootles being reported than verbal statements. Secondly, both internal and external validity were limited due to two teachers having to restrict the use of notecards to certain times during the day as they caused a distraction for several of the students. Traditional tootling procedures indicate that notecards are to be provided at the beginning of class and students are allowed to turn them in throughout the day; however, in this study, because two teachers had to withhold the notecards until the students were ready to write their prosocial behaviors down, their distribution times differed from the other two studies, hurting procedural integrity (Sherman, 2012).

In 2015, Lambert, Tingstrom, Sterling, Dufrene, and Lynne conducted a study that produced results which were consistent with the findings of Cihak and colleagues (2009). Results indicated that implementing tootling across fourth and fifth-grade general education classrooms decreased both class-wide and individual disruptive behaviors. In addition, findings showed substantial increases in appropriate behavior that maintained

through follow-up phases during which the intervention was no longer in place (Lambert et al, 2015). A consistent limitation found in both articles was that the intervention consisted of multiple common components meaning the results cannot be attributed to any one or multiple components (Cihak et al., 2009; Lambert et al., 2015).

Lum, Tingstrom, Dufrene, Radley, and Lynne (2017) aimed to build upon previous literature by Skinner and colleagues (2000) and Cashwell and colleagues (2001) by implementing tootling with an interdependent group contingency and public posting feedback in a high school setting. Participants in this study included three general education classrooms whose teachers had reported high levels of disruptive behaviors (DB). Thus, class-wide disruptive behavior was selected as the primary dependent variable. Consultants used the *Problem Identification Interview* (Kratochwill and Bergan, 1990) to interview teachers in order to identify the three most disruptive behaviors in the classroom to be assessed throughout the study. The three disruptive behaviors identified were out of seat behavior, playing with objects, and inappropriate vocalizations. Class-wide disruptive behavior data were aggregated across the three discrete behaviors. Academically engaged behavior (AEB) was coded as the secondary dependent variable. An ABAB withdrawal design was used to evaluate the effects of the tootling intervention on decreasing disruptive behaviors and increasing AEB. Results of this study demonstrated meaningful decreases in disruptive behaviors and increases in AEB. Therefore, this study extended the tootling literature by providing evidence that the intervention is effective with an older school-aged population. In addition, social validity measures completed by the teachers indicated that all three found the intervention effective and that they would be willing to implement it again in the future.

Several limitations of this study were identified by the authors. First, external validity was limited due to all three participating classrooms being from a single high school. This is a limit to external validity because the data cannot be reliably generalized to students at other high schools without replications of this study being conducted in more high school settings. Second, treatment integrity relied heavily on self-report data from the teachers. While treatment integrity levels were relatively high at 82% or above, it is unknown what level of treatment integrity is needed for tootling to be effective. Additionally, when integrity checks were completed by the observer, performance feedback had to be given several times for multiple teachers for missing important treatment steps. The steps teachers often missed included not reviewing the tootling instructions, not reading a minimum of five tootles at the end of class, and not reviewing or updating the class's progress chart each day. The authors noted that it is unknown what components are crucial for tootling to be implemented effectively. In terms of limitations to the dependent variables, disruptive behavior data were aggregated across the three discrete behaviors identified by the teachers; therefore, there is no way of knowing which specific behaviors the intervention effected. Lastly, Lum and colleagues (2017) highlighted a limitation that was also consistent in Cihak and colleagues (2009) and Lambert and colleagues (2015), that several components make up the tootling intervention and the study's design and methodology could not attribute effectiveness to any one or multiple of these components. Lum and colleagues (2017) recommended that future research look at the individual components of tootling to examine which components are having the biggest effect on student behaviors and which are not (Lum, 2017).

In 2016, McHugh, Tingstrom, Radley, Barry, and Walker looked to assess the effectiveness of the tootling intervention on class-wide behavior as well as individual student behavior by utilizing a goal that could be reached on a daily basis rather than one takes longer, such as a weekly goal. Similar to previous literature, this study utilized an ABAB withdrawal design with a multiple baseline element across three lower-elementary school classrooms and an interdependent group contingency with a public posting component. During the study, the teacher nominated a student with high levels of disruptive behaviors to be monitored independently while still participating in the class intervention. Results for both class-wide and individual student behaviors indicated decreases in DB and increases in AEB, with moderate to large effect sizes. Effect sizes were calculated using Nonoverlap of All Pairs (NAP; Parker & Vannest, 2009). For classroom A, effect size calculations indicated a moderate effect (NAP= .92) for decreasing DB and a strong effect (NAP= .98) for increasing AEB. Effect size calculations for the student in classroom A, indicated a moderate effect (NAP=.92) for decreasing DB and a moderate effect (NAP= .92) for increasing AEB. For classroom B, effect size calculations indicated a strong effect (NAP= 1.00) for decreasing DB and a strong effect (NAP= 1.00) for increasing AEB. Effect size calculations for the student in classroom B, indicated a strong effect (NAP=1.00) for decreasing DB and a strong effect (NAP= 1.00) for increasing AEB. For classroom C, effect size calculations indicated a strong effect (NAP= 1.00) for decreasing DB and a strong effect (NAP= 1.00) for increasing AEB. Effect size calculations for the student in classroom C, indicated a moderate effect (NAP=.89) for decreasing DB and a strong effect (NAP= .97) for increasing AEB. In terms of limitations, the current study did not demonstrate a direct

comparison between the ABAB design and methodology and the effects of a daily tootling goal when compared to prior studies (Cihak et al., 2009; Lambert et al., 2015; Lum et al., 2017). In addition, as noted by previous studies, the current study was unable to attribute the results of the tootling intervention to any one or multiple components of the intervention (Cihak et al., 2009; Lambert et al., 2015; Lum et al., 2017). The authors noted that future research should look to include a technological component, such as ClassDojo, to decrease the amount of time the intervention took to implement (McHugh et al., 2016).

In an aim to extend previous literature on tootling, McHugh, Radley, Tingstrom, Dart, and Barry (2019) decided to incorporate a technology component known as ClassDojo. The ClassDojo website was used to record the student's "tootles" and project them on the board for the entire class to see. The addition of this component eliminated aforementioned issues with using notecards, providing a more efficient and feasible method for classroom teachers to implement the intervention. An ABAB design with a multiple baseline element across three fifth-grade classrooms was used to assess the effects of tootling with ClassDojo on decreasing DB and increasing AEB. Results of the study were consistent with the previous tootling literature in that the tootling intervention resulted in substantial increases in AEB and decreases in DB during implementation. Additionally, effect sizes were calculated using Tau-U (Parker, Vannest, & Davis, 2011). Effect Sizes of the intervention indicated strong effects for both decreasing DB and increasing AEB across all three classrooms. Omnibus effect sizes for all three classrooms indicate strong effects (Tau-U= 1.0) for decreasing DB, as well as, strong effects (Tau-U= .98) for increasing AEB. Overall, results of the current study indicate positive effects

for DB and AEB when ClassDojo was added to the tootling intervention. However, results of the current study could not be directly attributed to any one or multiple components of the tootling intervention, which is consistent with the previous literature (Cihak et al., 2009; Lambert et al., 2015; Lum et al., 2017). This study pointed out that future research should assess the effects of tootling with ClassDojo with younger and older groups of students (McHugh et al., 2019).

Wright (2016) extended the research conducted by Lum and colleagues (2017) by implementing a traditional tootling procedure with an interdependent group contingency and public posting feedback in a high school setting. The major difference in this study was that the primary dependent variable chosen was AEB. All prior literature on tootling (Skinner et al., 2000; Cashwell et al., 2001; Cihak et al., 2009; Lum et al., 2017) examined DB as their primary dependent variable. DB and passive off-task behavior (POT) were coded as secondary dependent variables. The focus of this study was to increase AEB and decrease DB and POT. In addition, this study aimed to increase students' awareness to the occurrence of prosocial behaviors. An A/B/B+C design across four general education classrooms was used to assess tootling and tootling with a public posting component. The authors reported substantial increases in AEB and DB across all four classrooms with moderate to very large effect sizes for both interventions. Effect sizes were calculated using Tau-U (Parker, Vannest, & Davis, 2011). For classroom A, effect size calculations indicated a very large effect for AEB (Tau-U= 1.00) and DB (Tau-U= .90) across both intervention conditions. Effect size calculations for POT indicated a moderate effect (Tau-U= .36) for the tootling alone condition and a very large effect (Tau-U= 1.0) for the tootling plus public posting condition.

For classroom B, effect size calculations for AEB indicated a large effect (Tau-U= .72) for the tootling alone condition and a very large effect (Tau-U= 1.0) for the tootling plus public posting condition. Effect size calculations for DB indicated a very large effect (Tau-U= .93) across both intervention conditions. Effect size calculations for POT indicated a moderate effect (Tau-U= .47) for the tootling alone condition and a very large effect (Tau-U= .83) for the tootling plus public posting condition. For classroom C, effect size calculations indicated very large effects (Tau-U=1.0) for AEB across both intervention conditions. Effect size calculations for DB indicated a large effect (Tau-U=.73) for the tootling alone condition and a moderate effect (Tau-U= .40) for the tootling plus public posting condition. Effect size calculations for POT indicated a large effect (Tau-U= .80) for the tootling alone condition and a very large effect (Tau-U= .97) for the tootling plus public posting condition. Lastly, effect size calculations for classroom D indicated very large effects for AEB (Tau-U= .89) and DB (Tau-U= .89) across both intervention conditions. Effect size calculations for POT indicated moderate effects across both the tootling alone condition (Tau-U= .58) and the tootling plus public posting condition (Tau-U= .50).

The present study boasted higher levels of AEB than Lum and colleagues (2017) and were consistent with the findings of Lambert et al. (2015) and McHugh et al. (2016) in the elementary school setting. In regard to the tootling plus public posting component, the authors concluded that both AEB and DB were consistent with levels in the tootling alone condition. Therefore, the increase in data for the tootling plus public posting component was not substantial enough to need the addition of public posting in future studies. Additionally, social validity measures for both students and teachers indicated

high acceptability and effectiveness. Several limitations were indicated in the study. First, systematic direct observations were only 20 minutes in length during the same time each day, limiting the researcher's knowledge on student behavior during other times of the day. Student reactivity to the researchers may have had an effect on the accuracy of the data collected as it is a small sample of overall classroom behavior. Second, all four classrooms were located in one high school, limiting external validity. More replications of the study would need to be conducted in more high school settings before the results could be generalized to other high school students. Also, procedural integrity data were not 100% for classroom A and C and treatment integrity data were not 100% across all four classrooms. Procedural integrity was not 100% for classroom A and C due to time constraints, however, this did not negatively affect student training as students in all four classrooms were trained with 100% integrity. Treatment integrity data were not 100% across all four classrooms due to the teacher in classroom B having to be retrained after not updating the class feedback chart one day. Also, there was no training provided for teachers when transitioning into the tootling plus public posting condition. Lastly, the B+C phase, tootling plus public posting, always followed the B phase, tootling alone, thus it is not known whether order effects had an effect on the results of the study (Wright, 2016).

In 2019, Lum, Radley, Tingstrom, Dufrene, Olmi, and Wright sought to further extend the tootling literature in a high school setting by implementing the tootling intervention with a randomized independent group contingency. An ABAB withdrawal design across three general education classrooms to assess the effects of tootling on decreasing DB and POT and increasing AEB. During intervention, the students were

provided with a slip of paper to “tootle” on at the beginning of class. When they were ready to turn in their slip of paper, they would tear the slip of paper down the middle, separating their name from the “tootle”. The students then turned the two slips of paper into two different containers on the teacher’s desk, one for “tootles” and one for the submitter’s name. At the end of class, the teacher pulled three “tootles” from the designated container, read them aloud, and the students who performed the prosocial behavior received access to reinforcement. The teacher then pulled two slips of paper from the submitter’s container and those two individuals received access to reinforcement for participating in the intervention. Results of this study indicated that the tootling intervention with a randomized independent group contingency was effective at increasing AEB and decreasing DB; however, percentages of POT remained relatively consistent and stable across all four phases of the study. In addition, large effect sizes were calculated for both AEB and DB, while small to moderate effect sizes were calculated for POT. Effect sizes were calculated using Tau-U (Parker, Vannest, & Davis, 2011).

For classroom A, effect size calculations indicated large effects for AEB (Tau-U= .96) and DB (Tau-U= .94) across both the initial tootling and reimplementation conditions. Effect size calculations for POT indicated a moderate effect for the initial tootling conditions and a small effect (Tau-U= .03) for the reimplementation condition. Effect size calculations for classroom B indicated large effects for AEB (Tau-U=.88) and DB (Tau-U= 1.0) across both the initial tootling and reimplementation conditions. Effect size calculations for POT indicated moderate effects across both the initial tootling (Tau-U= .37) and reimplementation (Tau-U= .20) conditions. Lastly, classroom C effect size

calculations indicated large effects for AEB (Tau-U= .93) and DB (Tau-U= .87) across both the initial tootling and reimplementation conditions. Effect size calculations for POT indicated small effects across both the initial tootling (Tau-U= .16) and reimplementation (Tau-U= .10) conditions. Social validity measures yielded high scores indicating the intervention was considered acceptable and effective for both teachers and students, which were consistent with previous literature (Lambert et al., 2015; McHugh et al., 2016).

Several limitations were indicated in this study. External validity was limited due to the study being run at only one high school as well as the cost and time associated with edible rewards. Replication of the study in more high school settings would need to be conducted before the results could be generalized to other high school students. Rewards associated with the intervention could get costly which may not be feasible for educators wanting to utilize this intervention package. Another limitation was that DB percentages across all three classrooms was not as low in baseline and withdrawal phases as it was in screening observations. Thus, it is unknown whether the tootling intervention with a randomized group contingency would be effective in classroom settings with more severe percentages of DB. Additionally, classroom observations were conducted during the 20-minute period that the teacher indicated as the most disruptive. However, additional observations to support the teacher's claim were not conducted so it is unknown to the researchers what portion of the class period was truly the most disruptive. Lastly, the effect size calculation method, Tau- U, did not accurately represent the significance of behavior change developing from the intervention. The change in behaviors appeared

more meaningful and clinically significant than indicated by effect sizes scores (Lum et al., 2019).

Several studies have evaluated the effects of tootling, but prior literature has yet to assess what specific components of tootling are responsible for the observed decreases in class-wide DB and increases in class-wide AEB. Therefore, more research into tootling is necessary in determining what is driving changes in behavior. The current study looks to extend previous literature by analyzing one specific component of the tootling intervention, written material on the notecards. Previous research has utilized the notecard component of tootling to provide students with an element of anonymity when reporting their peers' prosocial behaviors. Sherman (2012) hypothesized but did not measure, that having "tootles" reported on notecards may make students more comfortable with participating in the intervention rather than reporting prosocial behaviors publicly. Traditionally, a goal for the number of "tootles" written is determined by the primary investigator and the teacher to encourage students to report their peers' prosocial behaviors. If the goal is met within a specified timeframe (e.g. a day, a week), students have the opportunity to access a reward. Upon learning this, students typically begin to change their behavior, so their peers can catch them exhibiting positive prosocial behaviors. Thus, the aim of the current study was to address the following question, is the material written on the notecards during the tootling intervention the component that is resulting in increases in academically engaged behavior and decreases in disruptive behavior?

Purpose

This study addressed the concern raised in previous literature (e.g., Cihak et al., 2009; Lambert et al., 2015; Lum et al., 2017; McHugh et al., 2016) in regard to whether a specific component of the tootling intervention is responsible for its effectiveness by comparing tootling to a comparison writing procedure, as well as, a no-treatment condition. The purpose of this study was to determine whether or not the material written on the notecards during intervention is the component that is resulting in increases in class-wide academically engaged behavior and decreases in class-wide disruptive behavior. The high school setting was the target setting used in this study as it was proven to be an appropriate setting for investigations of tootling (Lum et al., 2017; Wright, 2016; Lum et al., 2019). It was predicted that the traditional tootling intervention would lead to an increase in peer prosocial behavior when compared to some comparison writing procedure. Additionally, it was hypothesized that both treatments would lead to an increase in academically engaged behavior, as well as, a decrease in disruptive behaviors due to a placebo effect occurring during the comparison writing procedure.

Research questions:

1) Will the implementation of traditional tootling have a meaningful increase in academically engaged behavior when compared to a comparison writing procedure?
2) Will the implementation of traditional tootling have a meaningful decrease in disruptive behavior when compared to a comparison writing procedure?
3) Will the no-treatment control condition increase academically engaged behavior and decrease disruptive behavior?

4) Will there be a distinguishable difference in notecard responding between implementation of traditional tootling, a comparison writing procedure, and a no-treatment control condition?
5) Will the traditional tootling and the comparison writing procedure be identified as socially valid class-wide interventions by teachers and students?

CHAPTER II- METHODS

Participants and Setting

Participants in this study consisted of three general- education classrooms within a public high school located in a rural region in a southeastern state in the United States. The high school services roughly 580 total students in grades 9 through 12, 58% of whom are eligible for free or reduced lunch. Additionally, the high school is staffed with 51 faculty members. The high school schedule consists of four 90- minute block periods that students attend daily.

Prior to the start of the study, approval to conduct the study was obtained by the primary investigator from the University's Institutional Review Board (Appendix A). Approval to conduct the study was then obtained by the high school principal (Appendix B). The classrooms were identified for this study based on administrative referrals of low academically engaged behavior (AEB). Upon receiving the classroom referrals, the primary investigator met with each teacher to provide an overview of the study and consent to participate was obtained (Appendix C). Each classroom underwent an initial screening observation in which data were collected for class-wide levels of AEB, DB, and POT. The criterion for inclusion in this study mandated that students exhibit AEB during less than 70% of observation sessions during a 20-minute session. Once three classrooms met the inclusionary criterion, the screening process was terminated. During the screening process, no referred classrooms were screened that did not meet criterion; however, if any referred classrooms had not meet the criterion then they would not have been included in this study and other services outside the scope of this study would have been offered upon request.

Classroom A was an English course taught during the 3rd block of the school day, classroom B was a Creative Writing course taught during the 1st block of the school day, and classroom C was an Algebra course taught during the 2nd block of the school day. Observations for all three classrooms were conducted the last 20- minutes of the 90-minute block period. A demographic information form (Appendix D) was completed by each teacher to collect information pertaining to their personal demographics, as well as, the demographic make-up of their class (refer to Table 1).

Table 1

Classroom Participant Demographic Data

Classroom	A	B	C
Teacher			
Race	Caucasian	Caucasian	Caucasian
Gender	Female	Female	Female
Number of Years Teaching	8	7	5
Highest Degree Earned	Bachelor's	Master's	Master's
Classroom			
Number of students	23	19	18
Male	15	14	11
Female	8	5	7
Race			
African American	14	4	5
Asian	0	0	0
Caucasian	5	10	12
Hispanic	4	5	1
Number of SPED Students	8	9	3
SPED classification			
ASD	1	1	0
SLD	4	7	1
EMD	1	0	0
OHI	2	1	2

Note. ASD = Autism Spectrum Disorder; SLD = Specific Learning Disability; OHI = Other health Impairment; EMD = Emotional Disability

Materials

Prior to the start of intervention, teachers participating in the study were provided with a script to guide them through the tootling intervention (adapted from Lambert, 2012 and Wright, 2016; Appendices E-I). During intervention sessions, each student was provided with two, 3 by 5-inch notecards to record their "tootles" (i.e., observed prosocial behaviors) or the information requested during the comparison writing procedure and no-treatment control condition. For discrimination to occur, the notecards for each condition were a different color. The notecard colors included green, yellow, and white. There was no significance regarding the selection of the colors used. Teachers had a small plastic container set on the corner of their desk where students could place their notecards once they were completed. Additionally, a 14 by14-inch white board was positioned at the front of the classroom on which teachers were instructed to write the class's predetermined goal and the number of appropriate notecards turned in each day. During data collection, a timer was used to ensure proper interval timing.

Problem Identification Interview (PII)

The *PII* is a questionnaire that is used to identify class wide problem behaviors (Kratochwill & Bergan, 1990; adapted from Wright, 2016). This measure includes questions regarding procedures the teacher follows when dealing with problem behavior. Prior to screening observations, the *PII* was administered to teachers to determine the disruptive behaviors that were occurring most often during class time. These behaviors were used to develop operational definitions of disruptive behaviors in this study. An example of a question on the *PII* is, "Tell me about what happens before the behavior? After the behavior occurs?". Although the *PII* is commonly cited in behavioral

consultation, its psychometric properties are unreported at this time (Zuckerman, 2005). Refer to Appendix J for an example of the *PII*.

User Rating Profile- Intervention Revised (URP- IR)

The *URP-I* (Briesch, Chafouleas, Neugebauer, & Riley-Tillman, 2013) was created to assess the factors that influence why a given intervention is implemented and whether or not use of the intervention is maintained over time. The *URP- I* is made up of 35 items and assesses five factors including, acceptability, understanding, feasibility, integrity, and personal enthusiasm. In order to address the environmental level of influence, a study was conducted in 2013 to revise the *URP-I* to include the identification of additional items. The revised version of the *URP-I*, known as the *URP-IR*, now consists of six factors: acceptability, understanding, feasibility, family-school collaboration, system climate, and system support. The *URP-IR* reports high internal consistency, with alpha coefficients for each of the factors of .95, .80, .84, .79, .91, and .72, respectively (Briesch et al., 2013). The *URP- IR* consists of 29 items based on a 6-point Likert scale ranging from 1 (*strongly disagree*) to 6 (*strongly agree*). An example of a question on the *URP-IR* is, "I understand how to use this intervention" (Briesch et al., 2013). The *URP-IR* was completed by all three teachers that implemented intervention at the conclusion of the study to assess for the overall quality of the intervention utilized. Refer to Appendix K for an example of the *URP-IR*.

Children's Usage Rating Profile (CURP)

The *CURP* (Briesch & Chafouleas, 2009) consists of 21 questions based on a 4-point Likert scale ranging from 1 (*I totally disagree*) to 4 (*I totally agree*). Similar to the *URP-IR*, higher scores on the *CURP* indicate greater student satisfaction with the

intervention. The CURP assesses three factors: personal desirability, understanding, and feasibility. The *CURP* reports high internal consistency, with alpha coefficients for each of the factors of .92, .82, and .75, respectively (Briesch & Chafouleas, 2009). An example of a question on the CURP is, “This is a good way to help students” (Briesch & Chafouleas, 2009). The *CURP* was administered to students participating in the current study to assess for student satisfaction with the intervention. A parental consent form was provided to every student who participated in the study to obtain consent to complete the rating scale. Refer to Appendix L for an example of the *CURP*. Refer to Appendix M for the parental consent form for the *CURP*.

Dependent Measures

The primary dependent variable being examined in this study was academically engaged behavior (AEB). One of the purposes of this study was to examine whether levels of AEB would be affected by implementation of both the traditional tootling and the comparison writing procedure. Throughout the study, visual analysis of percentage of intervals in which AEB occurred was used to determine when phase changes were appropriate. AEB was defined as any instance of a student “being actively involved or attending to (e.g., looking at) independent seatwork, teacher instruction, designated classroom activities, and/or engaging in task-related vocalizations with teachers and/or peers” (adapted from Lambert et al., 2015, p 418).

The secondary dependent variable in this study was disruptive behavior (DB). Disruptive behaviors included the disruptive behaviors predominately identified by teachers on the *PII* and were aggregated into a single class of behavior. The disruptive behaviors identified by classroom A and C’s teachers included off-task behavior,

inappropriate conversation, and out of seat behavior. The disruptive behaviors identified by classroom B's teacher included off-task behavior, inappropriate conversation, and rapping. Off-task behavior was defined as any instance of behavior that interferes with a student's ability to be academically engaged such as playing with objects (e.g., tapping pencil), making noises (e.g., humming), playing on a cell phone, and rummaging through a bag/backpack. Inappropriate conversation was defined as any instance of a student engaging in conversation with a peer during academic instruction; any instance of a student engaging in conversation with a peer that does not pertain to the current task demand (i.e., during individual or group work). Out of seat behavior was defined as any instance of a student leaving his/her seat without permission, sitting on top of a desk, or putting their legs/feet on top of a desk (i.e., bottom should be in the seat and feet should be on the floor). Rapping was defined as any instance of a student engaging in vocal expression involving rhythmic speech and rhyming with or without musical accompaniment. Passive off-task (POT) behavior was also examined as a secondary dependent variable. POT was defined as any instance of behavior in which a student is not actively engaged in the task demand nor exhibiting behavior consistent with operational definitions of disruptive behavior. Examples of these behaviors included a student having their head down on their desk, sleeping, and being disoriented to the task demand in place (e.g., gazing off).

A tertiary dependent variable was included and was referred to as notecard responding. Notecard responding was defined as the total number of notecards turned in during each intervention condition. Calculation of notecard responding consisted of taking a frequency count of the total number of notecards the students completed each

day, per classroom. At the end of each intervention session, the researcher counted the number of completed notecards that were collected and recorded that number on a specified data sheet. Completed notecards were defined as those that had an appropriate sentence or phrase answering the provided prompt. Notecards with incomplete thoughts, inappropriate language, or nothing written on them were deemed incomplete. Refer to Appendix N for an example of the data sheet used to record notecard responding.

Data Collection

AEB, DB, and POT were recorded by researchers each day for a 20- minute period using momentary time sampling with 10- second intervals. Momentary time sampling has shown to provide the best representation of behavior during an observation because behavior is accounted for only when it is occurring at the end of the interval, rather than at smaller intervals or throughout the entirety of the interval (Green, McCoy, Burns, & Smith, 1982; Radley, O'Handley, & LaBrot, 2015). Thus, momentary time sampling has been proven to provide less measurement error as compared to partial and whole interval recording. At the end of each 10- second interval, the observers looked up at a student and recorded whether he or she was academically engaged, engaging in disruptive behavior, or engaging in passive off-task behavior, according to the aforementioned operational definitions.

Students were observed using the individual-fixed method of group behavior assessment. The observers always started the observation with the first student in the first row on the left side of the classroom. From there, they went down each row from front to back, moving from the left side of the classroom to the right, until all students were observed. If the observers reached the last student with time remaining, then they went

back to the first student and followed the same order of observation again until the timer hit 20 minutes. Breisch, Hemphill, Volpe, and Daniels (2015) assessed the class-wide behavior assessment methods and found that observing student's individual instances of academic engagement during each interval resulted in the best approximation of mean academic engagement. Whereas, the study found that observing the class as a whole during each interval resulted in an underestimation of the mean of academic engagement (Breisch, Hemphill, Volpe, and Daniels, 2015). In 2016, Dart, Radley, Briesch, Furlow, and Cavell expanded on the previous study by conducting two analyses to assess the accuracy of eight different class-wide direct observation methods, individual- fixed, individual-random, group-fixed, group-random, and four planned activity checks varying in duration. In analysis 1, simulated data from 200 classrooms was used to compare the eight observation methods across five classroom types (very low academic engagement to very high academic engagement) to a criterion estimate. Results of analysis 1 were consistent with the findings of Breisch et al. (2015).

Both the individual-fixed and individual-random methods produced class-wide behavior estimates that were consistent with the criterion estimate. Additionally, both methods produced small absolute error with mean differences near zero compared to the criterion estimate. The error produced from both individual observation methods is hypothesized to be random and the methods could either underestimate or overestimate the exact error rate. In analysis 2, three 20- minute observations were conducted in a single classroom with the aim to back- up the results of analysis 1. Results of analysis 2, confirmed the findings in analysis 1 in that the mean difference for the individual- fixed method was +1.3, which was smaller than the absolute mean difference of 4.4. Overall,

the study concluded that six of the eight observation methods, including individual-fixed, were accurate methods of assessment of group behavior (Dart, Radley, Briesch, Furlow, and Cavell, 2016). Refer to Appendix O for an example of the student observation form.

Design

An alternating treatment design (ATD) embedded within a multiple-baseline design was used across the three participating classrooms to determine if the written component of the intervention conditions had an effect on class-wide levels of AEB and DB. The three conditions included in the study were a traditional tootling treatment, a comparison writing treatment, and a no-treatment control condition. All three classrooms began baseline observations at the same time; however, each class began intervention on staggered days of data collection in order to strengthen internal validity by accounting for any extraneous variables that could affect changes in behavior. As previously mentioned, phase changes were determined through visual analysis of AEB. Once levels of AEB were stable, intervention began. During intervention, all three classrooms were exposed to the traditional tootling intervention, comparison writing procedure, and no-treatment control condition. These three conditions were randomly alternated throughout the intervention phase. After intervention, whichever treatment had the largest effect on increasing levels of AEB in each classroom was carried out in a best treatment phase. During this phase, procedures remained the same as in intervention, but only one condition (i.e., the most effective) was implemented in each of the three classrooms. Treatment A represented the traditional tooting treatment, in which tootling was implemented in its traditional format with a reinforcement component. Treatment B represented the comparison writing procedure, in which the students were instructed to

write down two interesting things they learned that day with the same reinforcement component as treatment A. In the no-treatment control condition, treatment C, students were asked to again write down two interesting things they learned that day, without a reinforcement component in place.

Treatment conditions were determined for each class, every day, through sampling without replacement. In order to do this, the primary investigator selected pieces of paper with the letters A, B, and C on them out of a container. Each letter corresponded to the three intervention conditions. Once a letter was selected, it was not to be placed back in the container until all three conditions had been implemented once in a given classroom. Once all three conditions had been implemented in a classroom, the researcher would place the slips of paper back into the container and restart the process. This process was conducted separately for each classroom.

Procedures

Screening

First, a screening observation were conducted in each classroom to determine if they met criterion to be included in this study. During the screening procedure, teachers were instructed to conduct their typical classroom management procedures. Data regarding students' AEB, DB, and POT were collected, however, only data for AEB were assessed to determine eligibility. As mentioned previously, students had to exhibit levels of AEB less than 70% of observation intervals to meet inclusion criterion. If a classroom met criterion, the screening observation was then used as the first baseline observation of the study.

Baseline

Prior to baseline, the teachers were instructed to run class as they normally would until told otherwise by the primary investigator. During baseline, 20- minute observations were conducted, during which time, data were collected for AEB, DB, and POT for a minimum of five sessions per classroom as outlined in the What Works Clearinghouse standards for data collection (Kratochwill, Hitchock, Horner, Levin, Odom, Rindskopf, & Shadish, 2010). Training on intervention components was withheld until after baseline was completed. Completion of baseline was determined by visual analysis of AEB for each classroom.

Training

Before the first day of intervention for any classroom, a training session was held to train the teachers on the intervention procedures. The primary investigator contacted each teacher to arrange a time that would be most convenient for them to complete the training session. All three teachers opted to complete training during their designated planning period during the day. Teachers were instructed that the duration of training was dependent on them attaining mastery on each intervention condition but was expected to last approximately 45 minutes. However, all three teachers completed training in approximately 25 minutes. Each teacher was trained on the tootling treatment, the comparison writing procedure, and the no-treatment control condition within the same training session, as all three treatments were implemented throughout the intervention phase. During training, the primary investigator provided the teachers with a training script for both interventions, as well as, the no-treatment control condition. The primary investigator went through each script one at a time with the teachers. Once treatment

procedures were explained, the procedures were modelled by the experimenter and any questions or concerns the teachers had were addressed. Then, each teacher had the opportunity to demonstrate the intervention steps to the experimenter to ensure reliability and the experimenter provided corrective feedback. Once, each teacher demonstrated mastery of each training script, the intervention phase could begin. The teachers spent approximately 10 minutes at the beginning of the first day of each intervention session training the students on the intervention condition being conducted. The training script procedural integrity checklist was completed by researchers on the first day of intervention for each condition in each classroom. Once all intervention conditions had been implemented at least once in a classroom using the training script, a shortened script was provided for the remainder of the intervention phase to aid in time management; however, the script for the no-treatment control condition remained the same throughout both training and intervention.

Additionally, the primary investigator discussed with each teacher appropriate reinforcers to be used if and when students met their goal during the tootling and comparison writing procedure conditions. All three teachers chose candy as the reinforcer to be used throughout the intervention and best treatment phases of the study.

Intervention

Prior to entering the classrooms, the primary investigator selected the treatments being implemented in each classroom by pulling a slip of paper with the corresponding treatment letter from a container as previously explained. Treatment A represented the tootling condition, treatment B the comparison writing procedure condition, and treatment C the no-treatment control condition. Once the treatment was selected, the

primary investigator provided each teacher with the script corresponding to the treatment that was to be implemented in their classroom that day. If a teacher was implementing the tootling or comparison writing procedure, then a goal was determined for the class that can be reasonably achieved each week and the teacher wrote the goal on the board at the front of the classroom. If the teacher was implementing the no-treatment control condition, a goal was still determined for the class for the purpose of analyzing levels of notecard responding only. As previously mentioned, reinforcement was not in place during the no-treatment control condition. The goal was determined by the researchers each day by multiplying the number of students in each class by two because that was the maximum number of "tootles" that could be turned in by each student. Intervention was carried out until there was a clear, stable, and distinguishable effect.

Traditional Tootling

When the tradition tootling condition was being implemented, the teachers began by reading the designated script and passing out two green notecards to each student. There was a limit of two notecards that each student could turn in each day during this condition. The teachers had a plastic container on the corner of their desks for the students to deposit their completed notecards. Students were given the opportunity to turn in their notecards during transition periods of class only. Toward the end of the class period the researchers counted the number of notecards in the container containing appropriate "tootles" and reported the number to the teacher to write on the board next to the goal. During the traditional tootling condition, the primary experimenter or a trained graduate student collected data for AEB, DB, and POT in 20- minute observation sessions using momentary time sampling. Additionally, notecard responding was

recorded on a separate data sheet. Classroom implementation format remained the same throughout the study across all three classrooms.

Comparison Writing Procedure

When the comparison writing procedure was being implemented, the teachers began by reading the designated script and passing out two yellow notecards to each student. Similar to the traditional tootling treatment condition, there was a limit of two notecards that each student could turn in each day during this condition. The teachers had a plastic container on the corner of their desks for the students to deposit their completed notecards. Students were given the opportunity to turn in their notecards during transition periods of class only. Toward the end of the class period the researchers counted the number of notecards in the container containing appropriate "learned items" and reported the number to the teacher to write on the board next to the goal. During the comparison writing procedure, the primary investigator or a trained graduate student collected data for AEB, DB, and POT in 20- minute observation sessions using momentary time sampling. Additionally, notecard responding was recorded on a separate data sheet. Classroom implementation format remained the same throughout the study across all three classrooms.

No-treatment Control

A no-treatment control condition was embedded within the ATD design. When this condition was being implemented, students were provided with two white notecards. During this condition, teachers followed the corresponding script, instructing the students to write down two interesting things they learned that day; same as in the comparison writing procedure. During transition periods, students were given the opportunity to place their notecards in the container, located on the teacher's desk. The purpose of this phase

is for comparison of the tootling and comparison writing procedure conditions. During the no-treatment control, the primary investigator or a trained graduate student collected data for AEB, DB, and POT in 20- minute observation sessions using momentary time sampling. Additionally, notecard responding was recorded on a separate data sheet. Classroom implementation format remained the same throughout the study across all three classrooms.

Interobserver Agreement (IOA)

During all phases, interobserver agreement (IOA) was collected by the primary experimenter or colleagues who had been previously trained on the data collection procedures. IOA was collected for a minimum of 33% of sessions for each phases of the study, in which a secondary trained observer helped with data collection. The secondary observer had to meet a criterion of 90% agreement with the primary observer throughout the study. If the criterion of 90% agreement was not met, the observer would have been retrained to ensure accuracy and consistency. Retraining was not necessary for any of the secondary observers throughout the duration of the study. IOA was calculated by taking the total number of agreements across all the intervals for each dependent variable, divided by the total number of intervals, and multiplied by 100 (Cooper et al. 2007).

IOA for classroom A was collected during baseline for 44.44% of observations, during intervention for 33.33% of observations, and during the best treatment phase for 33.33% of observations. Total IOA across all three dependent variables was 98.78% (range= 95-100%). Average IOA for academically engaged behavior was 99.56% (range= 97.50-100%), average IOA for disruptive behavior was 98.33% (range= 92.5-100%), and average IOA for passive off-task behavior was 98.67% (range= 95-100%).

IOA for classroom B was collected during baseline for 38.46% of observations, during intervention for 33.33% of observations, and during the best treatment phase for 33.33% of observations. Total IOA across all three dependent variables was 97.60% (range= 93.89-100%). Average IOA for academically engaged behavior was 97.47% (range= 93.33-100%), average IOA for disruptive behavior was 98.01% (range= 95-100%), and average IOA for passive off-task behavior was 97.20% (range= 93.33-100%).

IOA for classroom C was collected during baseline for 36.36% of observations, during intervention for 33.33% of observations, and during the best treatment phase for 33.33% of observations. Total IOA across all three dependent variables was 98.06% (range= 93.33-100%). Average IOA for academically engaged behavior was 98.71% (range= 93.33-100%), average IOA for disruptive behavior was 97.80% (range= 93.33-100%), and average IOA for passive off-task behavior was 97.65% (range= 93.33-100%).

Procedural Integrity

Procedural integrity was assessed by researchers during 100% of training sessions and baseline observations for each classroom teacher. IOA for procedural integrity was collected during 100% of training sessions and 33.33% of baseline observations. For teacher training a checklist was created for each of the three intervention conditions to ensure teachers could implement each condition with integrity. The tootling and comparison writing procedure procedural integrity checklists both consisted of 15 steps, while the no-treatment control procedural integrity checklist consisted of 9 steps. An example of a statement that appeared on all three checklists was, “Explain the daily writing procedure”. During baseline, researchers completed a checklist consisting of 4 “yes” or “no” statements. An example of a statement that appeared on the procedural

integrity checklist was, “The primary observer did not provide feedback to the teacher or students during data collection.” The researchers calculated procedural integrity by counting the number of steps completed by the teacher and dividing it by the total number of steps on the checklist. The answer was then multiplied by 100 to get a percentage for accuracy of steps completed by the teacher. Refer to appendices Q-S for examples of the procedural integrity checklists used during training (adapted from Wright, 2016). Refer to appendix T for an example of the baseline procedural integrity checklist.

Treatment Integrity

Teacher treatment integrity was evaluated by creating a checklist of the steps the teacher followed during intervention implementation. Treatment integrity was recorded during 100% of intervention and best treatment phases across all three classrooms. To ensure accuracy, IOA for treatment integrity was collected during 33.33% of observations during intervention and best treatment phases across all three classrooms. The treatment integrity checklist consisted of 9 “yes” or “no” statements that corresponded with the steps teachers followed throughout implementation of both intervention conditions. An example of a statement that appeared on the treatment integrity checklist was, “The teacher provided the students with the appropriate colored notecard for the treatment being conducted.” Similar to procedural integrity, treatment integrity was calculated by counting the number of steps the teacher completed and dividing it by the total number of steps on the checklist. The answer was then multiplied by 100 to get a percentage for accuracy of steps completed by the teacher. Treatment integrity averaged 97.97% (range= 85.71-100%) for classroom A, 98.59% (range= 85.71-

100%) for classroom B, and 98.15% (range= 88.89-100%) for classroom C. Refer to appendix P for an example of the checklist.

Data Analysis

Two analytic methods, visual analysis and Baseline-Corrected Tau were utilized in this study. First, visual analysis was used to assess for level, variability, and trend, immediacy of effect, the proportion of data points overlapping in level, the magnitude in changes in the DV, and the stability of data patterns across all conditions of the study. These criteria were used to determine whether a functional relationship exists between the DVs, in this case AEB and DB, and the IVs, tootling, the comparison writing procedure, and the no-treatment control condition (Horner et al, 2005). In this study an ATD design embedded within a multiple baseline design was utilized. For this, data patterns were assessed using visual analysis in baseline and once the level of AEB are determined stable in level, trend, and variability, implementation of intervention in the first classroom could begin. Intervention was staggered in each classroom to limit the effects of extraneous variables on changes in behavior. During intervention, the ATD design was implemented in which all of the previous criteria were continually assessed until there was a clear and distinguishable effect in the data pattern for each DV. Upon completion of intervention, another phase was implemented in which the best intervention in each classroom was implemented alone. Again, all criteria for visual analysis was assessed throughout the phase to determine if there was a carryover effect in the data patterns from intervention indicating a functional relationship between the DVs and the selected IV.

Second, an effect size measure, Baseline-Corrected Tau, was used to determine whether or not there was an intervention effect. Baseline-Corrected Tau was chosen for use in this study due to its conservative method of calculation. Baseline-Corrected Tau scores were calculated by comparing each phase to one another across classrooms. Therefore, baseline scores were compared to scores for Phase B (intervention), baseline scores were compared with scores for Phase C (best treatment condition), and Phase B (intervention) scores were compared with scores in Phase C (best treatment condition). This resulted in an effect size value between -1 and +1. Effect sizes indicate the strength of treatment effects. If the effect size value fell between -1 and 0 this means a negative relationship existed between the treatment and the outcome variable. Whereas, if the effect size value fell between 0 and +1 this means there was a positive relationship between the treatment and the outcome variable. Calculation of effect sizes was completed through an online calculator which indicated whether or not adjustments are needed to account for baseline trends (Tarlow, 2017). There are four categories in which effect sizes were reported; small, moderate, large, and large to very large. Scores below .2 indicate small effect sizes, scores between .2 and .6 indicate moderate effect sizes, scores between .6 and .8 indicate large effect sizes, and scores at or above .8 indicate large to very large effect sizes (Vannest & Ninci, 2015).

CHAPTER III- RESULTS

Classroom A

Classroom A met screen in criterion for this study with AEB at 18.33%. During the remainder of baseline, AEB was variable with a mean of 29.35% (range= 18.33-42.5%) across 9 observation session. Initially, classroom A could have moved into the intervention phase after the first 5 sessions due to stable low levels of AEB, however, the baseline phase had to be extended across all three classrooms due to two week-long breaks in the data collection schedule. The first break occurred between the 5th and 6th session and the second occurred between the 7th and 8th session of the baseline phase. It was determined after the 9th session to go ahead and move the class into the intervention phase. Although the phase ended with an increasing trend, after analyzing the baseline levels of AEB, the primary investigator determined it appropriate to change phases as it was hypothesized that extending the baseline phase further would not cause meaningful decreases in the data. During intervention, AEB for treatment A (tootling) initially increased from baseline levels, were variable for intervention sessions 2-4 and then stabilized during the 5th session. AEB during treatment A had a mean of 47.0% (range= 36.67-58.33%). For treatment B (comparison writing procedure), levels of AEB initially increased from baseline and then had a variable, decreasing trend across the following three sessions before returning to its initial high level for the last session. The mean for AEB during treatment B was 36.33% (range= 12.5- 49.17%). Levels of AEB during the no-treatment control condition, initially decreased during the first session of intervention before stabilizing across the remaining four session of intervention. The mean for AEB during the no-treatment control condition was 34.33% (range= 23.33- 42.5%). Treatment

A was selected for the best treatment phase for classroom A due to it having the highest mean levels of AEB. During the best treatment phase, levels of AEB started out stable from intervention levels and then decreased significantly across the following two sessions. The mean for AEB during the best treatment phase was 23.06% (range= 6.7-49.17%). Due to time constraints, only three data points were able to be collected during the best treatment phase. Refer to the top panel of figure 1 for classroom A's data regarding AEB.

Levels of DB for classroom A started out high during the baseline phase at 58.33% but had a variable yet steadily decreasing trend before transitioning to the intervention phase. DB had a mean of 38.80% (range= 15.83- 58.33%). Upon moving into the intervention phase, levels of DB across all three intervention conditions remained relatively stable with none of the condition's levels of DB increasing beyond the high end of the range for baseline levels. DB for treatment A slightly increased at the beginning of the intervention phase and was variable, ending on an increasing trend. The mean for DB during treatment A was 29.67% (range= 21.67- 38.33%). For treatment B, DB increased slightly at the beginning of the intervention phase and steadily increased across the next two sessions before increasing significantly during the 4th session and then stabilizing during the 5th session. The mean for DB during treatment B was 33.84% (range= 24.17-56.7%). For the no-treatment control condition, levels of DB slightly increased during the 1st intervention phase from baseline levels and were variable throughout intervention before ending on an increasing trend. The mean for DB during the no-treatment control condition was 31.93% (range= 23.0- 45.0%). During the best treatment phase, levels of DB for treatment A initially decreased during the 1st session, increased significantly

during the 2nd session, and then decreased back to levels consistent with intervention during the 3rd session. The mean for DB during the best treatment phase was 39.17% (range= 25- 60.8%). Refer to the top panel of figure 2 for classroom A's data regarding DB.

Classroom A's levels of POT started out low during baseline at 23.33% and then were variable. Levels of POT then increased to their highest-level during session 6 and had a steadily decreasing trend over the last three sessions of the baseline phase. The mean for POT during baseline was 31.85% (range= 23.33- 43.33%). During the intervention phase, levels of POT for treatment A decreased significantly during the 1st intervention session and remained stable with the exception of the 3rd session which increased significantly before levels stabilized. The mean for POT during treatment A was 23.33% (range= 15.0- 40.0%). During treatment B, levels of POT initially decreased during the 1st session of intervention and remained stable throughout the intervention phase. The mean for POT during treatment B was 29.83% (range= 23.33- 36.67%). For the no-treatment control condition, levels of POT were particularly variable, slightly increasing during the 1st session and ending on a decreasing trend. The mean for POT during the no-treatment control condition was 33.93% (range= 15.0- 53.33%). During the best treatment phase, levels of POT for treatment A increased from intervention levels and were stable across the first two sessions before making a significant jump during the 3rd session. The mean for POT during the best treatment phase was 37.78% (range= 25.8- 61.7%). Refer to the top panel of figure 3 for classroom A's data regarding POT.

In regard to the tertiary dependent variable, notecard responding, classroom A had a mean goal of 33.07 (range= 24- 38) notecards. For treatment A, mean responding was

22.4 (range= 10- 36) notecards. For treatment B, mean responding was 25.4 (range= 11-38) notecards. For the no-treatment control condition, mean responding was 13.4 (range= 8- 21) notecards. During the best treatment phase, classroom A had a mean goal of 29.33 (range= 28- 30) notecards and mean responding of 22 (range= 14- 28) notecards. Classroom A met their goal only twice, sessions 2 and 6, both of which occurred during treatment B. Classroom A did not meet their goal at all during the best treatment phase. Refer to Figure 4 for classroom A's data regarding notecard responding.

Classroom B

Classroom B met screen in criterion for this study with AEB at 61.67%. During the remainder of baseline, AEB was variable ending with a decreasing trend. The mean for AEB during baseline was 43.33% (range= 20.83- 61.67%) across 11 sessions. During intervention, AEB for treatment A (tootling) slightly increased during the 1st session from baseline levels and had a strong increasing trend until the 5th session in which levels of AEB decreased significantly but not below baseline levels. AEB during treatment A had a mean of 60.16% (range= 43.75-87.5%). For treatment B (comparison writing procedure), levels of AEB increased significantly from baseline levels before decreasing significantly during the 2nd session and maintaining an increasing trend through the remainder of intervention. The mean for AEB during treatment B was 51.23% (range= 28.33- 67.0%). Levels of AEB during the no-treatment control condition, initially increased during the 1st session of intervention before increasing across the next two sessions and maintaining a decreasing trend across the last three sessions. The mean for AEB during the no-treatment control condition was 47.53% (range= 31.0- 59.17%). Treatment A was selected for the best treatment phase for classroom B due to it having

the highest mean levels of AEB. During the best treatment phase, levels of AEB initially increased from intervention phase levels and then had a significantly decreasing trend over the remaining two sessions of the phase. The mean for AEB during the best treatment phase was 54.44% (range= 35.83- 74.17%). Due to time constraints, only three data points were able to be collected during the best treatment phase. Refer to the middle panel of figure 1 for classroom B's data regarding AEB.

Levels of DB for classroom B started out low and remained stable throughout the baseline phase, ending with an increasing trend. During the baseline phase, the mean for DB was 21.67% (range= 15.83- 27.5%). Upon moving into the intervention phase, DB for treatment A slightly increased and then had a decreasing trend throughout sessions 2-4, before slightly increasing again during the 5th session. The mean for DB during treatment A was 22.37% (range= 5.0- 35.56%). For treatment B, levels of DB decreased significantly at the beginning of the intervention phase, before significantly increasing during the 2nd session, and then steadily decreasing throughout the remainder of the phase. The mean for DB during treatment B was 24.45% (range= 6.0- 54.17%). For the no-treatment control condition, levels of DB slightly increased during the 1st intervention session from baseline levels, steadily decreased from sessions 2-4, and then increased again during the 5th session. The mean for DB during the no-treatment control condition was 23.14% (range= 10.0- 33.0%). During the best treatment phase, levels of DB for treatment A initially increased during the 1st session and maintained a slightly increasing trend throughout the remainder of the phase. The mean for DB during the best treatment phase was 24.72% (range= 20.0- 28.33%). Refer to the middle panel of figure 2 for classroom B's data regarding DB.

Classroom B's levels of POT started out low during baseline at 22.5% and then were variable. Levels of POT increased to their highest-level during session 6 and then decreasing to their lowest level during session 7. From there the data had an increasing trend and then stabilized over the last three sessions of the baseline phase. The mean for POT during baseline was 35.0% (range= 17.5- 55.0%). During the intervention phase, levels for POT for treatment A decreased significantly during the first intervention session and were variable throughout the remainder of the phase with levels of POT increasing significantly during the 5th session. The mean for POT during treatment A was 17.48% (range= 7.5- 35.0%). During treatment B, levels of POT decreased significantly during the 1st session of intervention, maintained a decreasing trend through the 4th session, and then increased again during the 5th session. The mean for POT during treatment B was 20.43% (range= 15.0- 28.0%). For the no-treatment control condition, levels of POT decreased significantly in the 1st intervention session and remained stable for the first three sessions before developing an increasing trend across the last two sessions of the phase. The mean for POT during the no-treatment control condition was 26.85% (range= 20.83- 37.0%). During the best treatment phase, levels of POT for treatment A initially decreased significantly before developing a steadily increasing trend across the remainder of the phase. The mean for POT during the best treatment phase was 20.83% (range= 5.83- 35.83%). Refer to the middle panel of figure 3 for classroom B's data regarding POT.

In regard to the tertiary dependent variable, notecard responding, classroom B had a mean goal of 28.53 (range= 24- 40) notecards during the intervention phase. For treatment A, mean responding was 21.8 (range= 10- 32) notecards. For treatment B,

mean responding was 21.6 (range= 20- 23) notecards. For the no-treatment control condition, mean responding was 14.8 (range= 8- 24) notecards. During the best treatment phase, classroom B had a mean goal of 24 notecards (range= 14- 30) and mean responding of 14 (range= 12- 16) notecards. Classroom B did not meet their goal for either treatment throughout the intervention phase. However, they did meet their goal once during the best treatment phase for treatment A. Refer to Figure 5 for classroom B's data regarding notecard responding.

Classroom C

Classroom C met screen in criterion for this study with AEB at 26.67%. During the remainder of baseline, AEB was particularly variable ending with a decreasing trend. The mean for AEB during baseline was 33.14% (range= 15.83- 61.67%) across 13 sessions. During intervention, AEB for treatment A (tootling) initially increased before decreasing significantly during the 2nd session and then steadily increasing to the highest level of any condition during intervention during the 5th session. AEB during treatment A had a mean of 26.0% (range= 8.0- 42.5%). For treatment B (comparison writing procedure), levels of AEB increased slightly from baseline levels and had an increasing trend across the first three session before declining into a decreasing trend across the 4th and 5th session. The mean for AEB during treatment B was 29.0% (range= 18.33- 37.5%). Levels of AEB during the no-treatment control condition, increased slightly during the 1st session of intervention and was variable across the remainder of the phase with an increasing trend across sessions 3-5. The mean for AEB during the no-treatment control condition was 31.76% (range= 18.0- 40.83%). The no-treatment control condition was selected for the best treatment phase for classroom B due to it having the highest

mean levels of AEB. During the best treatment phase, levels of AEB decreased significantly during the 1st session and again slightly decreased during the 2nd session before increasing significantly during the 3rd session. The mean for AEB during the best treatment phase was 23.6% (range= 18.3- 35.83%). Due to time constraints, only three data points were able to be collected during the best treatment phase. Refer to the bottom panel of figure 1 for classroom C's data regarding AEB.

Levels of DB for classroom C were variable throughout the baseline phase and ended on a decreasing trend. During the baseline phase, the mean for DB was 38.27% (range= 27.5- 50.83%). Upon moving into the intervention phase, levels of DB for treatment A slightly increased during the 1st session of the intervention phase and then made a significant jump during the 2nd session to the highest level of any intervention condition before decreasing to variable levels across the last three sessions. The mean for DB during treatment A was 42.23% (range= 26.0- 69.0%). For treatment B, levels of DB increased initially and were stable across all five sessions. The mean for DB during treatment B was 41.58% (range= 38.0- 45.0%). For the no-treatment control condition, levels of DB slightly increased during the 1st intervention session from baseline levels, were variable, and then had an increasing trend across the last three sessions. The mean for DB during the no-treatment control condition was 42.57% (range= 32.5- 53.33%). During the best treatment phase, levels of DB for the no-treatment control condition increased significantly during the 1st session and maintained a decreasing trend throughout the remainder of the phase. The mean for DB during the best treatment phase was 47.22% (range= 33.33- 57.5%). Refer to the bottom panel of figure 2 for classroom C's data regarding DB.

Classroom C's levels of POT were variable and stable until the 13th session when levels increased significantly. The mean for POT during baseline was 28.59% (range= 15.8- 56.67%). During the intervention phase, levels of POT for treatment A initially decreased during the 1st intervention session and were variable throughout the remainder of the phase ending on a decreasing trend. The mean for POT during treatment A was 31.6% (range= 20.83- 53.0%). During treatment B, levels of POT significantly decreased during the 1st session of intervention, remained stable across sessions 2-4, and then increased again during the 5th session. The mean for POT during treatment B was 29.94% (range= 25.0- 38.89%). For the no-treatment control condition, levels of POT decreased significantly in the 1st intervention session, continued to increase through the 3rd session, and then decreased again across the 4th and 5th sessions. The mean for POT during the no-treatment control condition was 25.67% (range= 18.33- 40.0%). During the best treatment phase, levels of POT for the no-treatment control condition initially increased and were variable throughout the phase. The mean for POT during the best treatment phase was 29.18% (range= 24.2- 32.5%). Refer to the bottom panel of figure 3 for classroom C's data regarding POT.

In regard to the tertiary dependent variable, notecard responding, classroom C had a mean goal of 30.73 (range= 15- 36) notecards during the intervention phase. For treatment A, mean responding was 26.8 (range= 18- 30) notecards. For treatment B, mean responding was 24.8 (range= 16- 34) notecards. For the no-treatment control condition, mean responding was 21.6 (range= 7- 32) notecards. During the best treatment phase, classroom C had a mean goal of 24 notecards (range= 20- 28) and mean responding of 6 (range= 4- 10) notecards. During intervention, classroom C met their

goal twice during implementation of treatment A, sessions 8 and 13. Classroom C also met their goal during implementation of treatment B twice, sessions 3 and 14. During the no-treatment control condition, classroom C turned in enough appropriate notecards to meet a goal during session 11 but did not turn in enough to meet a goal during any of the three best treatment sessions. Refer to Figure 6 for classroom C's data regarding notecard responding.

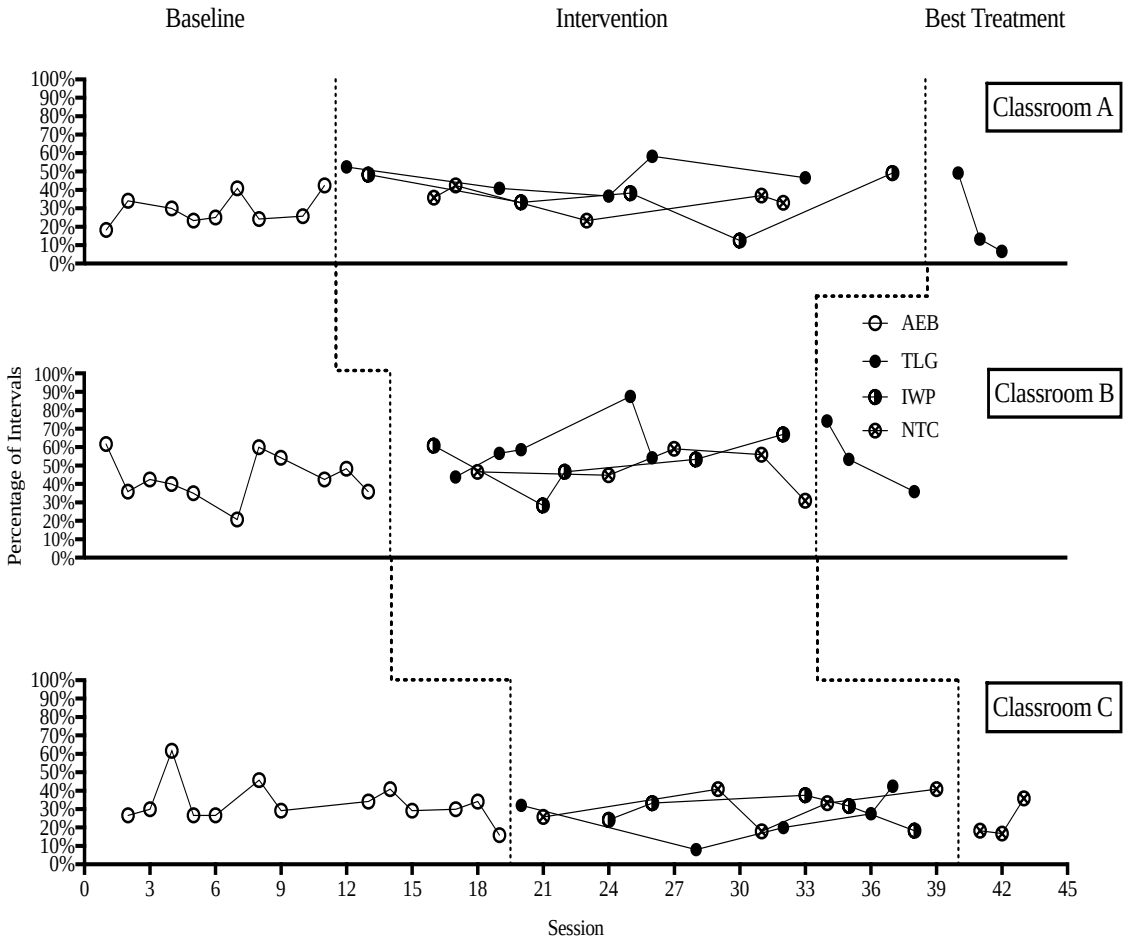

Figure 1. Percentage of intervals of academically engaged behavior (AEB) across all phases for three classrooms

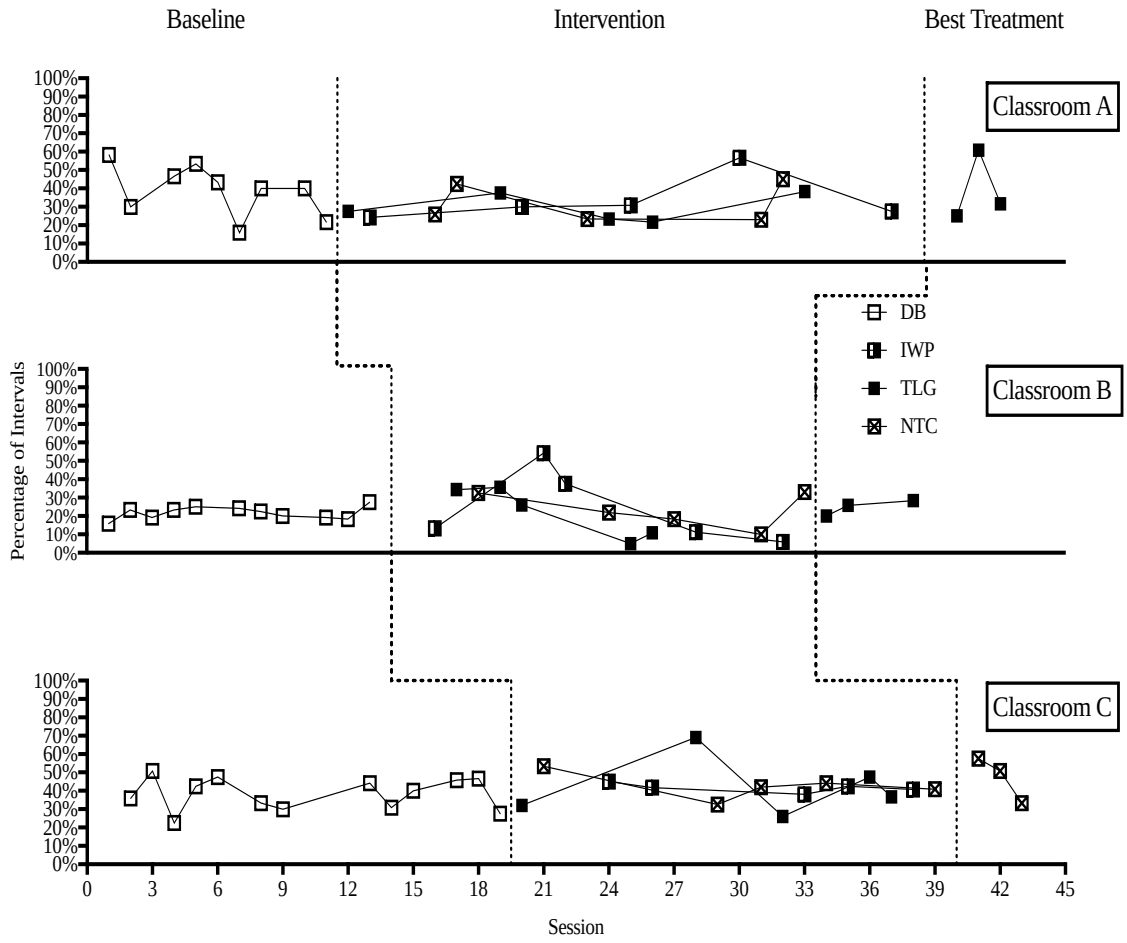

Figure 2. Percentage of intervals of disruptive behavior (DB) across all phases for three classrooms

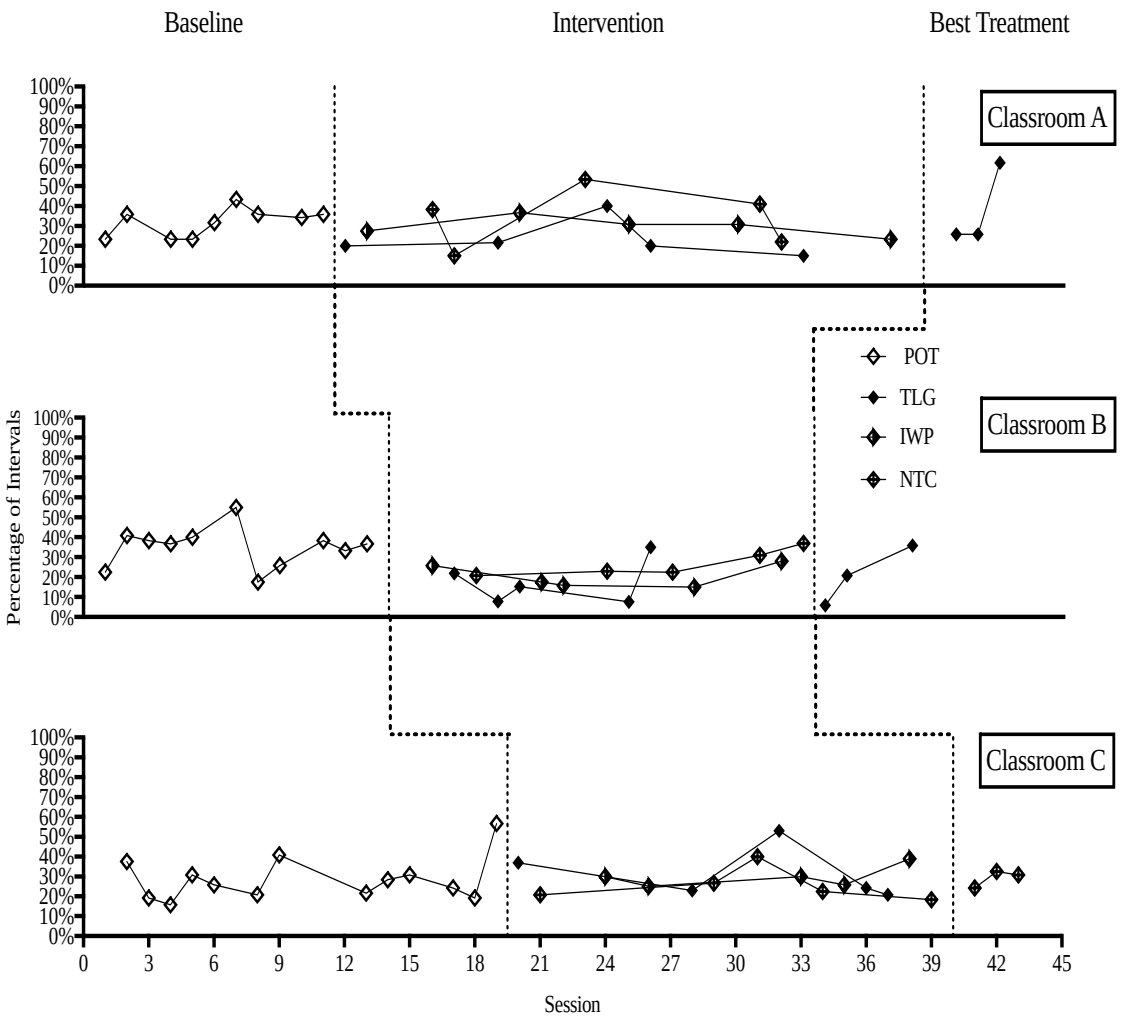

Figure 3. Percentage of intervals of passive off-task behavior (POT) across all phases for three classrooms

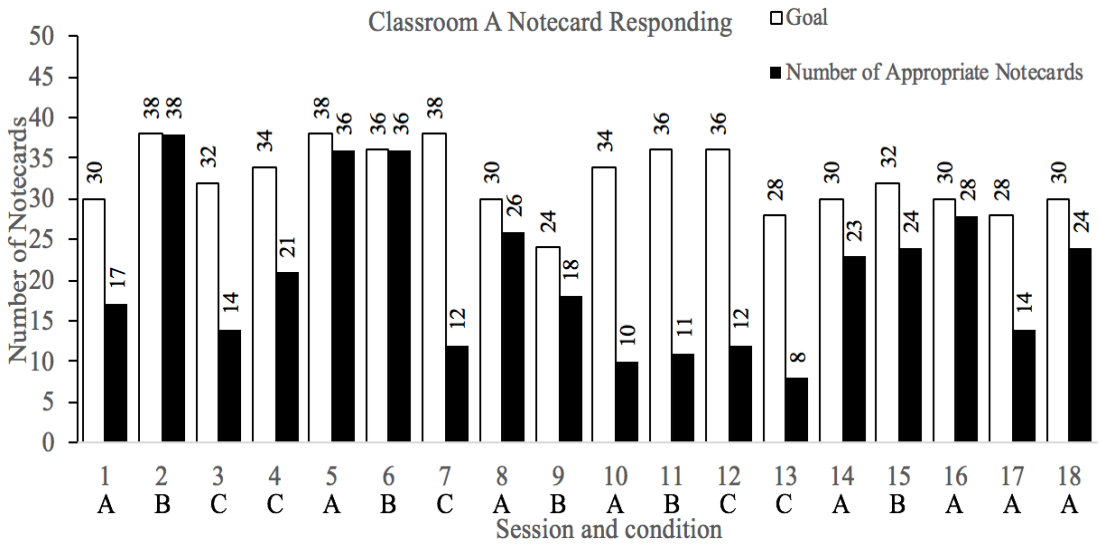

Figure 4. Number of notecards turned in across sessions and conditions for classroom A
Note. A= tootling, B= comparison writing procedure, C= no-treatment control

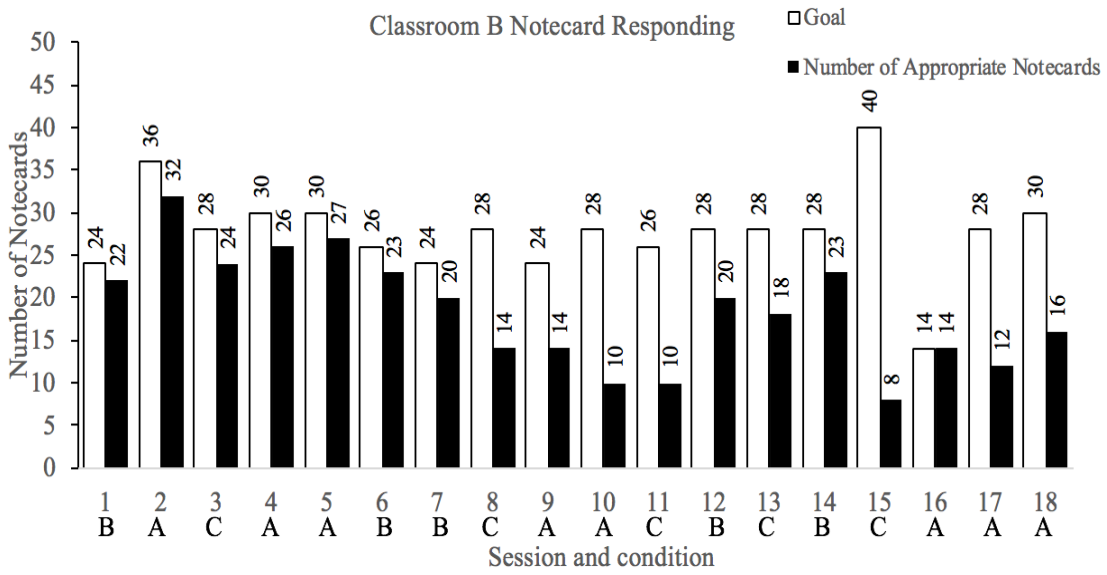

Figure 5. Number of notecards turned in across sessions and conditions for classroom B
Note. A= tootling, B= comparison writing procedure, C= no-treatment control

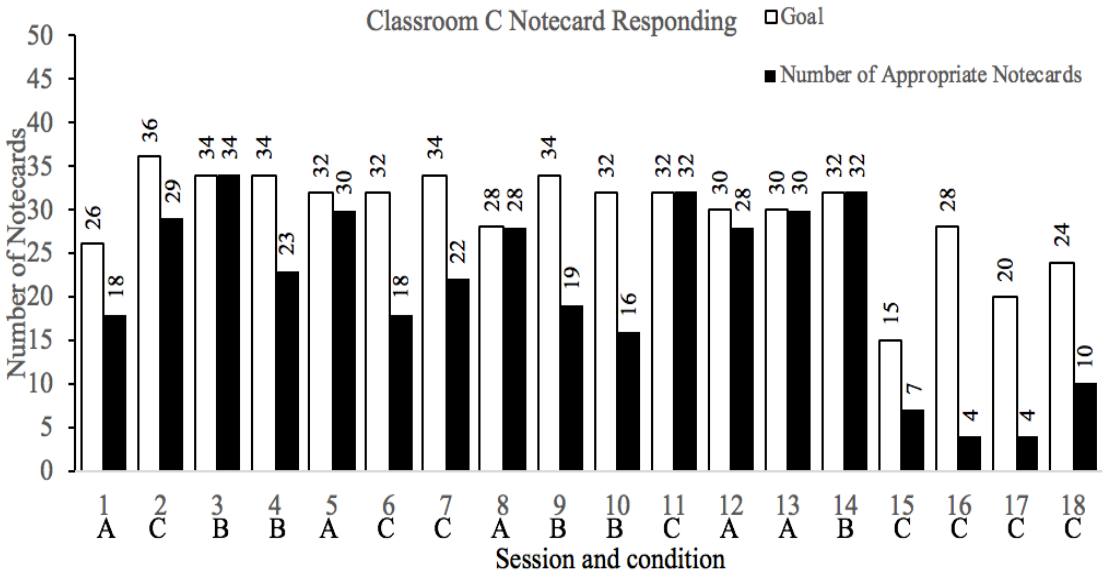

Figure 6. Number of notecards turned in across sessions and conditions for classroom C
Note. A= tootling, B= comparison writing procedure, C= no-treatment control

Effect Sizes

Baseline-Corrected Tau effect sizes were calculated across all three classrooms to determine if intervention conditions had an effect on student's levels of AEB, DB, and POT. None of the data across the three classrooms had to be corrected for baseline trend. Effect sizes for Classroom A indicate a large effect for AEB (TAU= .60) and a moderate effect for DB (TAU= .35) and POT (TAU= .47) during the tootling intervention when compared to baseline levels of behavior. For the comparison writing procedure, effect sizes indicate a moderate effect for AEB (TAU= .27) and a small effect for DB (TAU= .16) and POT (TAU= .13) when compared to baseline levels of behavior. Effect sizes for the no-treatment control condition indicate a moderate effect for AEB (TAU= .24) and a small effect for both DB (TAU= .17) and POT (TAU= .08) when compared to baseline levels of behavior. Effect sizes were also calculated to compare the effects of the condition implemented during the best treatment phase on AEB, DB, and POT to both

baseline and intervention levels of behavior. Effect sizes for tootling during the best treatment phase when compared to baseline levels of behavior indicate a moderate effect for AEB (TAU= .21) and a small effect DB (TAU= .02) and POT (TAU= .08). Effect sizes for tootling during the best treatment phase when compared to intervention levels of behavior were also moderate for AEB (TAU= .44), DB (TAU= .24), and POT (TAU= .55). Refer to Table 2 to for effect size calculations for classroom A.

Effect sizes for classroom B indicate a moderate effect for both AEB (TAU= .43) and POT (TAU= .56) during the tootling intervention and a small effect for DB (TAU= .11) when compared to baseline levels of behavior. Similar to the tootling intervention, effect sizes for the comparison writing procedure indicate a moderate effect for both AEB (TAU= .24) and POT (TAU= .54) and a small effect for DB (TAU= .14) when compared to baseline levels of behavior. Effect sizes for the no-treatment control condition indicate a small effect for both AEB (TAU= .14) and DB (TAU= .03) and a moderate effect for POT (TAU= .35) when compared to baseline levels of behavior. Effect sizes were also calculated to compare the effects of the condition implemented during the best treatment phase on AEB, DB, and POT to both baseline and intervention levels of behavior. Effect sizes for tootling during the best treatment phase when compared to baseline levels of behavior were moderate across AEB (TAU= .21), DB (TAU= .33), and POT (TAU= .42). Effect sizes for tootling during the best treatment phase when compared to intervention levels of behavior were moderate for AEB (TAU= .24) and small for both DB (TAU= .05) and POT (TAU= .05). Refer to Table 3 to for effect size calculations for classroom B.

Effect sizes for classroom C indicate a small effect across AEB (TAU= .17), DB (TAU= .06), and POT (TAU= .07) during the tootling intervention when compared to baseline levels of behavior. Similar to the tootling intervention, effect sizes for the comparison writing procedure indicate a small effect across AEB (TAU= .09), DB (TAU= .08) and POT (TAU= .14) when compared to baseline levels of behavior. Additionally, effect sizes for the no-treatment control condition indicate a small effect across AEB (TAU= .03), DB (TAU= .14), and POT (TAU= .10) when compared to baseline levels of behavior. Effect sizes were also calculated to compare the effects of the condition implemented during the best treatment phase on AEB, DB, and POT to both baseline and intervention levels of behavior. Effect sizes for the no-treatment control condition during the best treatment phase when compared to baseline levels of behavior were moderate for AEB (TAU= .23) and DB (TAU= .31) and small for POT (TAU= .16). Effect sizes for the no-treatment control condition during the best treatment phase when compared to intervention levels of behavior were moderate across AEB (TAU= .35), DB (TAU= .24), and POT (TAU= .34). Refer to Table 4 to for effect size calculations for classroom C.

Table 2

Baseline-Corrected Tau Effect Size Calculations for Classroom A

Dependent Variable	Baseline-Corrected Tau	Effect Size
AEB		
Baseline/ Tootling	0.60	Large
Baseline/ Comparison Writing Procedure	0.27	Moderate
Baseline/ No-treatment Control	0.24	Moderate
Baseline/ Best Treatment (Tootling)	0.21	Moderate
Tootling/ Best Treatment (Tootling)	0.44	Moderate

Table 2 (continued).

DB		
Baseline/ Tootling	0.35	Moderate
Baseline/ Comparison Writing Procedure	0.16	Small
Baseline/ No-treatment Control	0.17	Small
Baseline/ Best Treatment (Tootling)	0.02	Small
Tootling/ Best Treatment (Tootling)	0.24	Moderate
POT		
Baseline/ Tootling	0.47	Moderate
Baseline/ Comparison Writing Procedure	0.13	Small
Baseline/ No-treatment Control	0.08	Small
Baseline/ Best Treatment (Tootling)	0.08	Small
Tootling/ Best Treatment (Tootling)	0.55	Moderate

Note. Scores < .2 = small effect sizes; scores between .2 and .6 = moderate effect sizes; scores between .6 and .8 = large effect sizes; scores ≥ .8 indicate large to very large effect sizes

Table 3

Baseline-Corrected Tau Effect Size Calculations for Classroom B

Dependent Variable	Baseline-Corrected Tau	Effect Size
AEB		
Baseline/ Tootling	0.43	Moderate
Baseline/ Comparison Writing Procedure	0.24	Moderate
Baseline/ No-treatment Control	0.14	Small
Baseline/ Best Treatment (Tootling)	0.21	Moderate
Tootling/ Best Treatment (Tootling)	0.24	Moderate
DB		
Baseline/ Tootling	0.11	Small
Baseline/ Comparison Writing Procedure	0.14	Small
Baseline/ No-treatment Control	0.03	Small
Baseline/ Best Treatment (Tootling)	0.33	Moderate
Tootling/ Best Treatment (Tootling)	0.05	Small
POT		

Table 3 (continued).

Baseline/ Tootling	0.56	Moderate
Baseline/ Comparison Writing Procedure	0.54	Moderate
Baseline/ No-treatment Control	0.35	Moderate
Baseline/ Best Treatment (Tootling)	0.42	Moderate
Tootling/ Best Treatment (Tootling)	0.05	Small

Note. Scores < .2 = small effect sizes; scores between .2 and .6 = moderate effect sizes; scores between .6 and .8 = large effect sizes; scores ≥ .8 indicate large to very large effect sizes

Table 4

Baseline-Corrected Tau Effect Size Calculations for Classroom C

Dependent Variable	Baseline-Corrected Tau	Effect Size
AEB		
Baseline/ Tootling	0.17	Small
Baseline/ Comparison Writing Procedure	0.09	Small
Baseline/ No-treatment Control	0.03	Small
Baseline/ Best Treatment (No-treatment Control)	0.23	Moderate
No-treatment Control/ Best Treatment (No-treatment Control)	0.35	Moderate
DB		
Baseline/ Tootling	0.06	Small
Baseline/ Comparison Writing Procedure	0.08	Small
Baseline/ No-treatment Control	0.14	Small
Baseline/ Best Treatment (No-treatment Control)	0.31	Moderate
No-treatment Control/ Best Treatment (No-treatment Control)	0.24	Moderate
POT		
Baseline/ Tootling	0.07	Small
Baseline/ Comparison Writing Procedure	0.14	Small
Baseline/ No-treatment Control	0.10	Small

Table 4 (continued).

Baseline/ Best Treatment (No-treatment Control)	0.16	Small
No-treatment Control/ Best Treatment (No-treatment Control)	0.34	Moderate

Note. Scores < .2 = small effect sizes; scores between .2 and .6 = moderate effect sizes; scores between .6 and .8 = large effect sizes; scores ≥ .8 indicate large to very large effect sizes

Social Validity

URP-IR

Upon completion of the study, all three teachers completed the Usage Rating Profile- Intervention Revised (*URP- IR*). Teachers scores on the *URP- IR* can be found in Table 5. Questions on the rating scale were based on a 6- point Likert scale ranging from 1 (*strongly disagree*) to 6 (*strongly agree*). Overall, all three teachers reported consistently high scores on the factors of acceptability, understanding, feasibility, and system climate. This means the teachers found the intervention components acceptable and feasible to implement in their classes, the intervention components were easy for them to understand, and they felt that the school administration was supportive of their participation in this study. The other two factors, home school collaboration and system support had variable scores across teachers. The teacher in classroom A scored items on these factors higher indicating that collaboration between student's home and school is important in the implementation of this intervention and that administrative support in terms of having appropriate resources to carry out the intervention would be necessary. Whereas, the teachers in classrooms B and C's scores on these two factors indicate that home school collaboration and system support are of small to moderate importance to them during the implementation of the intervention components in this study.

Table 5

Mean Teacher Rating on the Usage Rating Profile (URP-IR)

	Classroom		
Factor	A	B	C
Acceptability	4.56	4.44	5.22
Understanding	5	4.33	4.67
Home School Collaboration	4.33	2.33	3.33
Feasibility	5	5	4.83
System Climate	4	4.25	4.25
System Support	5	3.33	3.33

CURP

The Children's Usage Rating Profile (*CURP*) was completed by 20% of students across all three classrooms upon conclusion of the study. Students were given a consent form for their parents to sign in order to participate in the completion of the rating scale. The primary investigator provided the consent forms at the beginning of the best treatment phase in each classroom. Students had one week to return the consent forms and teachers were encouraged to provide prompts each day to get the consent forms turned in. However, a limited number of students returned their consent forms and thus were able to complete the rating scale. Therefore, the results of the *CURP* should be interpreted with caution. Results may be obtained from the author if so desired.

CHAPTER IV- DISCUSSION

Several studies have examined the effects of tootling on class-wide levels of academically engaged behavior and disruptive behaviors, however, there are not any studies in the literature currently that have looked into the individual components of tootling and how those play into the effectiveness of the intervention. The purpose of the current study was to evaluate whether the written component of tootling is effective at increasing class-wide levels of AEB and decrease class-wide levels of DB in a high school classroom setting. This was accomplished by comparing the traditional tootling intervention to a comparison writing procedure and a no-treatment control condition. It was hypothesized that implementation of the tootling intervention and the comparison writing procedure would lead to increases in academically engaged behavior and decreases in disruptive behavior. Limitation from this study and potential future directions are discussed.

Research Questions

Question 1

The first question of this study addressed whether the traditional tootling intervention had a meaningful increase in academically engaged behavior when compared to the comparison writing procedure. Using visual analysis of AEB in figure 1 and the effect size calculations (Tables 2-4), classrooms A and B had “meaningful” increases in AEB during the traditional tootling intervention when compared to the comparison writing procedure. Classroom C, on the other hand, had increased levels of AEB during the comparison writing procedure when compared to the tootling intervention. The results for classroom A showed mean levels of AEB of 60.16%

compared to the comparison writing procedures mean levels of AEB of 51.23%. Similarly, classroom B had mean levels of AEB of 47% during the tootling intervention when compared to the comparison writing procedure's mean levels of AEB of 36.33%. Classroom C conversely had mean levels of AEB of 29% during the comparison writing procedure compared to the tootling intervention's mean levels of AEB of 26%.

Looking at the effect size calculations for AEB, classroom A had a large effect size for the tootling intervention (TAU= .60) and a small effect for the comparison writing procedure (TAU= .27) during the intervention phase when compared to levels of behavior in the baseline phase. Effect size calculations for AEB during the tootling intervention for the best treatment phase were moderate (TAU= .21) when compared to levels of behavior in the baseline phase and moderate (TAU= .44) when compared to levels of behavior in the intervention phase. For classroom B, effect size calculations for AEB were moderate for the tootling intervention (TAU= .43) and the comparison writing procedure (TAU= .24) when compared to levels of behavior in the baseline phase. During the best treatment phase, effect size calculations for AEB were moderate (TAU= .21) when compared to levels of behavior in the baseline phase and moderate (TAU= .24) when compared to levels of behavior in the intervention phase. Effect size calculations for classroom C indicated small effect sizes for both the tootling intervention (TAU= .17) and the comparison writing procedure (TAU= .09) for AEB during the intervention phase when compared to baseline levels of behavior.

Question 2

The second question of this study addressed whether the implementation of the traditional tootling intervention had a meaningful decrease in DB when compared to the

comparison writing procedure. Using visual analysis of DB in figure 2 and the effect size calculations (Tables 2-4), none of the three classrooms had a “meaningful” decrease in DB during the traditional tootling intervention when compared to the comparison writing procedure. Classrooms A had a slightly lower mean levels of DB of 29.67% during the tootling intervention when compared to the comparison writing procedure’s mean levels of DB of 33.84%. Classroom B had a slightly lower mean levels of DB of 22.37% during the tootling intervention when compared to the comparison writing procedure’s mean levels of DB of 24.45%. Lastly, Classroom C had a slightly higher mean levels of DB of 42.23% during the tootling intervention when compared to the comparison writing procedure’s mean levels of DB of 41.58%.

When considering the effect size calculations for DB, classrooms A had a moderate effect size for the tootling intervention (TAU= .35) and a small effect for the comparison writing procedure (TAU= .16) during the intervention phase when compared to levels of behavior in the baseline phase. Effect size calculations for DB during the tootling intervention for the best treatment phase were small (TAU= .02) when compared to levels of behavior in the baseline phase and moderate (TAU= .24) when compared to levels of behavior in the intervention phase. For classroom B, effect size calculations for DB were small for the tootling intervention (TAU= .11) and the comparison writing procedure (TAU= .14) when compared to levels of behavior in the baseline phase. During the best treatment phase, effect size calculations for DB were moderate (TAU= .33) when compared to levels of behavior in the baseline phase and small (TAU= .05) when compared to levels of behavior in the intervention phase. Effect size calculations for classroom C indicated small effect sizes for both the tootling intervention (TAU= .06)

and the comparison writing procedure (TAU= .08) for DB during the intervention phase when compared to baseline levels of behavior.

Question 3

The third question of this study addressed whether implementation of a no-treatment control condition had a meaningful increase in AEB and decrease in DB. In regard to AEB for classroom A, visual analysis indicated a slight increase in AEB from baseline to intervention with mean levels of AEB increasing from 29.35% to 34.33%. For DB in classroom A, there was a moderate decrease in mean levels from baseline to intervention from 58.33% to 31.93%. Effect size calculations were consistent with these findings with a moderate effect for AEB (TAU= .24) during intervention when compared to baseline levels of behavior and a small effect for DB (TAU= .17) during intervention when compared to baseline levels of behavior. For classroom B, the opposite effect of what was hypothesized occurred. Visual analysis for classroom B indicated a slight decrease in AEB from baseline to intervention with mean levels of AEB decreasing from 61.67% to 47.53%. Mean levels of DB for classroom B slightly increased from 21.67% during baseline to 23.14% during intervention. Effect size calculations for classroom B were consistent with these findings in that there were small effects for both AEB (TAU= .14) and DB (TAU= .03) during intervention when compared to baseline levels of behavior.

For classroom C, visual analysis of AEB indicated a slight increase from baseline to intervention with mean levels of AEB increasing from 26.67% to 31.76%. Levels of DB also displayed an increase in mean levels from 38.27% during baseline to 42.57% during intervention. Effect size calculations for classroom C indicated a small effect for

AEB (TAU= .03) and DB (TAU= .14) during intervention when compared to baseline levels of behavior. Additionally, the no-treatment control condition was selected for the best treatment phase for classroom C. This is not the expected result from a no-treatment control condition given that there was not a reinforcement component in place as there was with the other two intervention conditions. During the best treatment phase, mean levels of AEB decreased form intervention levels to 23.6% and DB increased to 47.22%. Effect size calculations were consistent with these findings indicating a moderate effect for AEB (TAU= .23) during the best treatment phase when compared to baseline levels of behavior and a moderate effect (TAU= .35) when compared to intervention levels of behavior. Effect size calculations for DB were moderate (TAU= .31; TAU= .24) during the best treatment phase when compared to both the baseline and intervention phases levels of behavior respectively. Overall, results for the no-treatment control condition indicated variable findings with none of the three classes having meaningful increases in AEB or decreases in DB.

Question 4

The fourth question of this study addressed whether a distinguishable difference in notecard responding existed between implementation of traditional tootling, a comparison writing procedure, and a no-treatment control condition. When analyzing the data for the tertiary dependent variable, notecard responding, all three classrooms demonstrated variable responding for the traditional tootling intervention, the comparison writing procedure, and the no-treatment control condition. Notecard responding for classroom A was variable and while the class never met their goal during the tootling condition, it was the condition selected for the best treatment phase due to it having the

highest mean levels of AEB. Classroom A had a slightly higher mean for notecard responding during the comparison writing procedure compared to the tootling intervention and met their goal twice during the comparison writing procedure. In comparison to the tootling and comparison writing procedure, the no-treatment control condition had a significantly lower mean for notecard responding in classroom A which was expected due to no reinforcement being in place during this condition. Similar to classroom A, classroom B's notecard responding during the tootling intervention was variable with the class never meeting their goal, but it was the condition selected for the best treatment phase. Classrooms B had a slightly higher mean for notecard responding during the tootling intervention compared to the comparison writing procedure and did meet their goal once during the best treatment phase. In regard to the no-treatment control condition, classroom B also had a significantly lower mean for notecard responding in comparison to the tootling and comparison writing procedure.

For classroom C, notecard responding during the tootling intervention had the highest mean levels of appropriate responding compared to the other two classes and the class had a slightly higher mean for notecard responding during the tootling intervention compared to the comparison writing procedure. Additionally, the class met their goal twice during the tootling condition and twice during the comparison writing procedure condition. However, neither the tootling intervention nor the comparison writing procedure were selected for the best treatment phase. The no-treatment control condition for classroom C was selected for the best treatment phase due to the condition resulting in the highest mean levels of AEB, although the no-treatment control condition had the lowest mean for notecard responding of the three conditions. Students in classroom C

turned in the appropriate number of notecards to meet their goal one time during the no-treatment control condition, however, there was no reinforcement contingency in place during this condition. Furthermore, during the best treatment phase, notecard responding was significantly low and did not come close the goal a single time. Overall, a distinguishable difference in notecard responding was shown to exist in classrooms A and B with responding during the tootling and comparison writing procedures to be higher than the no-treatment control condition. However, results for notecard responding in classroom c indicated variable, high levels of responding across all three conditions.

Question 5

The fifth question of this study addressed whether the traditional tootling intervention and comparison writing procedure were identified by teachers and students as socially valid class-wide interventions. According to the results of the *URP- IR* (Table 5), all three classroom teachers rated the intervention high across four of the six factors, in particular, they moderately agreed that the intervention components were acceptable for their classrooms with a mean for classroom A of 4.56, a mean for classroom B of 4.44, and a mean of classroom C of 5.22. In regard to the *CURP*, students across all three classes rated the intervention components more variably. Due to the limited sample of students who completed the measure it is difficult to interpret the outcomes reliably.

Limitations

There are several limitations that need to be taken into consideration when evaluating this study. The first limitation was this study had a small sample size with only three general education classrooms used from one rural high school. A more robust sample is needed to be able to generalize any results to students across other school

settings. A second limitation was that caution is warranted when interpreting the findings for student social validity (*CURP*) due to the limited number of students who were able to fill out the measure across all three classrooms. As previously mentioned, classroom A only had 2 students fill out the measure, classroom B had 1 student fill out the measure, and classroom C had 9 students fill out the measure. A more robust sample is needed to determine the accuracy of the social validity scores. A third limitation was the sampling without replacement method used to determine the order of conditions during intervention was not implemented correctly for Classroom B during session 5. Treatment A was replaced and reselected for implementation before all three treatments had been run. However, this error was corrected during session 15 by implementing the no-treatment control condition as it was the only condition that had yet to be implemented five times throughout the intervention phase. Sampling without replacement was implemented correctly for the other two classrooms.

A fourth limitation was observations were only 20 minutes in length providing a limited sample of behavior occurring in each classroom. It is unknown what behaviors looked like during the other 70 minutes of class time. Therefore, results should not be generalized to the entire course duration. A fifth limitation was three of the 20- minute observations were cut 3 minutes short due to time constraints in class schedules. Classroom B's teacher in particular often waited a while to begin implementation of the intervention after researchers entered the classroom. When this occurred for a second time, the primary investigator met with the teacher to determine possible changes or fixes that could be made to ensure that observations were conducted in a timely manner. One observation was also cut short in in classroom C due to an unexpected fire drill occurring

during the observation. A sixth limitation was that there were potential carryover effects between treatment A (tootling), treatment B (comparison writing procedure), and the no-treatment controls on notecard responding. Although students were provided explicit instructions by the teacher at the beginning of each observation using a script, there was often still inaccurate notecard responding. Researchers found that students were turning in notecards during treatment A ("write down two things you see your peers doing that are good or helpful") with responses to treatment B and C's prompt ("write down two things you learned today") and vice versa. Students would on occasion ask the teacher to repeat instructions during observations and reported becoming confused with what to write down. Teachers would simultaneously become frustrated with students for not being engaged when the script was read. This led to classes not meeting their daily goals. Upon analyzing the results of this study, it is though that the classroom goal, the number of students in attendance multiplied by the number of notecards each student could turn in (2), may have been too stringent and that the routine changing of conditions may have led to unclear discrimination between conditions.

Another limitation was that the majority of observations were conducted at the end of class time for all three classrooms, however, two teachers twice requested for observations to be moved to the beginning of class. Reasons included changes in the school schedule due to inclement weather, school assemblies, teachers having to leave school early, and state testing. The change in observation time may have resulted in changes in levels of AEB, DB, POT, and notecard responding due to differences in student motivation. For example, students may have been more motivated to be engaged in class instruction and intervention implementation at the beginning of class time versus

the end of class time potentially increasing levels of AEB and concurrently decreasing levels of DB and POT. The last limitation was that candy was the only reinforcer utilized throughout the study. Other forms of reinforcement were also suggested, however, teachers selected candy due to feasibility. This is a limitation because strength of reinforcement may have had an impact on student's motivation to participate in the study during both the intervention and best treatment phases.

Future Directions

While the traditional tootling intervention has been proven effective in the literature (e.g., Cihak et al., 2009), a consistent limitation in prior literature (e.g., Cihak et al., 2009; Lambert et al., 2015; Lum et al., 2017; McHugh et al., 2016) is that results of the tootling intervention cannot be attributed to any one component of the intervention. This study was the first to examine an individual component of the intervention to determine its effectiveness on class-wide levels of behavior. Future studies should continue to examine the individual components of the intervention to determine what is truly driving behavior change. A replication of this study to address some of the limitations would be beneficial to determine whether a more clear and distinguishable effect is established. Additionally, a full component analysis would be beneficial to examine the effects of more than one component of the tootling intervention.

Conclusion

In conclusion, the findings of this study warrant the question, what impact does the written component of the tootling intervention have on levels of behavior, specifically AEB, DB, POT, and notecard responding? The results of the study demonstrated no meaningful differences between the three intervention conditions. Thus, when compared

to a comparison writing procedure and no-treatment control condition, the tootling intervention did not have clinically significant effects on levels of AEB and DB consistent with previous studies implemented in the high school setting (Wright, 2016; Lum et al., 2019). These findings are consistent using both visual analysis and effect size calculations. Therefore, results demonstrated that the written component of tootling is not the component of tootling that promotes behavior change. More research on the written component of the tootling intervention is warranted to confirm these findings.

APPENDIX A- Institutional Review Board Approval Letter

ORI | Office of Research Integrity
118 COLLEGE DRIVE #5116 • HATTIESBURG, MS | 601.266.6756 | WWW.USM.EDU/ORI

INSTITUTIONAL REVIEW BOARD
118 College Drive #5147 | Hattiesburg, MS 39406-0001
Phone: 601.266.5997 | Fax: 601.266.4377 | www.usm.edu/research/institutional-review-board

NOTICE OF COMMITTEE ACTION

The project below has been reviewed by The University of Southern Mississippi Institutional Review Board in accordance with Federal Drug Administration regulations (21 CFR 26, 111), Department of Health and Human Services regulations (45 CFR Part 46), and University Policy to ensure:

- The risks to subjects are minimized and reasonable in relation to the anticipated benefits.
- The selection of subjects is equitable.
- Informed consent is adequate and appropriately documented.
- Where appropriate, the research plan makes adequate provisions for monitoring the data collected to ensure the safety of the subjects.
- Where appropriate, there are adequate provisions to protect the privacy of subjects and to maintain the confidentiality of all data.
- Appropriate additional safeguards have been included to protect vulnerable subjects.
- Any unanticipated, serious, or continuing problems encountered involving risks to subjects must be reported immediately, but not later than 10 days following the event. Problems should be reported to ORI via the Incident template on Cayuse IRB.
- The period of approval is twelve months. An application for renewal must be submitted for projects exceeding twelve months.

PROTOCOL NUMBER: IRB-18-66

PROJECT TITLE: Why Does Tootling Work? A Comparison of the Group Contingency to an Inert Writing Procedure

SCHOOL/PROGRAM: School of Psychology, Psychology

RESEARCHER(S): James Derieux
Evan Dart

IRB COMMITTEE ACTION: Approved
CATEGORY: Exempt
PERIOD OF APPROVAL: November 6, 2018 - November 6, 2019

Edward L. Goshorn, Ph.D.
Institutional Review Board Chairperson

REFERENCES

Barrish, H. H., Saunders, M., & Wolf, M. W. (1969). Good Behavior Game: Effects of individual contingencies for group consequences on disruptive behavior in a classroom. *Journal of Applied Behavior Analysis*, 2, 119-124.

Briesch, Amy M., & Chafouleas, S. M. (2009). Exploring student buy-in: Initial development of an instrument to measure likelihood of children's intervention usage. *Journal of Educational and Psychological Consultation, 19,* 321–336.

Briesch, Amy M., Chafouleas, S. M., Neugebauer, S. R., & Riley-Tillman, T. C. (2013). Assessing influences on intervention use: Revision of the Usage Rating Profile Intervention. *Journal of School Psychology,* 51, 81–96.

Briesch, Amy M., Hemphill, E. M., Volpe, R. J., & Daniels, B. (2015). An evaluation of observational methods for measuring response to class-wide intervention. *School Psychology Quarterly*, *30*, 37.

Cashwell, T.H., Skinner, C.H., & Smith, E.S. (2001). Increasing second-grade students' reports of peers' prosocial behaviors via direct instruction, group reinforcement, and progress feedback: A replication and extension. *Education and Treatment of Children*, 24, 161-175.

Cihak, D. F., Kirk, E. R., & Boon, R. T. (2009). Effects of class wide positive peer "tootling" to reduce the disruptive classroom behaviors of elementary students with and without disabilities. *Journal of Behavioral Education*, *18*, 267-278.

Cooper, J. O., Heron, T.E., & Heward, W.L. (2007). *Applied Behavior Analysis* (2nd Ed). Upper Saddle River, NY: Pearson Prentice.

Dart, Evan H., Fischer, Aaron., Polakoff, Ben., Richardson, Dylan., Wright, Sarah. (In press). *A Comparison of In-Vivo and Digital Systematic Direct Observation*. The University of Southern Mississippi, Hattiesburg, MS.

Dart, E. H., Radley, K. C., Briesch, A. M., Furlow, Christopher M., & Cavell, Hannah J. (2016). Assessing the Accuracy of Classwide Direct Observation Methods: Two Analyses Using Simulated and Naturalistic Data. *Behavioral Disorders,41*(3), 148-160.

Elliott, S., & Von Brock Treuting, M. (1991). The behavior intervention rating scale: Development and validation of a pretreatment acceptability and effectiveness measure. *Journal of School Psychology, 29*, 43-51.

Emmer, E. T., & Sabornie, E. J. (2015). Introduction to the section edition. In E. T. Emmer & E. J. Sabornie (Eds.), *Handbook on classroom management* (2nd ed.). New York, NY: Routledge.

Fairbanks, Sarah, Sugai, George, Guardino, David, Lathrop, Margaret (2007). Response to Intervention: Examining Classroom Behavior Support in Second Grade. *Exceptional Children*, Vol 73, Issue 3, pp. 288- 310.

Gamm, S., Casserly, M., Uro, G., Walston, D., Hall, R., et al. (2012). Common Core State Standards and Diverse Urban Students: Using Multi-Tiered Systems of Support. Washington D.C.: Council of the Great City Schools.

Green, S. B., McCoy, J. F., Burns, K. P., & Smith, A. C. (1982). Accuracy of observational data with whole interval, partial interval, and momentary time-sampling recording techniques. *Journal of Behavioral Assessment*, *4*, 103-118.

Gresham, F. M., & Gresham, G. N. (1982). Interdependent, Dependent, and Independent Group Contingencies for Controlling Disruptive Behavior. *The Journal of Special Education, 16*(1), 101-110.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, 71, 165-179.

Horner, R.H., & Sugai, G. (2000). School-wide behavior support: An emerging initiative. *Journal of Positive Behavior Interventions*, 2, 231-232.

Hoglund, W. L., Klingle, K. E., & Hosan, N. E. (2015). Classroom risks and resources: Teacher burnout, classroom quality and children's adjustment in high needs elementary schools. *Journal of School Psychology, 53*, 337–357.

Kazdin, A.E. (1993). Treatment of Conduct Disorders: Progress and directions in psychotherapy research. *Development and Psychopathology*,5, 277-310.

Kratochwill, T. R., & Bergan, J. R. (1990). *Behavioral consultation in applied settings: An individual guide*. New York, NY: Plenum Publishing Corporation.

Kratochwill, T. R., Hitchock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D.M. & Shadish, W.R. (2010). *Single-case designs technical documentation*. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf.

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial and Special Education*, *34*(1), 26–38.

Lambert, A. M. (2012). *Evaluating the effects of tootling of disruptive and appropriate behaviors in elementary school children*. (Master's thesis). The University of Southern Mississippi, Hattiesburg, MS.

Lambert, A. M, Tingstrom, D.H., Sterling, H.E., Dufrene B.A., & Lynne, S. (2015). Effects of tootling on class wide disruptive and appropriate behavior of upper-elementary students. *Behavior Modification, 39,* 413-430.

Lane, K. (2007). Identifying and supporting students at risk for emotional and behavioral disorders within multi-level models: Data driven approaches to conducting secondary interventions with an academic emphasis. *Education and Treatment of Children, 30,* 135–164.

Lewis, T.J., & Sugai, G. (1999). Effective behavior support: A systems approach to proactive school-wide management. *Focus on Exceptional Children,* 31, 1-24.

Litow, L., & Pumroy, D. K. (1975). A brief review of classroom group-oriented contingencies. *Journal of Applied Behavior Analysis, 8,* 341–347.

Lum, J. D. K., Tingstrom, D. H., Dufrene, B. A., Radley, K. C. and Lynne, S. (2017). Effects of Tootling on Classwide Disruptive and Academically Engaged Behavior of General-education High School Students. *Psychology in the Schools, 54,* 370-384.

Lum, J. D. K., Radley, K. C., Tingstrom, D. H., Dufrene, B. A., Olmi, D. J., & Wright, S. J. (2019). Tootling with a Randomized Independent Group Contingency to Improve High School Classwide Behavior. *Journal of Positive Behavior Interventions,21,* 93-105.

Maggin, D. M., Pustejovsky, J. E., & Johnson, A. H. (2017). A Meta-Analysis of School-Based Group Contingency Interventions for Students with Challenging Behavior: An Update. *Remedial and Special Education,* 0741932517716900.

McHugh, M. B., Tingstrom, D. H., Radley K. C., Barry, C. T., & Walker, K. M. (2016). Effects of Tootling on Classwide and Individual Disruptive and Academically Engaged Behavior of Lower-elementary Students. *Behavioral Interventions, 31*, 332-354.

McHugh Dillon, M. B., Radley, K. C., Tingstrom, D. H., Dart, E. H., & Barry, C. T. (2019). The Effects of Tootling via ClassDojo on Student Behavior in Elementary Classrooms. *School Psychology Review, 48,* 18-30.

Parker, R. I., & Vannest, K. J. (2009). An improved effect size for single case research: Non-overlap of all pairs (NAP). Behavior Therapy, 40, 357–367.

Parker, R. I., Vannest, K. J., & Davis, J. L. (2011). Effect size in single-case research: A review of nine nonoverlap techniques. Behavior Modification, 35, 303–322.

Radley, K. C., O'Handley, R. D., & LaBrot, Z. C. (2015). A comparison of momentary time sampling and partial-interval recording for assessment of effects of social skills training. *Psychology in the Schools, 52,* 363-378.

Sherman, Julie Christine, "Positive Peer Reporting and Positive Peer Reporting Combined with Tootling: A Comparison of Interventions" (2012). Dissertations. 857.

Skinner, C.H., Cashwell, T.H., & Skinner, A.L. (2000). Increasing Tootling: The Effects of a Peer- Monitored Group Contingency Program on Students' Reports of Peers' Prosocial Behaviors. *Psychology in the Schools*, 37, 263-270.

Sugai, G., & Horner, R. H. (2000). Introduction to the special series on positive behavior support in schools. *Journal of Emotional and Behavioral Disorders, 10*, 130-135.

Tarlow, K. R. (2017). An improved rank correlation effect size statistic for single-case designs: Baseline corrected tau. *Behavior Modification*, *41*(4), 427–467.

Theodore, L. A., Bray, M. A., Kehle, T. J., & Dioguardi, R. J. (2004). Contemporary review of group-oriented contingencies for disruptive behavior. *Journal of Applied School Psychology,* 20(1), 79–101.

Tingstrom, D. H., Sterling-Turner, H. E., & Wilczynski, S. M. (2006). The Good Behavior Game: 1969–2002. *Behavior Modification, 30*, 225–253.

Vannest, K. J., & Ninci, J. (2015). Evaluating Intervention Effects in Single-Case Research Designs. *Journal of Counseling & Development, 93*(4), 403-411.

Von Brock, M. B., & Elliott, S. N. (1987). Influence of treatment effectiveness information on the acceptability of classroom interventions. *Journal of School Psychology*, *25*, 131-144.

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

Wright, Sarah J. (2016). *The Effects of Tootling Combined with Public Posting in High School Classrooms.* (Master's thesis). The University of Southern Mississippi, Hattiesburg, MS.

Zuckerman, E. L. (2005). *Clinician's thesaurus: The guide to conducting interviews and writing psychological reports* (6th ed.). New York, NY: Guilford Press.