

Managing Student Behavior in a Middle School Special Education
Classroom using CW-FIT Tier 1

Robyn K. Orr, EdS (candidate), robyn.k.orr@gmail.com

Paul Caldarella, PhD, paul_caldarella@byu.edu

Blake D. Hansen, PhD, blake_hansen@byu.edu

Brigham Young University

Provo, UT 84602

Howard P. Wills, PhD, hpwills@ku.edu

University of Kansas

444 Minnesota Ave #300

Kansas City, KS 66101

Author Note:

Address all correspondence to Paul Caldarella, PhD, Center for the Improvement of Teacher Education and Schooling, Brigham Young University, 149 MCKB, Provo, UT 84602. Telephone: 801-422-5081. Fax: 801-422-0199. Email: paul_caldarella@byu.edu

Ethical Statement:

The research reported in this manuscript was conducted in accordance with the ethical standards of the American Psychological Association. Approval was obtained from the Institutional Review Board of Brigham Young University, which also approved the informed consent procedures used. We declare no conflicts of interest. This research was funded in part by a grant from the Institute of Education Sciences and the U.S. Department of Education (R324A160279) awarded to the University of Kansas in cooperation with Brigham Young University. The opinions presented in this article are those of the authors, and no endorsement by the funding agency is intended or implied.

The citation information and final published version of this paper is as follows;

Orr, R. K., Caldarella, P., Hansen, B. D., & Wills, H. P. (2019). Managing student behavior in a middle school special education classroom using CW-FIT tier 1. *Journal of Behavioral Education*. Advance online publication. doi:10.1007/s10864-019-09325-w

Abstract

Middle school special education teachers often express concern about challenging student behavior. Class-Wide Function-Related Intervention Teams (CW-FIT), a behavior management program based on school-wide positive behavior support, has been effective in elementary general education classrooms. The present study, the first to apply it in a middle school special education setting, used an ABAC design to examine effects on student on-task behavior and teacher praise rates in a self-contained special education classroom for students with severe disabilities and their typically developing peer tutors. Results suggested that CW-FIT Tier 1 is associated with improvements in student on-task behavior and teacher praise rates, especially when peer tutors are included in the intervention and when the timer is silent. Both teachers and students reported the intervention to be socially valid. Study limitations and areas for future research are addressed.

Keywords: special education, middle school, classroom behavior management

Managing Student Behavior in a Middle School Special Education

Classroom using CW-FIT Tier 1

School-wide positive behavior interventions and supports (SWPBIS), the parent framework for Class-wide Function-related Intervention Teams (CW-FIT), can help decrease challenging behavior in middle school and in special education settings (Carr et al., 1999). Implemented according to a school's specific needs, SWPBIS utilizes positive behavior strategies in three tiers (Carr et al., 2002). Primary Tier 1 interventions focus on prevention, define and teach behavioral expectations, reward appropriate behavior, provide a continuum of consequences for problem behavior, and continuously collect and use data for decision making (Sugai & Horner, 2002; Sugai, Horner, & Lewis, 2009). Tier 2 provides additional behavior support for students who do not adequately respond to Tier 1. Tier 3 utilizes individualized education programs and functional behavior assessments to further support the approximately 5% of students whose needs are not met adequately from the first two tiers (Sugai & Horner, 2002).

Hawken and O'Neil (2006) recommended that in schools implementing SWPBIS, the students with disabilities should be included in interventions and assessments. Nearly 40% of students with Intellectual Disabilities (ID) spend at least half the day in special education classes (Office of Special Education Rehabilitative Services, 2016), and many students with severe disabilities spend much of their day in self-contained special education classrooms. SWPBIS programs with all three tiers should be implemented in these classrooms (Hawken & O'Neill, 2006). There is a need for additional studies in self-contained classrooms to obtain more accurate knowledge of SWPBIS applicability for all students (Carr, 2006).

One reason for limited evidence of SWPBIS in special education may be that some doubt students with severe disabilities would benefit from the same strategies as general education students. However, Hawken and O’Neill (2006) recommended that students with severe disabilities be grouped with typically developing students while expectations are being taught and that lesson plans review expectations to help students retain and generalize them. A student with a severe disability may imitate other students engaging in prosocial and respectful behavior (Carr, 2006) and thus benefit from the same behavior expectations as typically developing peers.

Many schools that claim to utilize SWPBIS fail to include students with severe disabilities (Hawken & O’Neill, 2006). In a survey of state coordinators who oversee 3,955 schools throughout the United States, 93% believed all students could benefit from SWPBIS, but 41% said students with severe disabilities were not addressed in trainings and personnel were not adequately prepared to meet their needs using SWPBIS (Landers, Courtade, & Ryndak, 2012). The respondents thought special educators would know how to deliver interventions; however, Carr (2006) was concerned about this attitude in schools utilizing SWPBIS, noting that the difficulty of modifying universal interventions for students with severe disabilities could create a climate leading to “devolution of SWPBIS into a bifurcated system of regular verses special education” (p. 55). Carr acknowledged that the only way to have SWPBIS support all students in the least restrictive environment is for special and general education teachers to be prepared to implement all levels of SWPBIS. If students are placed in a self-contained classroom due to low cognitive and adaptive functioning, the default behavior support they receive is likely a tertiary intervention that is not cumulative of primary and secondary interventions (Carr, 2006; Kurth & Enyart, 2016). This is not consistent with best practice recommendations that students must fail

to adequately respond to primary and secondary interventions before tertiary behavior interventions are introduced (Sugai & Horner, 2002).

Class-wide Function-Related Intervention Teams

CW-FIT is a multilevel group-contingency that serves “as a proactive approach in keeping with positive behavioral support intended to enable children’s school success . . . by addressing alterable contingencies at group and individual levels” (Wills et al., 2010, p. 165). It is function-related because the use of the interdependent group contingency addresses attention, which is a common function of behavior (Kamps et al., 2011). Other interdependent group contingencies, such as the Good Behavior Game (GBG), have been used in 9th grade special education classrooms (Flower, McKenna, Muething, Pedrotty Bryant, & Bryant, 2014; Salend, Reynolds, & Coyle, 1989). There are some differences between the GBG and CW-FIT. One of the most notable differences is that the GBG traditionally tracks the number of negative behaviors a group has, as infractions must stay beneath a certain number. The teacher does not tell the students how many infractions they are allowed to still be able to receive the reward, so it is difficult for students to self-monitor. The GBG also does not actively seek to increase the teacher’s rate of praise.

Wills et al. (2010) addresses the following components of implementing CW-FIT with fidelity. To address the most common reasons for problem behavior, CW-FIT teachers explicitly teach three social skills: getting the teacher’s attention, following directions, and ignoring peers’ inappropriate behavior (minimizing social reinforcement). Students are grouped as flexible teams. A child whose misbehavior is preventing the team from receiving points can be designated as a one-person team, allowing all to receive points and encouraging the child to comply. The teacher informs the class of the number of points they must earn to receive the

reward, posts the daily point goal, and awards points on a visual chart, so students can more easily monitor teacher ratings of their behavior. When the timer set by the teacher sounds (every three to five minutes), teams that have all members on task receive a point, along with behavior-specific praise. The timer helps the teacher remember to award points and shows students that expected behavior is consistently noticed and recognized. Points earned are never taken away. Teachers are also taught to increase behavior-specific praise and decrease reprimands. If students need further support, determined through observations that the student consistently contributes to their team not earning points, teachers can implement Tier 2, which consists of providing these students with a self-management chart to mark their own points and/or a help card to obtain assistance from peers or teachers.

A study testing CW-FIT on more than 35 classrooms in mostly Title 1 elementary schools (over 700 students; Wills et al., 2010) found that on-task behavior in three of the schools (16 classrooms) increased by 21.67%—leading to increased academic engagement. At-risk students using the Tier 2 intervention showed nearly a 50% reduction of disruptive behaviors. Most teachers in the study found that implementing CW-FIT helped them stay positive and the intervention “protects teaching time by increasing engagement, decreasing disruptions, and avoiding reactive or punitive strategies that can result in students being referred to the office or otherwise losing instructional time” (p. 169). Of the students, 85% reported that CW-FIT was fun, that their teacher was positive, and they liked working as a team to earn rewards.

CW-FIT has also been successful in elementary school special education classrooms. Bolt (2015) conducted a CW-FIT study in a class consisting of three students, two with ASD and one with “other health impairment” (OHI). All three students’ behavior improved: The number of disruptions decreased, and levels of engagement increased, particularly when the program

offered high rates of opportunities to respond. Weeden et al. (2016) implemented CW-FIT in a self-contained elementary school classroom for students with emotional and behavioral disorders (EBD). The participating teacher hesitated to stop CW-FIT after initial implementation, so the researchers included brief withdrawals between implementations. Baseline measures averaged 55% for on-task behavior; intervention measures averaged 90% for on-task behavior. Reprimand rates decreased by an average of 5.1 occurrences and praise rates increased by an average of 36.5 occurrences during the intervention phases. The researchers recommended conducting additional studies of CW-FIT in other special education classes.

Although CW-FIT has been successful in elementary general education and special education classes, there has only been one study examining CW-FIT in a middle school general education classroom: This study showed that two general education seventh grade classrooms' on-task behavior increased by almost 50% following the implementation of the intervention (Conklin, Kamps, & Wills, 2017). However, CW-FIT has yet to be evaluated in a middle school special education context.

The current study was conducted to better our understanding of how primary level SWPBIS can be implemented in special education classrooms. The primary research question examined what impact CW-FIT would have on students' on-task behavior when implemented in a self-contained special education classroom. Two secondary research questions were included: (a) Can a middle school special education teacher implement CW-FIT with fidelity, including increasing praise and decreasing reprimands? (b) Do the classroom teacher and students (both students with disabilities and their peer tutors) find CW-FIT socially valid?

Method

Setting and Participants

This study occurred in a self-contained special education classroom at a suburban Title I middle school in Utah, where 56% of students received free or reduced-price lunch. The teacher, a 28-year-old Caucasian female with a bachelor's degree in special education, had been teaching special education for eight years. The class consisted of 12 students with disabilities: 58% in seventh grade, 17% in eighth grade, and 25% in ninth grade. The classroom population consisted of 50% female and 50% male; 58.33% Caucasian, 33.33% Hispanic, and 8.3% Native American. Students in special education were classified with ID (58.33%), Specific Learning Disability (SLD; 16.67%), Autism Spectrum Disorder (ASD; 8.33%), Speech Language Impairments (SLI; 8.33%), or Other Health Impairment (OHI; 8.33%). While not typical to have students with SLD or SLI in a self-contained classroom, the school IEP teams had determined that this was the appropriate placement. Sixty-seven percent of the students had IQ scores below 71, and 33% had IQ scores in the 71-84 range. Similarly, 58% of the students had adaptive behavior composite scores that were low and 42% students had adaptive behavior composite scores that were moderately low, based on the Vineland Adaptive Behavior Scales, Second Edition (Sparrow, Cicchetti, & Balla, 2005). The range for students' reading achievement based on the Brigance (2010) was pre-primer to 3rd grade. Amongst other data collection, a rubric grading the student across multiple areas including social skills, academic skills, and classroom independence was also used to determine placement.

All secondary schools in the district had self-contained classrooms with typically developing peer tutors, who were all considered participants in the study. This class included 11 peer tutors: 73% were in 8th grade and 27% were in 9th grade; 36% were males and 64% were

females; 27% were Hispanic and 73% were Caucasian. One peer tutor received resource services for a developmental disorder during another class period. While tutors received coaching during the first week of school by a school staff member on how to tutor a student with a disability, the researchers did not train peer tutors in CW-FIT since they were not involved in any way with the implementation of the program. Tutors sat among their peers with disabilities to assist in academic areas, model good behavior, and build relationships. The teacher reported that influence and relationships of some past peer tutors had remained through high school. Peer tutors had also learned how to advocate for persons with disabilities.

Context

The teacher taught the special education students for the entire day, with peer tutors attending the class for an hour at a time as an elective. A minimum of six peer tutors and a mean of 20 students (including peer tutors) were present each day. CW-FIT was implemented during the 75-minute period that the teacher identified (via interview) as being the most behaviorally challenging—the period she taught functional independent life skills, such as using money. Before the intervention, the teacher organized the class into four groups according to those who seemed likely to work best together. She based this decision upon her prior observations of which students worked well together. There were three classroom aides whose primary job was to assist students with their classwork. They would occasionally remind students of classroom behavior expectations, but mostly assist with academics. Group desks were arranged as rectangles. Two of the lowest functioning students with disabilities were assigned specific peer tutors, but most peer tutors sat in the groups and helped whenever the teacher asked students with disabilities to complete a task. These two students with the most difficult behavior problems were at the front of the classroom with their assigned peer tutor. One student was often distracted

from her work and did not want to practice giving verbal answers. The researchers hypothesized that the function of her behavior was avoidance. The other student often giggled and tried to get her classmates' attention, so researchers hypothesized that the function was attention. No formal functional behavior assessment was conducted for either student.

Procedures

The teacher was selected after a school-wide recruitment meeting. Clearance was obtained from the school district and institutional review board (IRB) and approved consent forms were obtained. All researchers followed ethical protocol and procedures as approved by the IRB.

Baseline. Baseline data were collected during the teacher's normal classroom routines, during the same 75-minute period as the interventions. During the first week of school, the teacher had taught the steps of her various classroom rules ("Listen to directions the first time," "Keep body and objects to self," "Use kind words," "Work hard," and "Be prepared"), but had not reviewed the specific steps again after initially teaching them. During baseline, the rules were listed on the wall, and students with disabilities who struggled the most with compliance used tracking sheets (daily behavior chart) referencing them. These tracking sheets included a list of the teacher's classroom rules. The expectations were the same for each student and included answering *yes* or *no* to the following statements: I followed directions; I kept my body and objects to myself; I used kind words; I worked hard; I brought materials/notes from home. Peer tutors helped these students complete the tracking sheets. The teacher or peer tutor placed a sticker on the tracking sheet at the end of each period if the student followed these rules. Those students who had enough stickers could earn five minutes of free time at the end of the period.

Training. The researchers taught the teacher the rationales and intervention strategies for CW-FIT, conducting a two-hour training between baseline data collection and CW-FIT implementation. The teacher received scripted lessons to introduce the three target skills, with opportunities to role-play them using the accompanying posters, timers, and point charts, and to receive feedback from the researchers. The training included videos of teachers modeling Tier 1 of CW-FIT as integrated in everyday instruction. The three classroom adult aides did not attend the training, but were present when the intervention was explained to the students.

Intervention 1. As described below, the Tier 1 portion of CW-FIT (Wills et al., 2010) consists of teaching social skills and implementing an interdependent group contingency. The teacher did not use CW-FIT Tier 2 in either intervention phase, because the class on-task percentages during Tier 1 improved to an acceptable level and she did not identify any students in need of additional support. During this phase, the peer tutors sat amongst the students with disabilities, but the teacher did not address them when teaching expectations or include them in her decision to award team points. She did not focus on increasing praise or decreasing reprimands to the peer tutors during this phase. The peer tutors were not included in the first intervention, because the teacher's primary concern was the behavior of the students with disabilities and how this was negatively affecting the classroom environment. During this phase, students who had required tracking sheets during baseline continued to receive tracking sheets, as CW-FIT was implemented as an addition to already established classroom routines. CW-FIT provided the teacher with a Tier 1 intervention while she continued using the Tier 2 tracking sheets for some students.

Social skills lessons. The teacher began implementing CW-FIT Tier 1 by teaching her class the three social skills from the research-based version (Wills et al., 2010): (a) follow

directions the first time, (b) get teacher's attention the right way, and (c) ignore inappropriate behavior. In consultation with the research staff, she taught one 10-15-minute social skill lesson a day for three days, following the teaching script. The lessons included the rationale for each target behavior, explanations of the steps, role-playing by students, and teacher-class recitation of the steps. The written skills with corresponding pictures were displayed on posters visible to all students. After her first CW-FIT Tier 1 implementation, the teacher received feedback. Since the teacher was able to teach the skills with fidelity above 80%, no additional support was needed. Data were collected during the three days of training; throughout training and interventions, researchers were available for consultation, which she received four times, averaging two and a half minutes of discussion.

Teams. Students were assigned to one of five teams during CW-FIT. Each team had four to five students. This grouping is reflective of how previous CW-FIT studies structured their groups (Jolstead et al., 2016; Nelson et al., 2018; Wills et al., 2010). Teams were organized by ability level (academic and adaptive abilities), with approximately a 1:1 ratio of students with disabilities to peer tutors. One student was on a team with only her peer tutor because of her very low verbal abilities. Observers noted that both peer tutors and students with disabilities were often off-task.

Timer. As done in previous CW-FIT studies (Jolstead et al., 2016; Nelson et al., 2018; Wills et al., 2010) the teacher set an audible timer to beep at a variable interval schedule; the timer was set for three to five minutes.

Goals, points, and praise. At the beginning of each session, the teacher set a goal number of points for earning team rewards. The goal was based on 75% to 85% of total timer beeps, as has been done in previous CW-FIT studies (Jolstead et al., 2016; Nelson et al., 2018; Wills et al.,

2010). The point goal changed every day depending on how much time there was to play CW-FIT. For example, if there were 35 minutes for CW-FIT and the timer would beep 9 times, an appropriate point goal would be 7. At each timer beep, the teacher gave behavior-specific praise and awarded points to groups in which students in special education were displaying the social skills they had been taught. She would also occasionally praise and award points whenever she noticed good behavior regardless of the timer. The teacher tracked points on a wall poster clearly visible to all students.

Reward. At the end of CW-FIT Tier 1, during the last five minutes of class, the teacher tallied points to see which teams had met the point goal and earned the promised reward—either a tangible item or a desired activity. Only occasionally were there teams that did not achieve the class goal. The teacher chose two reward options each day and had the class vote on which they preferred. Common rewards included Bean Boozled, silent karaoke, rap battles, and watching video clips. These and similar activities were used in both intervention phases. Unstructured free time, the reward for the sticker chart, was not used as a reward for CW-FIT.

Withdrawal. A withdrawal phase occurred after the first intervention phase, during which the teacher removed the social skills posters, stopped reviewing the skills, stopped using the timer and point chart, did not identify the students by groups, and gave no points or rewards. She used the same procedures she had during baseline.

Intervention 2. After the withdrawal, the teacher once again implemented CW-FIT Tier 1 as she had during the intervention 1, but with two adaptations. First, she set the timer on silent vibrate during the timed intervals because researchers were concerned that many students were not paying attention until they heard the timer and then would stop paying attention after receiving their point. Second, the teacher involved the peer tutors more in this phase by

encouraging them to recite the rules along with the class and counted them in assigning points, making them a more integral part of the interdependent group contingency. Researchers decided to include the peer tutors in the second intervention, since their off-task behavior appeared to be contributing to the off-task behavior of the students with disabilities during intervention 1.

Post-intervention. Researchers met with the teacher after all data were collected. They showed the teacher graphs of on-task behavior, praise, and reprimand rates. The teacher and students completed a social validity survey.

Dependent Variables and Measures

Dependent variables and measures were chosen based on past CW-FIT studies (Jolstead et al., 2016; Nelson et al., 2018; Wills et al., 2010). Prior to the baseline phase, a university-based researcher used video-based training recordings to teach undergraduate and graduate student observers how to identify on-task behavior, praise, and reprimands. Watching videos of classrooms, observers marked groups of students as either on- or off-task, as well as tallying praise and reprimands. Each observation was compared against a key. When observers reached a minimum of 90% reliability, they were required to observe live classrooms with 90% reliability three times before collecting data for the study. Observers were positioned unobtrusively at the side of the classroom.

Group on-task behavior. The main dependent variable was student groups' on-task behavior. Observers identified on-task behavior as following directions, responding appropriately, asking questions appropriately, attending to the teacher, and working on an assignment. Off-task behavior included not following directions, not attending to the teacher, and talking out. On-task behavior was recorded only when all students in the group, including peer tutors in both phases, were on task; if only one student were off-task in a group, the whole group

was marked off-task. Paper and pencil methods were used to collect data in 20-minute observation increments at the beginning of class. Using a momentary time sampling method, every 30 seconds the observer looked at each group, one at a time, and marked it as on or off task before going to the next—quickly so children would not be aware of the observers' pattern. Teams were observed in the same order each session (Team 1, Team 2, then Team 3, etc.). To calculate a class on-task percentage, observers added the total number of on-task marks and divided it by the total number of observed intervals.

Treatment fidelity. A fidelity measure used in previous CW-FIT studies was used in the current study (Jolstead et al., 2016; Nelson et al., 2018; Wills et al., 2010). At the end of each observation period, before leaving the classroom, observers completed a 13-item treatment fidelity checklist by circling a “yes” or “no” option for components that demonstrated that the teacher was implementing the intervention correctly. Items included using the timer as expected, awarding points as earned, explaining to groups not earning points which expectation they had not met, and giving the reward at the end of the class period. Before entering the classroom, observers were trained to define and identify correct use of CW-FIT Tier 1 procedures and to reference the definitions in completing the sheets. Fidelity was calculated by dividing the number of “yes” responses recorded by the total number possible.

For all “yes” responses, observers rated the quality of use ($3 = \textit{full fidelity}$, $2 = \textit{good fidelity}$, $1 = \textit{partial fidelity}$). For example, the item “precorrects on skills at beginning of session” asks if the teacher reviewed the expectations and steps prior to starting the intervention. A “yes” response would be quality rated as follows: ($1 = \textit{Teacher minimally reviews skills}$, $2 = \textit{Teacher reviews some skills, but not all}$, $3 = \textit{Teacher reviews all skills}$). A startup fidelity form was also completed, evaluating whether the teacher sufficiently explained the intervention and taught the

social skills. Overall quality ratings were determined by adding the quality ratings and dividing by the total possible. During the training phase, the observers used a training fidelity checklist with its own expectations and definitions.

Behavior-specific praise and reprimands. Behavior specific praise was defined as any verbal statement acknowledging a correct response and suggesting approval: for example, “I like that everybody has eyes on me” and “Jed, thank you for raising your hand.” “That is correct, Leo” is a non-example, as what Leo actually did is not acknowledged. A reprimand was defined as a punitive statement or suggestion of displeasure in behavior: “This is your reminder to have eyes on me,” or “C.J., you need to sit back down.” The observers were trained to tally each statement of praise or reprimand directed to a student or group of students. The data were simultaneously collected, using paper and pencil methods, in 20-minute sessions with the group on-task behavior.

Social validity. When the study was completed, the teacher answered an 18-item social validity questionnaire that has also been used in previous CW-FIT studies (Jolstead et al., 2016; Nelson et al., 2018, Wills et al., 2010). Included were 15 items rated on a four-point Likert scale (1 = *very true* to 4 = *not true*) and three open-ended qualitative items asking what was most helpful, what could be improved, and what the teacher would change. She was asked if she found CW-FIT Tier 1 to be beneficial and practical for classroom use. Participating students completed a five-question social validity survey indicating their opinions of the intervention, as has also been done in past CW-FIT studies (Jolstead et al., 2016; Nelson et al., 2018; Caldarella, Williams, Hansen, & Wills, 2015). Due to the students’ overall low reading scores and the teacher’s concern that most students would not be able to write on or read the social validity questions, the researchers decided it would be most appropriate for an adult to help the students

complete these forms. The class was arranged in four groups, so that two graduate students, a classroom aide, and the teacher each sat with a group. The questionnaire consisted of two *yes* or *no* questions asking whether the students enjoyed the intervention and if they thought other children should play it in their classrooms. In addition, three open-ended questions asked what they did or did not like about CW-FIT Tier 1 and why others should or should not get it in their classrooms. The adult would read each question aloud and gave ample time for students to indicate their answers before moving on. The adults wrote the answers for the students with limited writing skills. For students whose verbal abilities were limited, teachers and researchers would offer options. For example, when a student shook his head “yes” to the question of whether other students should play CW-FIT Tier 1, the researcher asked “why” by giving two options, “fun” and “easy,” making fists that the child could touch to indicate his preference. Peer tutors helped before filling out their own questionnaire.

Interobserver agreement (IOA). IOA for all dependent variables was calculated during 57% of the sessions. IOA for on-task was calculated by dividing the number of intervals in which observers agreed by the total number of intervals observed: the range for on-task was 90%-100%, with an average of 94.07%. IOA was also calculated for treatment fidelity observations, occurrence and quality, by dividing the number of agreeing intervals by the total number of intervals. The average IOA was 99.18%, with a range of 91-100%. IOA for praise data was calculated by dividing the total number of praise statements recorded by one observer by the total number of praise statements noted by the second; IOA for reprimands utilized the same method. IOA for praise statements and reprimands averaged 82.69% (range = 50-100%) and 91.23% (range = 50-100%), respectively. Compared to on-task behavior, praise and

reprimands occurred less frequently making it easier for an observer to miss an occurrence thus resulting in lower IOA for these variables.

Design and Analysis

A single subject design (ABAC) was used. Five data points were collected in both baseline phases and during intervention 2. Three data points were collected during training, with six data points collected during intervention 1. Researchers collected a minimum of three data points demonstrating fairly consistent levels of on-task behavior, before changing phases. Visual methods were used to analyze the graphical data for teacher praise/reprimand ratios and group on-task behavior, examining level, trend, and variability. Researchers calculated an average fidelity score by analyzing information from the fidelity checklist. Tau-U was used to compute differences between baseline and intervention averages. Tau-U is a non-parametric statistic that provides an estimate of effect size by analyzing non-overlapping data points between phases, which is appropriate for single-subject research (Parker, Vannest, Davis, & Sauber, 2010). Researchers used a Tau-U calculator (www.singlecaseresearch.org/calculators/tau-u) to compute the effect size and statistical significance of changes across phases. Using the Tau-U calculator, the baseline data were compared with the training and intervention 1 data, intervention 1 data was compared to the second baseline data, and the second baseline data were compared to intervention 2 data. Descriptive statistics and qualitative coding were used to summarize teacher and student social validity questionnaires.

Results

Group On-Task Behavior

The primary research question for this study asked, “What impact does CW-FIT have on students’ on-task behavior?” Visual analysis was conducted on level, trend, and variability

within phases and on overlap and consistency between phases. Average group on-task data (Figure 1) began with a baseline of 67.85% ($SD = 8.69$) with an upward trend and high variability. Group on-task averages increased to 77.98% ($SD = 2.64$) during training with a slight downward trend and low variability. During intervention 1, group on-task behavior averaged 76.76% ($SD = 6.79$) with a slight upward trend and moderate variability. When CW-FIT was withdrawn, the group on-task average decreased to 68.38% ($SD = 2.1$) with a stable trend and low variability. During intervention 2 the group on-task average increased to 88.63% ($SD = 5.2$) with an increasing trend and moderate variability.

Changes in on-task behavior were not significant between the first baseline phase and intervention 1 ($Tau u = .467, p = .201$). Changes were statistically significant between intervention 1 and the second baseline phase ($Tau u = -.833, p = .02$), and between the second baseline phase and the intervention 2 ($Tau u = .92, p = .023$).

Treatment Fidelity

A secondary research question in this study asked, “Can a middle school special education teacher implement CW-FIT with fidelity, including increasing praise and decreasing reprimands?” During the baseline phase, the teacher naturally implemented Tier 1 components with 3.64% fidelity ($SD = 4.98$). During the training phase of the study, her performance was at 84.5% fidelity ($SD = 3.85$). She implemented each component with 100% fidelity except for asking “Which school/classroom rules does this match?” and “What other ways can you...?” both of which she neglected to do all three days. She also neglected to provide rationale (“Why is it important to...”) on the final day of training. Because fidelity was acceptable (above 80%) during all three days, and because the components she did not implement were not components of the intervention phases’ checklist, her fidelity was considered sufficient to progress to the next

stage. During intervention 1, the teacher implemented CW-FIT Tier 1 with 98.61% ($SD = 3.40$) fidelity: Her fidelity was 100% each day except the fourth, when two skills were at 91% rather than 100%: “Corrections are instructive and refer to skills” and “points were tallied for teams.” When the intervention was withdrawn, the teacher’s fidelity average was 7.25% ($SD = 7.6$); when intervention 2 was introduced, her fidelity averaged 96.79% ($SD = 4.4$). Because the teacher was expected to award team points when she praised, we also report the average number of points awarded in each phase: baseline = 0 points ($SD = 0$), training = 28 points ($SD = 10.58$), intervention 1 = 21 points ($SD = 2.34$), baseline = 0 points ($SD = 0$), and intervention 2 = 40.8 points ($SD = 4.32$).

Praise and Reprimand Rates. During baseline, the teacher praised the students an average of 6.6 times ($SD = 6.6$) with a moderate downward trend and high variability and reprimanded an average of 6.2 times ($SD = 2.38$) with a moderate upward trend and moderate variability (see Figure 2). During training the teacher praised students an average of 18.33 times ($SD = 10.01$) with a slight downward trend and high variability, reprimanding at an average of 9.67 times ($SD = 6.8$) with a slight downward trend and high variability. During intervention 1 of CW-FIT Tier 1, she praised students an average of 14.3 times ($SD = 8.5$) with a slightly upward trend and high variability, but reprimanded only an average of 4.67 times ($SD = 1.63$) with a slightly upward trend and low variability. While the intervention was withdrawn, the teacher averaged 10.8 praise statements ($SD = 9.78$) with a slight downward trend and moderate variability, along with 7.1 reprimands ($SD = 4.09$) with a slight upward trend and low variability. When intervention 2 of CW-FIT Tier 1 was implemented again, the praise rates averaged 26.88 ($SD = 13.98$), with a slight downward trend and high variability, while reprimands averaged 4.4 ($SD = 3.2$) with a slight downward trend and moderate variability. Similarly, a previous CW-FIT

study had an increase in praise after the second baseline phase, along with a downward trend (Nelson et al., 2018). Changes in average praise to reprimand ratios across phases are reported in Table 1.

Changes in praise rates were significant between the first baseline phase and intervention 1 ($Tau u = .80, p = .029$). When intervention 1 was compared to the second baseline phase, changes in praise rates were not significant ($Tau u = -.633, p = .083$). Between the second baseline phase and intervention 2, changes in praise rates were significant ($Tau u = .76, p = .047$). The same comparisons for reprimands showed only one with statistically significant differences: between the second baseline phase and the intervention 2 ($Tau u = -.76, p = .047$).

Social Validity

Teacher. The last question in this study asked, “Do teachers and students find CW-FIT socially valid?” The teacher’s responses were *mostly true* to whether she enjoyed being a CW-FIT teacher, whether using teams and assigning points for appropriate behaviors were helpful in improving students' behavior, whether she learned new skills, whether she will use the CW-FIT skills with future classes, whether her students enjoyed it, and whether her students were more focused and engaged during CW-FIT. Her answer was *very true* when asked if CW-FIT was easy to implement in her classroom and if she would recommend it to colleagues. She reported that reminders to give praise were helpful, though she indicated only *somewhat true* to whether using the timer was manageable during instruction. She noted that some students only got on task when they heard the timer, but when she put it on vibrate she benefited from being reminded without the sound of the timer prompting students to quickly get on-task. She also reported that explicitly going over classroom rules and expectations was helpful. The teacher would have found it helpful to modify the program for students with disabilities. She reported that, for some students,

simply giving any kind of communicative response was more important than raising their hands to get her attention. She suggested that the intervention would be better if teachers chose their own rules.

Students with disabilities. Ten of the special education students (83.33%) were surveyed. Researchers were unable to survey the two students who were absent, since a major school break started the next day. Of the ten students present, nine (90%) said that other students should have the opportunity to play CW-FIT. The most common reason given ($n = 7$) was that it is fun. Six (60%) of the students said that they liked playing CW-FIT. When they were asked, “What do you like about CW-FIT?” the most common answers were rewards/prize/treats ($n = 6$) and getting points ($n = 2$). When asked if there was anything they did not like about it, the most common responses were “boring” ($n = 2$) and “losing points” ($n = 2$). In CW-FIT, points are never taken away, but students who failed to earn points interpreted this as losing points. The one student who reported not liking CW-FIT, said, “It is hard to understand.”

Peer tutors. Seven peer tutors (63.64%) were surveyed, five (71%) of whom said that other classes should be able to play. Their explanations of why they recommended it for others were similar: “It is a great way to earn praise,” “It gets them to behave,” “They might be able to learn and listen easier,” “It might make them focus on their work if they know they are going to get a reward at the end of class,” and “It helps children pay attention.” The two students who said that others should not play CW-FIT remarked, “There may be a better, more fun way to help students interact in their learning” and “Because it's something special so it's magical in here.” Four peer tutors (57%) said that they liked playing CW-FIT. The most common specification of why they liked it was rewards/prize/treats ($n = 4$). Other answers included “I like how it made me and others pay attention” and “the students are interactive.” When peer tutors were asked if

there was anything that they did not like about it, answers varied: “the points,” “it takes a long time,” “the timer part,” “having to stop every few minutes,” “having to earn a lot of points,” and “I don’t like how if one person is not following rules the table misses a point.”

Discussion

The purpose of this study was to examine the impact CW-FIT had on students’ on-task behavior when implemented in a self-contained middle school special education classroom. Prior studies have found CW-FIT effective in improving on-task behavior and increasing praise in general education classrooms (Caldarella et al., 2015; Wills et al., 2010; Nelson et al., 2018) and elementary special education classrooms (Bolt, 2015; Weeden et al., 2016). This was the first study of CW-FIT implemented in a self-contained middle school special education classroom. The results suggest CW-FIT had a positive effect on the classroom.

First, group on-task behavior improved significantly during intervention phases. On-task behavior improved 9% from the first baseline to intervention 1. While this initial increase in on-task behavior was not as high as in prior CW-FIT studies, the rate was still higher than baseline. On-task behavior improved 20% from the second baseline to intervention 2, a result more consistent with prior studies (Caldarella et al., 2015; Nelson et al., 2018; Wills et al., 2010). Higher on-task behavior during intervention 2 may be due to the modifications made by the teachers, namely using a silent timer and including the peer tutors more during the intervention. In addition, the teacher praise to reprimand ratio was much higher during the intervention 2, as noted below.

Second, results showed that the teacher implemented CW-FIT Tier 1 with fidelity, consistent with fidelity results of prior studies (Caldarella et al., 2015; Jolstead et al., 2016; Nelson et al., 2018). Only two areas were not implemented with 100% fidelity and high-quality

ratings: “corrections are instructive and refer to skills” and “points tallied for teams” — both at 92%. All other areas were not only implemented, but also implemented well. While praise statements increased significantly during both intervention 1 and 2, not all reprimand decreases were statistically significant. During both baseline phases, one praise statement was given for every reprimand. During intervention 1, the praise per reprimand ratio was 3:1; during intervention 2, it was approximately 6:1—consistent with other studies finding that praise-to-reprimand ratios improve significantly during CW-FIT (Jolstead et al., 2016; Nelson et al., 2018; Wills et al., 2010). One hypothesis for the increase in praise during intervention 2 is that the teacher included the peer tutors and expressed that she felt more confident with CW-FIT than during intervention 1. It is unclear why there was a downward trend in praise rates during intervention 2, though a previous CW-FIT study had similar findings (Nelson et al., 2018).

Finally, the teacher and students found CW-FIT Tier 1 to be socially valid—also consistent with previous studies’ findings (Jolstead et al., 2016; Nelson et al., 2018; Wills et al., 2010). The teacher found CW-FIT easy to implement and enjoyed using it in her classroom, though she did not like the audible timer and felt the classroom expectations may have been too high for her students. A large majority of both the students in special education and the peer tutors liked playing CW-FIT and thought other students should play it, liking the prizes and finding the activity fun.

Limitations and Areas for Future Research

While the results of this study were positive, there were several limitations. First, on-task behavior data were collected for mixed student groups consisting of those with disabilities and typically developing peer tutors. As students with disabilities and peer tutors interact, researchers thought it was appropriate to measure the combined impact of their behaviors. While researchers

did not gather separate data for the two groups, anecdotally researchers observed that the peer tutors were often more off-task than the students with disabilities. Researchers hypothesized that, because the teacher taught life skills that the peer tutors likely already knew, they would stop paying attention while the teacher taught. While it is natural to measure both peer tutors and students with disabilities' behavior in the classroom, since the behaviors can influence each other, it is recommended that future researchers study CW-FIT in self-contained classrooms without peer tutors.

Second, this study implemented only Tier 1 of CW-FIT. Although prior studies implemented Tier 2 for students who made little progress with Tier 1 (Caldarella et al., 2015; Nelson et al., 2018; Wills et al., 2010), other CW-FIT studies, including the present study, utilized only Tier 1 (Bolt, 2015; Jolstead et al., 2016; Kamps et al., 2015). The teacher in the present study did not think Tier 2 was necessary for her class; she also explained that her students with disabilities would require assistance using the help cards and self-management charts and that any peer tutors needing Tier 2 would require extra time helping their assigned students. We recommend that both tiers be implemented in special education classes that include individuals or groups of students who need an intervention that is more intensive and do not need major assistance managing the Tier 2 materials. The question of whether students with severe disabilities can be “fully included in primary and secondary levels of support” without “intensive, individualized” support remains unanswered (Hawken & O’Neill, 2006, p. 52).

Due to the ABAC design of this study, and lack of replications, we cannot be certain about causality thus the results were inconclusive. However, there appears to be an association between the delivery of CW-FIT and improved on-task behavior. Researchers originally intended to do an ABAB design. However, given the changes needed in intervention 2 to improve

effectiveness, the study changed to an ABAC. Because the school year ended shortly thereafter, there was not time to conduct additional replications. The teacher significantly increased the amount of praise and points given during the intervention 2, which was associated with increased group on-task behavior. Another difference was that the timer was inaudible for intervention 2, though this may be a necessary alteration in a primary intervention for students with moderate to severe disabilities. We recommend that future researchers examine the effects of manipulating the amount of praise and points given, as well as the use of audible versus non-audible timers, during CW-FIT implementation, to determine the specific effects of such manipulations on student behavior.

A fourth limitation was that social validity was not collected from all students in the same way. The peer tutors completed surveys on their own, whereas the students with disabilities received the help of others (i.e., tutors, aides, teacher). However, this was necessary given the adaptive level of the students with disabilities. For students with verbal limitations, the three open-ended questions were reduced to two options that researchers had seen as common answers to the open-ended questions in previous CW-FIT studies. However, students may have wanted to choose different answers but did not have the opportunity to do so. The use of peer tutors in this stage made the process more efficient but could have biased students' ratings. In the future, researchers should systematically survey students one at a time to answer such questions.

Fifth, the teacher used a Tier 2 intervention (i.e., tracking sheets) in both the baseline and intervention phases. While consistent throughout phases, it would have been appropriate to have the teacher stop using the tracking sheets during all phases so that CW-FIT as a Tier 1 intervention could be compared to no interventions during baseline, instead of a Tier 2 intervention during baseline, and a Tier 1 and Tier 2 intervention during intervention phases. It is

most appropriate to implement Tier 1 interventions with at least 80% success and add then Tier 2 interventions for students who do not respond to Tier 1 alone (Sugai, Horner, & Gresham, 2002).

A final limitation was that the study was conducted in only one classroom with one teacher. Caution is advised in generalizing from one to other classrooms. Most of the students in the special education class had ID, so generalizing the outcomes to populations with other disabilities should be done so with caution. This study should be replicated for students with more severe disabilities and those with mild/moderate disabilities. It should also be replicated when the entire school is utilizing CW-FIT, so that results in special education and general education classrooms can be compared.

Implications

This study helps fill the gap in empirical evidence supporting the benefits of SWPBIS primary interventions for students in self-contained classrooms. Tier 1 of CW-FIT was shown to be feasible and effective in helping a struggling middle school special education teacher in a self-contained classroom improve her praise-reprimand behavior and increase on-task student behavior for both general education students and students receiving special education services. During the two interventions, the teacher implemented the intervention with fidelity and increased her praise-reprimand rates. Although replications are needed to confirm CW-FIT effectiveness, including iterations using Tier 2 and ABAB studies, results of this study suggest optimism for middle school students in self-contained classes when interventions include predictable environments with rationales for expectations, posted expectations, opportunities to practice expectations, instruction in building social skills, and meaningful reinforcement (Mitchell, 2014).

References

- Bolt, T. D. (2015). *The effects of varied opportunities to respond embedded in a group contingency program* (Master's thesis). Retrieved from <http://scholarsarchive.byu.edu/studentpub/165>
- Brigance, A. H. (2010). *Brigance Inventory of Early Development II (IED II)* [Assessment instrument]. North Billerica, MA: Curriculum Associates
- Caldarella, P., Williams, L., Hansen, B., & Wills, H. (2015). Managing student behavior with class-wide function-related intervention teams: An observational study in early elementary classrooms. *Early Childhood Education Journal*, *43*, 357–365.
doi:10.1007/s10643-014-0664-3
- Carr, E. G. (2006). SWPBIS: The greatest good for the greatest number, or the needs of the majority trump the needs of the minority? *Research and Practice for Persons with Severe Disabilities*, *31*, 54-56.
- Carr, E. G., Dunlap, G., Horner, R. H., Koegel, R. L., Turnbull, A. P., Sailor, W., & Fox, L. (2002). Positive behavior support: Evolution of an applied science. *Journal of Positive Behavior Interventions*, *4*, 4–16. doi: 10.1177/109830070200400102
- Carr, E. G., Horner, R. H., Turnbull, A. P., Marquis, J. G., McLaughlin, D. M., McAtee, M. L., & Braddock, D. (1999). Positive behavior support for people with developmental disabilities: A research synthesis. *American Association on Mental Retardation Monograph Series*. Washington, DC: American Association on Mental Retardation.
- Conklin, C. G., Kamps, D., & Wills, H. (2017). The effects of Class-Wide Function-related Intervention Teams (CW-FIT) on students' prosocial classroom behaviors. *Journal of Behavioral Education*, *26*, 75-100 DOI: 10.1007/210864-016-9252-5

- Flower, A., McKenna, J., Muething, C. S., Bryant, D. P., & Bryant, B. R. (2014). Effects of the good behavior game on classwide off-task behavior in a high school basic algebra resource classroom. *Behavior Modification, 38*, 45–68. doi:10.1177/0145445513507574
- Hawken, L. S., & O’Neill, R. E. (2006). Including students with severe disabilities in all levels of school-wide positive behavior support. *Research and Practice for Persons with Severe Disabilities, 31*, 46-53
- Jolstead, K. A., Caldarella, P., Hansen, B. D., Korth, B. B., Williams, L., & Kamps, D. (2016). Implementing positive behavior support in preschools: An exploratory study of CW-FIT Tier 1. *Journal of Positive Behavior Interventions, 19*, 48–60.
doi:10.1177/1098300716653226
- Kamps, D., Wills, H., Dawson-Bannister, H., Heitzman-Powell, L., Kottwitz, E., Hansen, B., & Fleming, K. (2015). Class-wide function-related intervention teams “CW-FIT” efficacy trial outcomes. *Journal of Positive Behavior Interventions, 17*, 134–145.
doi:10.1177/1098300714565244
- Kamps, D., Wills, H. P., Heitzman-Powell, L., Laylin, J., Szoke, C., Petrillo, T., et al. (2011). Class-wide function-related intervention teams: Effects of group contingency programs in urban classrooms. *Journal of Positive Behavior Interventions, 13*, 154–167.
doi:10.1177/1098300711398935.
- Kurth, J. A., & Enyart, M. (2016). Schoolwide positive behavior supports and students with significant disabilities: Where are we? *Research and Practice for Persons with Severe Disabilities, 4*, 216-222.

- Landers, E., Courtade, G., & Ryndak, D. (2012). Including students with severe disabilities in school-wide positive behavior supports: Perceptions of state coordinators. *Research and Practice for Persons with Severe Disabilities, 37*, 1-8.
- Mitchell, D. (2014). *What really works in special and inclusive education: Using evidence-based teaching strategies*. New York, NY: Routledge.
- Nelson, M. A., Caldarella, P., Hansen, B. D., Graham, M. A., Williams, L., & Wills, H. P. (2018). Improving student behavior in art classrooms: An exploratory study of CW-FIT Tier 1. *Journal of Positive Behavior Interventions, 20*, 1-12. doi: 10.1177/1098300718762744
- Office of Special Education Rehabilitative Services. (2016). *Thirty-eighth Annual Report to Congress on the Implementation of the Individuals with Disabilities ACT (IDEA)*. Washington, DC: OSERS
- Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2010). Combining non-overlap and trend for single case research: Tau-U. *Behavior Therapy, 42*, 284-299. doi: 10.1016/j.beth.2010.08.006
- Salend, S. J., Reynolds, C. J., & Coyle, E. M. (1989). Individualizing the good behavior game across type and frequency of behavior with emotionally disturbed adolescents. *Behavior Modification, 13*, 108–126. doi:10.1177/01454455890131007
- Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. (2005). *Vineland Adaptive Behavior Scales, Second Edition (Vineland-II)*. Circle Pines, MN: American Guidance Service.
- Sugai, G., & Horner, R. H. (2002). Introduction to the special series on positive behavior support in schools. *Journal of Emotional and Behavioral Disorders, 10*, 130–135. doi:10.1177/10634266020100030101

- Sugai, G., Horner, R. H., & Gresham, F. (2002). Behaviorally effective school environments. In M. R. Shinn, G. Stoner, & H. M. Walker (Eds.), *Interventions for academic and behavior problems: Preventive and remedial approaches* (pp. 315–350). Silver Spring, MD: National Association for School Psychologists.
- Sugai, G., Horner, R. H., & Lewis, T. (2009). *Is school-wide positive behavior support an evidence-based practice?* Retrieved from Positive Behavior Interventions and Supports OSEP Technical Assistance Center website: <http://www.pbis.org/research>
- Weeden, M., Wills, H. P., Kottwitz, E., & Kamps, D. (2016). The effects of a class-wide behavior intervention for students with emotional and behavioral disorders. *Behavioral Disorders, 42*, 285–293. doi: 0.17988/bd-14-12.1
- Wills, H. P., Kamps, D., Hansen, B., Conklin, C., Bellinger, S., Neaderhiser, J., & Nsubunga, B. (2010). The class-wide function-based intervention team program. *Preventing School Failure, 54*, 164–171. doi: doi.org/10.1080/10459880903496230

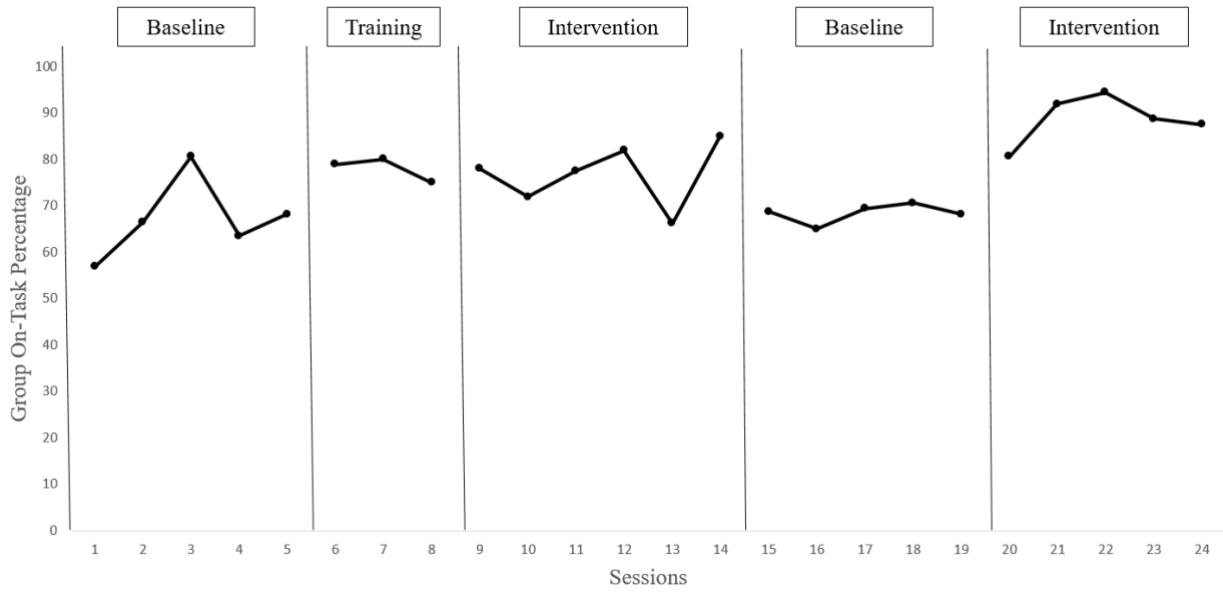


Figure 1: Average student group on-task behavior percentages across phases

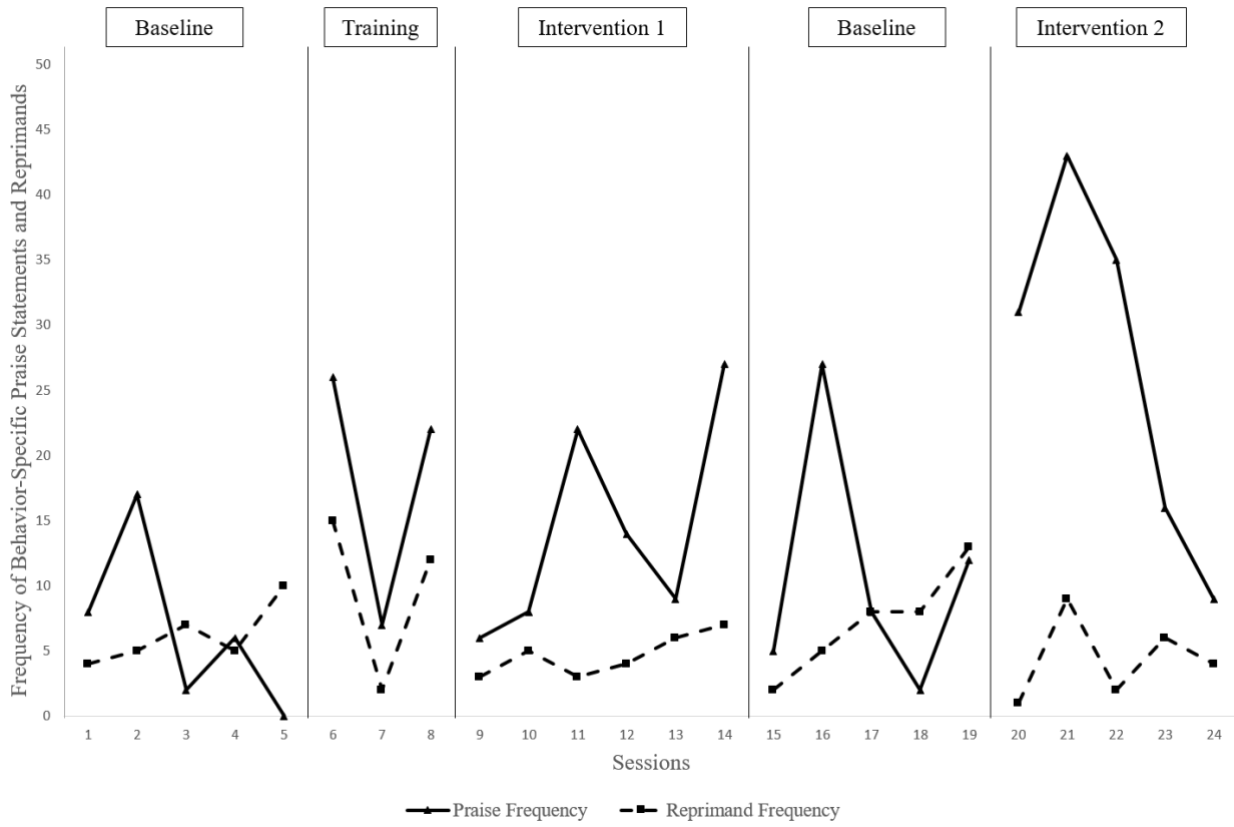


Figure 2: Comparison of behavior-specific praise statements and reprimands across phases

Table 1

Average teacher praise to reprimand ratios across phases

	Baseline	Training	Intervention 1	Baseline	Intervention 2
Average teacher praise to reprimand ratio	1.06:1	1.89:1	3.06:1	1.51:1	6:1
