Effects of a Behavior Management Strategy, CW-FIT, on High School Student and Teacher Behavior

Renee Speight1 · Suzanne Kucharczyk1 · Peggy Whitby1

Accepted: 23 December 2020 / Published online: 11 January 2021
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021, corrected publication 2022

Abstract
Challenging classroom behavior can interfere with learning. Fortunately known, positive, and proactive approaches to behavior management can improve student responding. Class-Wide Function-related Intervention Teams (CW-FIT) have led to improvement in student behavior across elementary and middle school contexts. However, little is known of the impact of the intervention on high school student behavior. This study evaluated CW-FIT’s utility in improving high school student and high school teacher behavior in a co-taught learning environment. A single-subject withdrawal design was used to evaluate improvements in on-task behavior for 14 high school students in one co-taught classroom. The impact on praise and reprimand statements of two high school teachers was also assessed. The findings showed improvement to student and teacher behavior and sustainability of the intervention. Further, teachers and students expressed satisfaction with the intervention and teachers maintained high levels of implementation fidelity. Limitations of the evaluation and areas for future research are presented.

Keywords On-task behavior · Interdependent group contingency · Behavior management · High school · Co-teaching

Introduction
Despite significant literature to inform implementation of evidence-based classroom management strategies to improve student behavior (Simonsen et al. 2008), teachers often lack the knowledge or training to implement effective classroom management strategies reducing problem behavior responses (Moore et al. 2017). Because display of off-task or disruptive behavior can impede learning for all students in the
classroom, supporting teachers in implementation of effective management methods is essential. This is of particular importance in high school contexts which report the highest rates of suspensions and expulsions used to manage challenging behaviors (Flannery et al. 2014). Group contingency interventions (e.g., Good et al. 2019) have proven effective at decreasing behavior in high school settings, although issues of sustainability and adherence to principles of positive behavior support remain. This study addresses the need for further evaluation of group contingencies in high school contexts by evaluating the utility of an interdependent group contingency, Class-Wide Function-related Intervention Teams (CW-FIT), in improving on-task behavior of high school students in a co-taught learning environment (Wills et al. 2019).

**Group Contingencies**

Group contingencies are an effective strategy to manage student classroom behavior (Maggin et al. 2017; Simonsen et al. 2008). Group contingencies involve clear antecedents and consequences to differentially strengthen behavioral responses (Litoe and Pumroy 1975). To implement group contingencies in classroom environments, a behavior criterion is set for a group of students. When the set criteria are met, students earn access to the reward individually or as a group. There are three types of group contingencies: dependent, independent, and interdependent (Litoe and Pumroy 1975). For dependent group contingencies, group access to the reward is contingent on the behavior of one or more selected individuals within the group. For independent group contingencies, individual access to the reward is contingent on the behavior of the individual. For interdependent group contingencies, group access to the reward is contingent on behavior of the entire group. Of the three, interdependent group contingencies may be more preferred by teachers due to increases in cooperative behavior among students (Groves and Austin 2017). However, few evaluations of interdependent group contingencies have included high school students (Maggin et al. 2017).

To date, the impact of one interdependent group contingency the Good Behavior Game (GBG) has been evaluated in high school contexts and has shown improvement in high school student behavior such as decreasing talking out of turn (Stratton et al. 2019), off-task behavior (Flower et al. 2014), inappropriate contact with objects and classmates, inappropriate noises, seat leaving (Donaldson et al. 2015), and aggression (Kleinman and Saigh 2011). To implement the GBG, participating students learn the behavior expectations, are divided into teams, and work together to gain access to a reward (Donaldson et al. 2015). In high school contexts, the GBG has been implemented with a punishment procedure in which teams are given points for demonstration of behaviors not consistent with the behavior expectations (Donaldson et al. 2015; Flower et al. 2014; Mitchell et al. 2015; Kleinman and Saigh 2011; Stratton et al. 2019). Despite improvement in behavior with implementation of the GBG, the intervention has not demonstrated sustainability in high school contexts (Flower et al. 2014; Kleinman and Saigh 2011). Further, the punitive procedures utilized are not well-aligned with positive behavior management systems.
Kleinman and Saigh (2011) suggest it is necessary to determine if similar effects can be observed with interdependent group contingencies that reward demonstrations of appropriate behavior.

**Class-Wide Function-Related Intervention Teams**

Recently, researchers have evaluated an interdependent group contingency, Class-Wide Function-related Interventions (CW-FIT), with some components similar to the GBG (e.g., identification of target behaviors, interdependent group contingency, group access to reward). In contrast to the GBG’s delivery of points for interfering behaviors, CW-FIT emphasizes teaching appropriate behavior and rewarding students for the demonstration of such behavior through delivery of points for target replacement behaviors (Caldarella et al. 2015; Kamps et al. 2015; Wills et al. 2018, 2019). To implement CW-FIT, teachers first utilize direct instruction to teach students expected classroom behaviors. Students are then grouped into teams as part of a game to earn points. Teams are awarded points at the sound of a timer set at an interval of 3–5 min. When the timer beeps, the teacher scans the classroom, vocally acknowledges demonstration of expected behaviors and records points for teams on a team point chart. During game play, teachers are expected to increase levels of behavior-specific praise. At the end of the game, teams who met the point goal criteria established at the onset of the game earn access to a highly preferred item, activity, or edible from a reinforcer menu.

CW-FIT has strong empirical support in elementary classrooms with emerging evidence in middle school settings. Kamps et al. (2015) evaluated the impact of CW-FIT across 17 elementary schools. With implementation of the intervention, Kamps et al. noted significant improvement in whole class behavior with on-task behavior increasing over 30% from baseline. Additionally, teacher behavior showed improvements with increased praise statements and decreased reprimands (Kamps et al. 2015). Further, teachers maintained high levels of treatment fidelity, reported high levels of satisfaction with the intervention, and nearly half of participating teachers continued to implement the intervention beyond study phases. In an evaluation across four classrooms, Conklin et al. (2017) reported increases in appropriate classroom behavior (e.g., on-task behavior, hand raising, and compliance) and decreases in inappropriate behavior (e.g., out of seat behavior and talking out) for kindergarten, second grade, and seventh-grade students. The researchers also reported increases in teacher praise and slight decreases in teacher reprimand. Wills et al. (2018) noted similar increases in on-task behaviors and decreases in disruptive behaviors when evaluating CW-FIT across 324 elementary students in a randomized control trial. As with prior investigations, participating teachers and students rated the intervention favorably and teachers maintained high levels of fidelity.

Emerging evidence supports the utility of CW-FIT in middle school settings (Conklin et al. 2017; Orr et al. 2019; Speight et al. 2020; Wills et al. 2019). Orr and colleagues evaluated CW-FIT using a single-subject withdrawal design and noted improvements in on-task behavior for 12 middle school students in a self-contained classroom. Further, increases in praise statements and decreases in reprimands were
also observed. The teacher also maintained high levels of fidelity of implementation. The teacher and students also rated the intervention favorably. Despite improvement in teacher behavior and student behavior, the authors note caution is warranted given the research ABAC research design. In another study, Wills et al. (2019) evaluated CW-FIT in three middle school classrooms. With application of the intervention, student on-task behavior increased in all classrooms. Teacher behavior improved in two classrooms with increased praise. Further, teachers maintained high levels of implementation fidelity and consumers (i.e., teachers and students) viewed the intervention favorably. Speight et al. (2020) found similar results when implementing CW-FIT in an ethnically and socioeconomically diverse middle school. On-task behavior and teacher praise showed improvement with CW-FIT. Further, teachers and students rated the intervention favorably.

Despite the positive effects in elementary and middle school settings, little is known of the efficacy of CW-FIT in high school contexts. The purpose of this study was to evaluate the generalizability of the effects of CW-FIT on high school populations. Specifically, this study addressed the following research questions: (a) How does CW-FIT impact high school students on-task behavior class-wide? (b) How does CW-FIT impact high school teacher behavior? (c) Can high school teachers implement CW-FIT Tier 1 with fidelity? (d) Are high school teachers and students satisfied with CW-FIT as a classroom management strategy?

Method

Setting and Participants

This study was conducted at a high school in the Southern USA. Informed consent was obtained for 14 students (58%) of a ninth-grade English language arts class. The class was co-taught which is defined as a general education and special education teacher partnering to instruct and support students with and without disabilities (Friend et al. 2010). Of the study sample, five students (36%) were ethnically diverse, two students (14%) received free or reduced lunch, three students (21%) were English language learners, and five students (36%) had IEPs. Two high school teachers volunteered to participate in the study. Teacher 1 was a special education teacher with a bachelor’s degree completing his first year of teaching. Teacher 2 was a general education teacher with a master’s degree completing her fourth year of teaching.

Additionally, three target students were identified by the special education teacher as at-risk for externalizing or internalizing behavior using Stage 1 of the Systematic Screening for Behavior Disorders (SSBD; Walker and Severson 1992). Stage 2 and Stage 3 of the SSBD were not completed due to time constraints. All target students were 14 years old at the time of the study. Target 1, nominated for demonstrating externalizing behavior, was an English language learner and received IEP services under the IDEA autism category. Target 2, nominated for demonstrating externalizing behavior, was also an English language learner. Target 3, nominated
for demonstrating internalizing behavior, received IEP services under the IDEA autism category.

Context

The instructional period lasted 90 min with teachers implementing the game for approximately 80 min after the initial ten minutes of individual silent reading. Students were seated at separate desks, clustered in groups of three or four, facing or sitting adjacent to each other. Data were collected across various literacy-based classroom activities such as students attending to teacher-led instruction (e.g., modeling, read aloud), students completing tasks in small groups (e.g., completing worksheets, creating posters), and students completing tasks individually (e.g., completing worksheets, annotating text, writing essays). During classroom activities, student materials included individual laptops, texts, and printed or digital worksheets.

Dependent Variables

Dependent variable data were coded during 30-min live observation sessions when at least 10 of the 14 students with consent were present. Baseline session four was terminated after 25 min when only nine students remained in the classroom. Data were collected on three dependent variables: (a) whole group on-task behavior, (b) target student on-task behavior, and (c) teacher praise and reprimand statements.

A 30-s momentary time-sampling procedure was used to record occurrence of on-task behavior for whole group and target students. On-task behavior was defined as: (a) attending to the teacher or materials (e.g., looking at the teacher or materials, taking notes, waiting for instruction); (b) completing tasks (e.g., reading materials, writing, or participating in group discussions); and (c) following teacher instruction (e.g., gathering materials). At the end of each 30 s interval, observers scanned each group and coded “+” if all group members were demonstrating on-task behavior. Observers recorded “−” if one or more students in each group were observed not demonstrating on-task behavior. Target student data were recorded immediately after the coding of whole group on-task behavior following the same 30 s momentary time-sampling procedure.

Behavior-specific praise (e.g., “Nice work showing respect by staying on-task.”) and reprimands (e.g., “Shhhh”, “Sit down.” “Stop talking.”) were coded as frequency. Upon occurrence across each live observation session, observers tallied praise and reprimand statements across both the Teacher 1 and Teacher 2.

Procedural Fidelity

Consistent with previous evaluations of CW-FIT (e.g., Orr et al. 2019; Wills et al. 2019) fidelity measures were completed after each session of implementation. Nine intervention components (e.g., classroom expectations posted, team point chart displayed, daily point goal posted, pre-corrects on skills, timer used and set on appropriate intervals, points awarded to teams, praise-to-reprimand ratio 4:1, behavior/
skill specific praise and reprimands, reward delivered) were rated on a scale of one to three to render a score out of 27. Fidelity data were used to provide constructive feedback to teachers on implementation of the procedures immediately following the intervention and to pre-correct teacher behavior before implementation of the intervention. Additionally, procedural fidelity IOA were recorded 42.9% of intervention phase one, and 50% of intervention phase two.

**Inter-Observer Agreement**

During recording of IOA, two data collectors, the primary researcher and a secondary observer, simultaneously recorded data on the dependent variables during live classroom observations. Two secondary observers (graduate student and university faculty member) were trained by the primary researcher until agreement criteria of 90% was met. For IOA training, secondary observers memorized dependent variable definitions and collected data in the target classroom during 10-min live observation sessions using the dependent variable coding procedures. To code IOA, the primary researcher named each group and target student, “Group one, group two…, target one, target two…,” and both observers scanned and recorded the behavior of the named group or student.

Across all experimental phases for the full duration of observational data coding (i.e., 30 min), IOA was collected at least 33% of sessions for whole group on-task, teacher praise statements, and teacher reprimand statements, and at least 25% for target student data due to student absences. For whole group on-task data, IOA averaged 95% (range = 93–96%) during baseline, 95% (range = 92–97%) during treatment phase one, 83% during withdrawal, and 94% (range = 90–98%) during treatment phase two. For teacher reprimands IOA averaged 83% (range = 50–100%) during baseline, 74% (range = 50–100%) during treatment phase one, 80% during withdrawal, and 8% (range = 0–33%) during treatment phase two. For teacher praise, IOA averaged 50% (range = 0–100%) during baseline, 88% (range = 67–100%) during treatment phase one, 33% during withdrawal, and 90% (range = 71–100%) during treatment phase two. Low frequency of praise and reprimand statements combined with the missed recording of a single instance of behavior could have attributed to lower levels of agreement. For target student data, IOA averaged 97% (range = 93–100%) during baseline, 96% (range = 93–100%) during treatment phase one, 90% (range = 86–95%) during withdrawal, and 97% (range = 90–100%) during treatment phase two.

**Consumer Satisfaction**

To assess consumer satisfaction, teachers and students completed surveys after the final session of the second intervention phase. A modified version of the Intervention Rating Profile-15 (IRP-15) was used to evaluate teacher satisfaction (Martens and Witt 1982). Reliability of the IRP-15 has been reported as 0.98 using Cronbach’s alpha (Martens et al. 1985). Teachers indicated their satisfaction with the intervention by selecting one of six options across 15 statements: 1 = *strongly disagree*,
2 = disagree, 3 = slightly disagree, 4 = slightly agree, 5 = agree, 6 = strongly agree (Martens and Witt 1982). Statements rated by teachers included “this would be an acceptable intervention for children’s problem behavior,” “I would be willing to use this intervention in the classroom setting,” and “I like the procedures used in this intervention.” A modified version of the Children Intervention Rating Profile (C-IRP) was used to evaluate student satisfaction (Kratochwill 1985). Reliability of the C-IRP has been reported to range from 0.75 to 0.89 (Finn and Sladeczek 2001). Students completed the survey by indicating agree or disagree across seven statements (Kratochwill 1985). Statements included “CW-FIT was a fair way to deal with behavior,” “CW-FIT is a good game to use with other kids,” and “I like CW-FIT.” As an additional measure of consumer satisfaction, data were collected across two follow-up sessions six weeks and 20 weeks after the final intervention phase to assess sustainability.

Research Design and Procedures

Research Design

A single-subject withdrawal ABAB (i.e., baseline, intervention, withdrawal, and intervention) design was used to evaluate the impact of CW-FIT. Phase changes occurred when stability across three data points was shown in whole group on-task data with at least three data points in each phase to meet standards with reservation (Kratochwill et al. 2010). Stability was defined as data points falling within a 20% range with no evidence of a trend (i.e., data points increasing or decreasing). Data were collected two to three times each week across all phases of the study. The length of each phase of the study ranged from two weeks (e.g., baseline, withdrawal, intervention phase two) to three weeks (e.g., intervention phase one).

Baseline

Baseline data were collected across five sessions. During baseline, Teacher 1 and Teacher 2 maintained their typical classroom procedures (i.e., reprimands, prompting task completion, and classroom removal). Data were collected at the time indicated by teachers in which students engaged in the highest rate of interfering behavior. This time period occurred after students completed their silent reading assignments which took place at the beginning of class.

CW-FIT Implementation

The intervention was integrated into the classroom system in three phases: teacher training, student training, and implementation of the interdependent group contingency and token system. Teacher 1 and Teacher 2 were trained to implement CW-FIT during a 45-min training session in Teacher 2’s classroom prior to implementation of the intervention. During training, the primary investigator reviewed the components of the CW-FIT intervention (a) behavior expectation posted
prominently in view of all students (Fig. 1); (b) clear display of team point chart; (c) team point chart includes daily point goal; (d) target skills pre-corrected by teachers (e.g., “What does it look like to be respectful in our classroom?”); (e) audible timer set to sound at 3–5 min intervals throughout the game; (f) teams awarded points for demonstrating the expectations; (g) a 4:1 praise to reprimand ratio maintained by teachers; (h) skill specific praise and corrective feedback; and (i) reward delivered at the end of game to groups meeting criteria. Additionally, during training, the teachers viewed videos depicting implementation of the intervention and reviewed the fidelity tool and the expectations instruction script. During implementation of CW-FIT, follow-up coaching was provided when fidelity levels dropped below 80%. Lower levels of fidelity were attributed to teachers not maintaining a 4:1 praise to reprimand ratio, absence of pre-correction of target behavior, and non-specific praise and reprimands.

After teachers were trained to implement the intervention, the teachers taught the students the expected behavior (i.e., Be Respectful). To teach the behavior, the teachers discussed what the behavior looked like while recording descriptions on a
poster (see Fig. 1), provided students the opportunity to discuss why it was important to demonstrate the behavior in the classroom setting, directed the students to demonstrate examples and non-examples of the behavior, and displayed the poster describing the behavior on the wall. After teaching the target behavior, the classroom teachers shared with the students they would be playing a game in class called CW-FIT. The students learned they would be working in teams to meet a goal and upon meeting that goal, and they would have access to a reward. The teachers then worked with the students to identify immediate (e.g., free-time, device-time, snacks) and delayed (e.g., class party, gym time) tangibles or activities to create a reinforcement menu.

After identification of reinforcers, the teachers initiated CW-FIT. To begin game play, the teachers set the goal, identified the reward, and provided pre-corrects (i.e., reminded students to demonstrate respect to earn points). Then, the Teacher 1 set an audible timer on his watch to sound every 3–5 min. At the end of each 3–5-min interval, teachers scanned the classroom, praised students for demonstrating target behaviors, and recorded points for teams. Points were tallied on a chart drawn on the board by the teachers. If teams were not demonstrating target behaviors, the teacher would use reminders such as “Remember to demonstrate respect by following directions Team Four.” Both teachers also awarded bonus points to teams or individual students for demonstrating the target skill (e.g., “Team One earned a bonus point for showing respect by working together.”). Across each session, teams had approximately 16 opportunities to earn points. Immediately after game play, teachers identified the teams who met criteria by tallying points and provided those teams with access to the reward (e.g., phone time, snacks, or point toward class party).

Results

Impact of CW-FIT on High School Student Behavior

Whole Class On-Task

Group on-task data are graphically displayed in Fig. 2. Visual analysis was conducted on level, trend, and variability within phases. Immediacy of effect was assessed along with percentage of non-overlap (PND) from baseline to intervention phase 1 and withdraws to intervention phase 2. PND are depicted in Table 1. Average group on-task behavior during baseline was 55.2% (range = 46–63%) with a slight upward trend and low variability. Whole group on-task behavior increased to 86.6% (range = 80–94%) upon initiation of the intervention showing an immediate treatment effect with a stable trend and low variability. Upon withdrawal, group on-task decreased to 66.7% (range = 64–70%) with a stable trend and low variability. With reintroduction of CW-FIT, whole group on-task increased to an average of 86.5% (range = 81–80%) with an immediacy of effect, stable trend, and low variability. During follow-up sessions, whole class on-task behavior averaged 88.5% (range = 88–89%) indicating increased levels of on-task behavior maintained six weeks and 20 weeks after the end of the second intervention phase. PND between
baseline to intervention 1 and withdraw to intervention 2 indicate a strong treatment effect (Table 1). The data suggest a functional relation between high school student on-task behavior and CW-FIT.

**Target Student On-Task**

To explore the impact of CW-FIT on behavior of students at-risk for internalizing and externalizing behavior problems, data were collected across phases for three target students. Data representing on-task levels of Target Students 1, 2, and 3, are presented in Fig. 3. Visual analyses of target student on-task data did not determine condition changes but instead were used to explore the presence of a functional relation by analyzing data for improvement in behavior of students nominated by teachers for demonstrating low levels of on-task. As shown in Fig. 3, Target Student 1 baseline levels of on-task behavior averaged 64.2%
Target Student 1 on-task behavior increased to 89.3% (range = 85–93%) with a stable trend and moderate variability. With CW-FIT, Target Student 1 average on-task behavior increased 87% (range = 82–97%) with an upward trend and low variability. Upon withdrawal, Target Student 1 on-task behavior averaged 87% (range = 82–97%) with a stable trend and low variability. Upon reintroduction of the intervention, Target Student 1 on-task behavior averaged 87.5% (range = 83–92%) with a stable trend and low variability. At follow-up, Target Student 1 on-task behavior averaged 84%; however, only one data point was collected limiting analysis. PND suggest a strong treatment effect from baseline to intervention 1 with no evidence of effect from withdrawal to phase 2 (Table 1). It appears Target Student 1 had carry-over effects from intervention 1 to the withdrawal phase.

Target Student 2 on-task behavior averaged 62% (range = 50–67%) with a decreasing trend and moderate variability. With implementation of CW-FIT, Target Student 2 average on-task behavior increased to 83.7% (range = 73–97%) showing an immediacy of effect with an increasing trend and low variability. Upon withdrawal, Target Student 2 on-task behavior averaged 69.7% (range = 53–82%) with a slight decreasing trend and moderate variability. With reintroduction of the intervention, Target Student 2 on-task behavior increased to an average of 93.3% (range = 92–95%) with a stable trend and low variability. Target Student 2 average on-task behavior at maintenance was 80.5% (range = 68–93%) indicating on-task behavior maintained at intervention levels six weeks and 20 weeks after the end of the second intervention phase. However, maintenance data showed variability with a decreasing trend. PND suggest a strong treatment effect between phases (Table 1).
At baseline as depicted in Fig. 3, Target Student 3 on-task behavior averaged 63.8% (range = 44–85%) with an increasing trend and high variability. With introduction of CW-FIT, Target Student 3 average on-task behavior increased to 93.3% (range = 82–98%) showing an immediacy of effect with an increasing trend and low variability. Upon withdrawal, Target Student 3 on-task behavior averaged 78.7% (range = 75–82%) with a stable trend and low variability. With reintroduction of CW-FIT, Target Student 3 on-task behavior increased to an average of 93% (range = 85–97%) with a stable trend and low variability. During maintenance, Target Student 3 average on-task behavior was 97.5% (range = 97–98%) indicating on-task behavior was maintained at intervention levels six weeks and 20 weeks after the end of the second intervention phase. PND suggest a strong treatment effect between phases (Table 1).

**Impact of CW-FIT on High School Teacher Behavior**

To explore the effect of CW-FIT on teacher behavior, data on frequency of praise and reprimands were collected. Frequency of Teacher 1 and Teacher 2 praise and reprimand statements is graphically displayed in Figs. 4 and 5. In addition to individual teacher data, collective teacher praise and reprimand statements are displayed.

**Teacher Praise Statements**

As depicted in Fig. 4, at baseline Teacher 1’s praise statements averaged 0.0 (range = 0–0) with a stable trend and no variability. With implementation of CW-FIT, Teacher 1’s praise behavior statements increased to an average of 14.4 (range = 9–22) with a slight decreasing trend and moderate variability. Data showed an immediacy of effect. Upon withdrawal of the intervention, Teacher 1 praise statements averaged 3.3 (range = 3–4) with no trend and low variability. With reintroduction of CW-FIT, Teacher 1 praise statements increased to an average of 6.8 (range = 5–9) showing an immediacy of effect with no trend and low variability. At follow-up, Teacher 1’s praise statements averaged 4.5 (range = 4–5) indicating praise occurred above baseline levels yet below intervention levels six weeks and 20 weeks after the end of the second intervention phase.
after the second intervention phase. PND indicated a strong treatment effect from baseline to intervention phase 1 and withdrawal to intervention phase 2 (Table 1).

At baseline, Teacher 2’s praise statements averaged 0.0 (range = 0–0) with a stable trend and no variability. With implementation of CW-FIT, Teacher 2 praise statements averaged 4 (range = 1–9) and showed a decreasing trend and high variability. Upon withdrawal, Teacher 2’s praise averaged 0.3 (range = 0–1) with a decreasing trend and low variability. With CW-FIT reintroduction, Teacher 2 praise statements averaged 3.25 (range = 1–6) and showed a decreasing trend and low variability. During follow-up, Teacher 2’s praise statements averaged 6 (range = 5–7) indicating praise was maintained at intervention levels six weeks and 20 weeks after the second intervention phase. PND suggest a strong treatment effect from baseline to intervention 1 and a moderate treatment effect from withdrawal to intervention phase 2 (Table 1).

As shown in Fig. 4 at baseline, collective teacher praise statements averaged 0.0 (range = 0–0) with a stable trend and no variability. With implementation of CW-FIT, collective teachers’ praise statements averaged 18 (range = 12–31) with a decreasing trend and moderate variability. Upon withdrawal of the intervention, collective teacher statements averaged 3.6 (range = 3–5) with a stable trend and low variability. With reintroduction of CW-FIT, collective teacher praise statements increased to an average of 10 (range = 8–13) with a slight decreasing trend and low variability. During maintenance, collective teachers praise statements averaged 10.5 (range = 9–12) indicating praise statements maintained at intervention levels six weeks and 20 weeks after the end of the second intervention phase.

**Teacher Reprimand Statements**

As displayed in Fig. 5 at baseline Teacher 1’s reprimand statements averaged 0.4 (range = 0–2) with a slight decreasing trend and low variability. With implementation of CW-FIT, Teacher 1 reprimand statements increased to an average of 2.6 (range = 0–5) with a stable trend and moderate variability. Upon withdrawal of the
intervention, reprimand statements averaged 2.3 (range = 0–5) with an increasing trend and low variability. With reintroduction of CW-FIT, Teacher 1’s reprimand statements decreased to an average of 1 (range = 0–3) with a slight decreasing trend and low variability. During follow-up, Teacher 1’s reprimand behavior averaged 1.5 (range = 1–2). PND suggest no treatment effect between phases (Table 1).

At baseline, Teacher 2’s reprimand statements averaged 4.4 (range = 1–10) with a decreasing trend and moderate variability. Upon implementation of CW-FIT, Teacher 2 reprimand behavior averaged 2.1 (range = 0–4) with an increasing trend and high variability. Upon withdrawal, Teacher 2 average reprimand behavior was 1.7 (range = 0–5) with an increasing trend and moderate variability. With reintroduction of CW-FIT, Teacher 2 reprimand statements averaged 1.5 (Range = 0–3) with an increasing trend and moderate variability. At follow-up, Teacher 2’s reprimand behavior averaged 2 (range = 1–3). PND indicate no treatment effect between phases (Table 1).

Collective teacher reprimand statements at baseline averaged 4.8 (range = 1–10) with a decreasing trend and moderate variability. Upon implementation of CW-FIT, collective teacher reprimand statements averaged 4.7 (range = 3–8) with an increasing trend and moderate variability. Upon withdrawal of the intervention, collective teacher reprimand statements averaged 4 (range = 0–7) with an increasing trend and moderate variability. With reintroduction of CW-FIT, collective teacher reprimand statements decreased to 2.5 (range = 0–4) with an increasing trend and moderate variability. At follow-up, collective teacher reprimand statements averaged 3.5 (range = 2–5) with an increasing trend and low variability.

### High School Teacher Procedural Fidelity

Procedural fidelity data were collected for 100% of sessions across both intervention phases. With phase one of intervention, procedural fidelity levels averaged 94% (range = 74–100%). At phase two of intervention, procedural fidelity levels averaged 92% (range = 81–100%). Across both phases of implementation, teachers averaged high levels of fidelity defined as 85% or higher (Wills et al. 2019). Interobserver agreement for procedural fidelity was collected 42% of sessions in intervention phase one and 50% of sessions intervention phase two. Across both phases IOA for procedural fidelity averaged 98% (range = 96–100%).

Additionally, fidelity data were collected for 66% of withdrawal phase sessions. At withdrawal of the intervention, procedural fidelity levels averaged 38.5% (range = 11–33%). During each session of withdrawal, the skill poster was prominently displayed. Additionally, during session one of the withdrawal phase, the teachers maintained a 4:1 praise to reprimand ratio. At follow-up, teachers maintained high levels of procedural fidelity averaging 96.5% (range = 93–100%).

### Direct Consumer Satisfaction

Social validity surveys and maintenance data were collected to assess direct consumer satisfaction with the intervention. After the final observation session
associated with the reintroduction of CW-FIT, teachers and students completed social validity surveys. The teachers rated the intervention very favorably indicating agree (5) and strongly agree (6) across all IRP-15 statements. The teachers found the intervention to be acceptable, appropriate for other teachers and a variety of students, and reasonable. Further, the teachers indicated satisfaction with procedures and found the intervention to be a fair way to handle behavior. Similar to the teachers, the students rated the intervention very favorably on the C-IRP. All students (N=13) indicated CW-FIT was a fair way to deal with behavior and agreed they liked CW-FIT. Additionally, most students (N=12) believed the game would be good to use with other kids. Finally, all students indicated disagreement with the statements, “My teacher was too harsh,” and “CW-FIT may cause problems with my friends.” As a final measure of consumer satisfaction, maintenance data were collected at six and 20 weeks after the final intervention session. The teachers continued to implement CW-FIT at both maintenance checks. Continued implementation is a strong indicator of consumer satisfaction.

Discussion

This study evaluated the impact of CW-FIT on student and teacher behavior in a co-taught ninth-grade English class. Consistent with previous evaluations of CW-FIT in elementary and middle school settings (Caldarella et al. 2015; Kamps et al. 2015; Orr et al. 2020; Wills et al. 2018, 2019) the results of the current study suggest CW-FIT outcomes may generalize to high school settings. The results of this study are discussed across the four research questions guiding the inquiry.

First, related to improvement in student behavior, whole group on-task behavior increased with implementation of CW-FIT. Although data for whole group on-task showed a slight increasing trend prior to implementation of the intervention, behavior improved an average of 30% from baseline to phase one of intervention. Such improvement in behavior is similar to findings in previous evaluations of CW-FIT (Kamps et al. 2015; Wills et al. 2018, 2019). Withdrawal of the intervention was associated with an average decrease of 20% in on-task behavior. With reimplementation of CW-FIT, behavior showed an average of 20% improvement. Further, maintenance data demonstrated improvement in whole group on-task behavior sustained well beyond phase two of the intervention. Such results provide strong evidence for a functional relation between CW-FIT and improvement in high school student behavior. Concerning exploration of the improvement in target student behaviors, the results suggest a functional relation for Target 2 with less compelling evidence for Target 1 and 3. Target 2 on-task behavior showed a decreasing trend prior to implementation of the intervention and improvement averaged 20% from baseline to treatment. Upon removal of the intervention, on-task behavior decreased an average of 14% yet variability was shown during the withdrawal phase. With reintroduction of the intervention, on-task behavior again improved approximately 25% with such improvement maintaining during follow-up observation sessions. Target 1 showed improvement averaging 25% from baseline to treatment; it should be noted prior to implementation of the intervention baseline data showed an increasing trend. Importantly, upon withdrawal of the intervention, on-task
behavior levels sustained. Absence of change in on-task behavior upon removal of the intervention may have been related to carry-over effects. During baseline, Target 3 on-task behavior displayed variability and an increasing trend. Yet, introduction of the intervention led to the highest level of improvement across target students, increasing from an average of 63.8% to 93.3%. Removal of CW-FIT was associated with decreased levels of on-task behavior that again improved from an average of 78.7% to 93% during the final phase of treatment. Average improvement in on-task behavior sustained for all three target students across both follow-up sessions.

Second, related to the impact of CW-FIT on teacher behavior, the intervention appeared to have the most notable impact on the Teacher 1’s frequency of praise. For both teachers, praise did not occur at any point during baseline. Upon implementation of the intervention, praise increased significantly for Teacher 1 averaging 14.4 statements per session. Less marked improvement in praise was shown for Teacher 2. Withdrawal of the intervention led to a decrease in praise for both Teacher 1 and Teacher 2, but again the decrease was less substantial for the Teacher 2. With reintroduction of the CW-FIT, Teacher 1 again showed higher frequency of praise, while praise levels remained consistently low for Teacher 2. For reprimands, there was minimal effect observed for Teacher 1 and Teacher 2 as average reprimand frequency levels persisted across phases. Despite limited evidence of significant change in individual teacher behavior, collectively, the teachers maintained the target 4:1 praise to reprimand ratio during both phases of treatment with a slightly lower ratio of 3:1 shown during both follow-up observations. Although no change to reprimand statements was observed, the increase in praise demonstrated by Teacher 1 was consistent with results from other evaluations of CW-FIT’s impact on teacher behavior (Caldarella et al. 2019; Wills et al. 2019).

Third, regarding implementation fidelity, the teachers maintained high levels across both phases of intervention. Phase 1 fidelity averaged 94% (range = 74–100%) with Phase 2 averaging 92% (range = 81–100%). Further, fidelity levels maintained at follow-up, averaging 96.5% (range = 93–100%). These findings are consistent with previous evaluations of CW-FIT showing teachers are able to implement the intervention with high levels of fidelity.

Finally, social validity results indicated teachers and students found the procedure favorable. Through their ratings on the 6-point Likert scale, teachers expressed satisfaction with the procedures and outcomes on the intervention. Teachers also indicated they would likely continue using the intervention. Similarly, student ratings on the dichotomous survey showed high satisfaction with the intervention. All students indicated they liked CW-FIT and thought it was a fair way to deal with behavior. Importantly, the intervention showed sustainability in this context as the teachers continued implementation of CW-FIT beyond the treatment window and have reported continued use of the intervention in the subsequent school year.

**Practical Implications**

Given the prevalence of problem behavior in secondary contexts and overuse of punitive measures in responding to such behavior (Flannery et al. 2014), identification of
feasible and effective strategies for teachers to use in high school settings is essential. Evidence from this study suggests CW-FIT may lead to improvement in student behavior in high school contexts. Further, similar to previous evaluations of CW-FIT (Caldarella et al. 2015; Kamps et al. 2015; Wills et al. 2018, 2019), teachers were able to maintain high levels of treatment fidelity and found favor with the intervention. Most importantly, the intervention showed sustainability as the teachers have continued use of CW-FIT well beyond the phases of the study.

Limitations and Future Research

Despite improvement in whole group on-task behavior, caution is warranted in interpretation of the results as there are limitations to these findings. First, related to the sample, the current study only included one classroom with two teachers. Thus, generalizability of the impact is not clear given the small sample size. Future research should expand evaluation of CW-FIT across other high school settings including additional classrooms and more students. Additionally, improvement in on-task behavior was only evidenced through whole group on-task data with some evidence of a functional relation for Target 2. Less of an effect was observed with the other two target students given variability of data with Target 3 and potential carry-over effects for Target 1. Further, target students were identified solely by teacher ranking as direct observations were not conducted to confirm low levels of on-task behavior. Finally, IOA was low for both teacher praise and reprimands, thus limiting the reliability of the data presented.

In addition, limitations exist with the research design and data collection procedures. Related to the withdrawal design used to evaluate experimental effect, limitations exist with this design in that return to baseline levels may not be possible given changes in the behavior of participants. This certainly appeared to be the case for Target Student 1 as minimal change was noted upon withdrawal of the intervention. Also, during the withdrawal phase components of the intervention sustained as evidenced by the fidelity percentages reported limiting assessment for experimental effect. As an additional limitation, per Kratochwill and Levin (2010), at least five data points per study phase are required to meet standards with three data points needed to meet standards with reservation. Although baseline and intervention phase 1 had sufficient number of data points, the withdrawal and intervention phase 2 had fewer data points, thus weakening the evidence of a functional relation. Future investigations should consider alternative research designs such as multiple baselines and ensure collection of at least five data points per study phase. Finally, the 30-s momentary time-sampling procedure to record on-task behavior may not have been sensitive to changes in on-task behavior across all participants.

Presenting an additional limitation, activities were not consistent across observation sessions. Although data were collected the same time each observation session, as is typical in classroom settings, the activities varied. As an example, the final baseline data points were collected, while students were completing an exam for the majority of the class period. Such contextual factors likely impacted on-task
behavior levels. Future evaluations of CW-FIT should more clearly define sessions in which data should be collected.

Finally, an additional limitation is related to intervention procedures. As is required with CW-FIT, a timer was set to sound every 3–5 min throughout game play. For this evaluation, the Teacher 1 held the timer and was first to give praise. This may have impacted the Teacher 2’s praise frequency as she may have thought she did not need to praise given Teacher 1 had just praised the students. Future evaluations should consider alternative methods for training co-taught teachers on implementation or keeping time to encourage both teachers to maintain high levels of praise.

Conclusion

The utility of CW-FIT across elementary settings has been well-established in the literature base (Caldarella et al. 2015; Kamps et al. 2015; Wills et al. 2018), and evidence is emerging for intervention efficacy in middle school classrooms (Caldarella et al. 2019; Orr et al. 2020; Speight et al. 2020; Wills et al. 2019). Though the results should be interpreted cautiously given the limitations, this study provides evidence CW-FIT may lead to similar improvements in high school student and high school teacher behavior. Fidelity measures also indicated high school teachers can efficiently integrate CW-FIT into their classrooms with relatively little training. Further, social validity surveys and follow-up sessions to assess sustainability showed high levels of consumer satisfaction. The findings suggest generalizability of CW-FIT to high school contexts.

Compliance with Ethical Standards

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

Informed Consent Informed consent was obtained from all individual participants included in the study.

References

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. https://doi.org/10.1007/s10643-014-0664-3.

Caldarella, P., Wills, H., Anderson, D., & Williams, L. (2019). Managing student behavior in the middle grades using class-wide function-related intervention teams. *Research in Middle Level Education, 42*, 1–15. https://doi.org/10.1080/19404476.2019.1654799.

Conklin, C., Kamps, D., & Wills, H. (2017). The effects of class-wide function-related intervention teams (CW-FIT) on students’ prosocial behaviors. *Journal of Behavioral Education, 26*, 75–100. https://doi.org/10.1007/s10864-016-9252-5.

Donaldson, J., Wiskow, K., & Soto, P. (2015). Immediate and distal effects of the good behavior game. *Journal of Applied Behavior Analysis, 48*, 685–689. https://doi.org/10.1002/jaba.229.
Finn, C. A., & Sladeczek, I. E. (2001). Assessing the social validity of behavioral interventions: A review of treatment acceptability measures. School Psychology Quarterly, 16, 176–206. https://doi.org/10.1521/spec.16.2.176.18703.

Flannery, K. B., Fenning, P., Kato, M. M., & McIntosh, K. (2014). Effects of school-wide positive behavioral interventions and supports and fidelity of implementation on problem behavior in high schools. School Psychology Quarterly: The Official Journal of the Division of School Psychology, American Psychological Association, 29, 111–124. https://doi.org/10.1037/spq0000039.

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. https://doi.org/10.1177/0145445513507574.

Friend, M., Cook, L., Hurley-Chamberlain, D., & Shamberger, C. (2010). Co-teaching: An illustration of the complexity of collaboration in special education. Journal of Educational & Psychological Consultation, 20, 9–27. https://doi.org/10.1080/10474410903535380.

Goves, E., & Austin, J. (2017). An evaluation of interdependent and independent group contingencies during the good behavior game. Journal of Applied Behavior Analysis, 50, 552–566. https://doi.org/10.1002/jaba.393.

Kamps, D., Wills, H., Bannister, H., Heitzman-Powell, L., Kottwitz, E., & Fleiming, K. (2015). Class-wide function-related intervention teams “CW-FIT” efficacy trial outcomes. Journal of Positive Behavior Intervention, 17, 134–145. https://doi.org/10.1177/1098300714565244.

Kleinman, K. E., & Saigh, P. A. (2011). The effects of the good behavior game on the conduct of regular education New York city high school students. Behavior Modification, 35, 95–105. https://doi.org/10.1177/0145445510392213.

Kratochwill, T. (1985). Advances in school psychology (Vol. IV). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Kratochwill, T. R., & Levin, J. R. (2010). Enhancing the scientific credibility of single-case intervention research: Randomization to the rescue. Psychological Methods, 15, 124–144. https://doi.org/10.1037/a0017736.

Litoe, L., & Pumroy, D. K. (1975). A brief review of classroom group-oriented contingencies. Journal of applied behavior analysis, 8, 341–347. https://doi.org/10.1901/jaba.1975.8-341.

Maggin, D., Pustejovsky, J., & Johnson, A. (2017). A meta-analysis of school-based group contingency interventions for students with challenging behavior: An update. Journal of Remedial and Special Education, 38, 353–370. https://doi.org/10.1177/0741932517716900.

Martens, B., & Witt, J. (1982). The intervention rating profile. University of Nebraska-Lincoln: Lincoln.

Martens, B. K., Witt, J. C., Elliott, S. N., & Darveaux, D. X. (1985). Teacher judgments concerning the acceptability of school-based interventions. Professional Psychology: Research and Practice, 16, 191–198. https://doi.org/10.1037/0735-7028.16.2.191.

Mitchell, R. R., Tingstrom, D. H., Dufrene, B. A., Ford, W. B., & Sterling, H. E. (2015). The effects of the good behavior game with general-education high school students. School Psychology Review, 44, 191–207. https://doi.org/10.17105/spr-14-0063.1.

Moore, T., Wehby, J., Oliver, R., Chow, J., Gordon, J., & Mahany, L. (2017). Teachers’ reported knowledge and implementation of research-based classroom and behavior management strategies. Remedial and Special Education, 38, 222–232. https://doi.org/10.1177/0741932516683631.

Orr, R., Caldarella, P., Hansen, B., & Wills, H. (2020). Managing student behavior in a middle school special education classroom using CW-FIT tier 1. Journal of Behavioral Education, 29, 168–187. https://doi.org/10.1080/10864-019-09325-w.

Simonsen, B., Fairbanks, S., Briesch, A., Myers, D., & Sugai, G. (2008). Evidence-based practices in classroom management: Considerations for research to practice. Education and Treatment of Children, 31, 351–380. https://doi.org/10.1353/etc.0.0007.

Speight, R., Whitby, P., & Kucharczyk, S. (2020). Impact of CW-FIT on student and teacher behavior in a middle school. Positive Behavior Interventions, 00, 1–15. https://doi.org/10.1177/109830720910133.

Stratton, K., Gadke, D., & Morton, R. (2019). Using the good behavior game with high school special education students: Comparing student- and teacher-selected reinforcers. Journal of Applied School Psychology, 35, 105–121. https://doi.org/10.1080/15377903.2018.1509920.
Walker, H., & Severson, H. (1992). Systematic screening for behavior disorders: SSBD. Longmont, CO: Sopris West.

Wills, H., Caldarella, P., Mason, B., Lappin, A., & Anderson, D. (2019). Improving student behavior in middle schools: Results of a classroom management intervention. Journal of Positive Behavior Interventions, 21, 213–227. https://doi.org/10.1177/1098300719857185.

Wills, H., Wehby, J., Caldarella, P., Kamps, D., & Swinburne Romine, R. (2018). Classroom management that works: A replication trial of the CW-FIT program. Exceptional Children, 84, 437–456. https://doi.org/10.1177/0014402918771321.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.