Hammill Institute on Disabilities

Journal of Positive Behavior Interventions
2015, Vol. 17(3) 134–145

DOI: 10.1177/1098300714565244
jpbi.sagepub.com
SAGE

Class-Wide Function-Related Intervention Teams "CW-FIT" Efficacy Trial Outcomes

Debra Kamps, PhD[1], Howard Wills, PhD[1], Harriett Dawson-Bannister, MS[1], Linda Heitzman-Powell, PhD[2], Esther Kottwitz, MS[3], Blake Hansen, PhD[4], and Kandace Fleming, PhD[5]

Abstract

The purpose of the study was to determine the efficacy of the Class-Wide Function-Related Intervention Teams (CW-FIT) program for improving students' on-task behavior, and increasing teacher recognition of appropriate behavior. The intervention is a group contingency classroom management program consisting of teaching and reinforcing appropriate behaviors (i.e., getting the teacher's attention, following directions, and ignoring inappropriate behaviors of peers). Seventeen elementary schools, the majority in urban and culturally diverse communities, participated in a randomized trial with 86 teachers (classrooms) assigned to CW-FIT, and 73 teachers (classrooms) assigned to the comparison group. Class-wide student on-task behavior improved over baseline levels in the intervention classes. Teachers were able to implement the intervention with high fidelity overall, as observed in adherence to 96% of the fidelity criteria on average. Teacher praise and attention to appropriate behaviors increased, and reprimands decreased. These effects were replicated in new classrooms each of the 4 years of the study, and for all years combined.

Keywords

classroom, intervention(s), group designs, data analysis, studies

Indicators of effective schools include a positive school climate, high expectations for learning, and well-trained teachers who manage their classrooms to support academic and social development for all students. One persistent challenge to providing effective instruction is managing classroom behaviors for increasingly diverse groups of children in today's schools (Chafouleas, Volpe, Gresham, & Cook, 2010; Stage & Quiroz, 1997). Epstein, Atkins, Cullinan, Kutash, and Weaver (2008) reported that discipline and managing behavior in schools is consistently cited as one of the top four concerns in public education; and that studies over the past 30 years indicate that at any given time, approximately 20% of children are at risk for behavior problems, and 10% of children may have mental illness (about 5 million children). Educators continue to rank disruptive behaviors and conduct problems in the classroom as an ongoing barrier to teaching their students (Simonsen, Fairbanks, Briesch, Myers, & Sugai, 2008). In addition, many teachers report receiving little training in behavioral interventions to address challenging classroom behaviors (Reinke, Stormont, Herman, Puri, & Goel, 2011).

Fortunately, several evidenced-based interventions are specifically designed to be implemented with groups of children in classrooms and school settings to manage behaviors (Maggin, Chafouleas, Goddard, & Johnson, 2011; Stage & Quiroz, 1997). Pertinent to the current study are group contingency interventions that include teaching pro-social skills and classroom rules, and differential reinforcement of expected behaviors (Maggin et al., 2011; Simonsen et al., 2008; Stage & Quiroz, 1997). Reviews showing positive outcomes from the use of group contingencies have been published beginning in the 1970s, with continuing reviews supporting their relevance in schools (see reviews by Embry, 2002; Litow & Pumroy, 1975; Theodore, Bray,

[1]University of Kansas, Kansas City, USA
[2]University of Kansas Medical Center, Kansas City, USA
[3]Kansas City, Kansas Public Schools, USA
[4]Brigham Young University, Provo, UT, USA
[5]University of Kansas, Lawrence, USA

Corresponding Author:
Debra Kamps, Senior Scientist, Life Span Institute, Juniper Gardens Children's Project, University of Kansas, 444 Minnesota Avenue, 3rd floor, Kansas City, KS 66101, USA.
Email: dkamps@ku.edu

Action Editor: Joshua Harrower

Kehle, & DioGuardi, 2004; Tingstrom, Sterling-Turner, & Wilczynski, 2006; Maggin, Johnson, Chafouleas, Ruberto, & Berggren, 2012). Group contingency programs refer to behavioral classroom interventions where one or several specified contingencies are applied to the same behavior for all students or groups of students within a classroom (Cooper, Heron, & Heward, 2007). Research from many group contingency programs provide evidence for the combined use of teaching behavioral and social rules, and reinforcement of appropriate behavior to improve student performance (Lea, Bray, Kehle, & DioGuardi, 2004; Leflot, vanLier, Onghena, & Colpin, 2013; Skinner, Cashwell, & Skinner, 2000; Thorne & Kamps, 2008).

Class-Wide Function-Related Intervention Teams (CW-FIT) Group Contingency

CW-FIT is a classroom management system based on teaching classroom rules/skills and use of a group contingency plan with differential reinforcement of appropriate behaviors, and minimized social attention to inappropriate behavior. Two studies have demonstrated that the CW-FIT group contingency program is beneficial for improving class-wide student behavior (Kamps et al., 2011; Wills, Kamps, Hansen, et al., 2010). The Wills article details implementation procedures and general outcomes for three urban elementary schools with economically disadvantaged and diverse student bodies. Findings indicated that on-task behaviors improved to 80% or higher during observations for 16 classrooms using CW-FIT. In addition, observations before and during the CW-FIT intervention were conducted for a sample of 25 students who were at risk for emotional/behavioral disorders. These students showed a nearly 50% reduction of disruptive behaviors during the intervention conditions.

In another recent study, the CW-FIT intervention was found to be effective in increasing on-task behaviors with more than 100 students in six elementary school classes (Kamps et al., 2011). Data were collected for 8 students at risk for emotional and behavioral disorders (EBD) in three of the six participating classes, with decreases in disruptive behaviors for all of the students during CW-FIT conditions. Teacher praise increased and reprimands (attention to inappropriate behaviors) decreased during intervention. Findings from these initial CW-FIT intervention studies and others support the use of group contingencies to improve classroom behaviors (Kamps et al., 2011; Mitchem, Young, West, & Benyo, 2001; Theodore et al., 2004; Wills, Kamps, Hansen, et al., 2010). Currently, studies are needed to demonstrate the efficacy of behavioral interventions in classroom management practices such as CW-FIT with large groups of students in randomized trials (Maggin et al., 2012).

While there is a long history of group contingency research, the vast majority has been conducted using single case designs. A recent systematic review of group contingencies using What Works Clearinghouse procedures (WWC) included 27 single case design studies (Maggin et al., 2012). Their findings indicated "sufficient rigor, evidence and replication to label the intervention as evidence-based" (Maggin et al., 2012, p. 625). The authors recommended additional research to provide clearer descriptions of students, to determine which students are best suited for the intervention, and to measure the specific contingency procedures and fidelity. They further emphasized the need for large, randomized trials of group contingencies. The authors reported that their search yielded 24 investigations of large-n design studies, but upon investigation, they found "a considerable range of quality, rigor, and focus" (Maggin et al., 2012, p. 628).

In summary, this brief review suggests the need for more large-scale studies of effective classroom management interventions including group contingencies that (a) have procedural components supported by research, (b) have evidence that they can be implemented with fidelity, (c) focus on teacher attention to appropriate behaviors, and (d) are acceptable to teachers. Interventions that meet these criteria can provide a mechanism for increasing effective classroom management practices, and improved student performance.

Purpose

This purpose of the study was to conduct a randomized trial for the CW-FIT group contingency intervention in a larger number of schools, and with more teachers and students than were involved in prior CW-FIT studies. Large group studies are needed to demonstrate more generalizable findings for group contingency interventions, including CW-FIT. In addition, the purpose was to demonstrate effects of CW-FIT in classrooms not concurrently implementing a School-Wide Positive Behavior Support intervention (SWPBS). Research questions for the study were as follows:

Research Question 1: What is the effect of the CW-FIT group contingency program on class-wide on-task behavior, and are these outcomes superior to comparison classes?

Research Question 2: What is the effect of the CW-FIT group contingency program on teacher attention to appropriate behaviors (praise/points) and reprimand frequencies, and does intervention teacher behavior differ from teacher behavior in comparison classes?

Method

Participants and Settings

Seventeen elementary schools, the majority in urban and culturally diverse communities, participated in the randomized trial of CW-FIT, 4 in Year 1, 5 in Year 2, 5 in Year 3, and 3 in Year 4. Each school participated in the study for 1 year. These schools were located in three districts. One was an urban district in a large city in the Midwest (12 schools), one was an adjacent district still considered part of the metropolitan area (3 schools), and the third was in a university community approximately 40 miles away (2 schools). School size averaged 382 students (range = 161–684); with a mean of 79% (range = 39%–97%) with free/reduced lunch status; and a mean of 65% (range = 36%–93%) minority status. District administrators referred schools to the research team each year. Researchers met with the principals to secure interest in participating. Principals spoke independently to their staff to confirm agreement to participate. All the schools comprised grades kindergarten through fifth. In 12 schools, the student body represented culturally diverse backgrounds, with less than half of the students describing themselves as White, non-Hispanic. Sixteen of the 17 schools served large numbers of children from low socioeconomic status families as indicated by numbers receiving free/reduced lunches (range = 63%–97%). Only one school appeared to be less economically disadvantaged (free/reduced: 39%). An average of eight teachers (classes) per school participated in the study each year.

A total of 159 teachers participated in the 4 years of study, 78 females and 8 males in the CW-FIT group, and 70 females and 3 males in the comparison group. Two additional teachers dropped out of the study because of concerns about using rewards for appropriate behaviors. Class sizes ranged from approximately 18 to 25 in both groups. Data, however, were not collected on class size. Given that both groups were assigned from each school, class size was not considered to confound the study. Prior to notification of their group assignment, each teacher in the study selected a time of the day with challenging student behaviors for study implementation. Teachers in the experimental group selected academic times as follows: 44 math (51%), 28 reading (33%), 10 writing (12%), 1 science (1%), and 3 other (3%). Teachers in the comparison groups selected the following: 34 math (47%), 23 reading (32%), 3 science (4%), 5 writing (7%), and 8 other (11%). The content areas were fairly equally distributed, with predominantly math and reading selected regardless of intervention group.

Building coaches served as CW-FIT trainers in the schools (see implementation). Coaches were district employees with salaries paid by a subcontract from the university through grant funding. Five persons, all females with varying levels of experience, served as coaches over the 4 years. Each school was assigned a 40% to 50% time coach. One coach served as the lead coach for the project. She had 26 years teaching and consulting experience in early childhood settings. The second coach served for 3 years and was a special education teacher with 12 years teaching experience working in the primary participating district. The third coach served for 3 years. This coach had 1 year of experience as a school social worker prior to working on the project. Two coaches served on the project for 1 year. One was an elementary school counselor with more than 20 years of experience. One was a school social worker with 3 years' experience. Once the grant funding ended, the coaches were no longer available to the schools, but the lead coach continued to liaise between the school personnel and the university to forward or answer any questions, or provide any intervention materials that were needed.

Measures and Classroom Observations

Group on-task data. Group on-task data were collected using a 30-s momentary time sample procedure. The data collection form contained a grid broken down by 30-s intervals up to 20 min across the page and number of teams/groups down the page. A silent, digital timer showing minutes and seconds was used to keep time. Every 30 s, the observer would scan the group and record a plus for each team (row or small group) of students if ALL students in the group were on-task. There was an average of four students per team with a range of anywhere between 2 and 5 students. Teams were rows or tables of students seated together during the activity, and team members were constant across sessions. Within each of the 30-s intervals, the observer would rotate from Team 1 (look/score), to Team 2 (look/score), to Team 3 (look/score), and so on using the same sequence until each team was scored, then begin the sequence again. If any one member of the team was off-task, the observer would score a minus in that box. All groups were scored sequentially during each 30-s time sample for every 20-min observation conducted. On-task was defined as follows: All students are appropriately working on the assigned/approved activity including (a) attending to the material/task, (b) making appropriate responses (writing, following rules of a game, looking at the teacher), (c) asking for assistance (where appropriate) in an acceptable manner (e.g., raising hand), and (d) waiting appropriately for the teacher to begin or continue with instruction (staying quiet and staying in seat). Group on-task data were collected for an average of 1-2 sessions per week per class during baseline (n=277) and intervention (n=975) for the experimental group; and for the comparison group 233 times during baseline in the fall and 420 during baseline 2 (winter to early spring).

Teacher behavior. Teacher praise statements, points, and reprimands were recorded on a frequency basis during the 20-min group on-task data session. Praise/attention to appropriate behavior was defined as follows: A verbal statement (e.g., "Nice work following directions!" "Team 1 is doing a great job staying in their seats!") or physical gesture of intended reinforcement (hugs, pats) or tangible rewards (tokens, points) that indicate approval of behavior. Delivery of points on the CW-FIT game chart was also recorded in the praise frequencies. Teachers were trained to give targeted praise statements specific to the behavior they wanted the students to repeat. Data, however, were not collected on specific versus general praise statements. Reprimands were defined as (a) verbal comments or negative statements about behavior with the intent to stop the student from misbehaving (e.g., "Everyone needs to get quiet!"), and (b) gestures used with the same intent as verbal reprimands. Tone was likely stern or punitive, although reprimands may have been delivered in a pleasant tone. Threats were also counted as reprimands.

Procedural fidelity. A 13-item procedural fidelity checklist was used to determine the use of CW-FIT intervention components during sessions (e.g., skills are prominently displayed on posters, pre-corrects on skills occur at beginning of session, point goal is determined, points are awarded to individuals/teams for use of the skills at set intervals, etc.). Each checklist item was scored as yes or no by the observers. The fidelity checklist probes were completed in conjunction with the group on-task data in both the baseline and intervention phases, and for both the experimental and comparison groups. Although the comparison teachers did not attend CW-FIT training, the checklist was administered for those teachers as well to measure any use of intervention components in the comparison classes. Fidelity data were collected 1 to 2 times per week concurrently with group on-task data or observations of individual students, for a total of 1,851 for experimental group and 975 for the comparison. Teachers in CW-FIT classrooms implemented the intervention with high fidelity averaging 92.4%. The use of procedures was low during baseline/non CW-FIT conditions for all groups (1%–2%).

Classroom management ratings. An eight-item checklist related to general classroom management (e.g., directions for class assignments are provided and clear, transitions are smooth with only minor disruptions, teacher ignores minor inappropriate behaviors, etc.) was completed during each observation (1–2 times per week). These items were rated 1 (*very low*) to 4 (*high*). The checklist was completed at the same time as the fidelity checklist in baseline and intervention and for the experimental and comparison groups. The measure, modified from the *Classroom Atmosphere Rating Scale* (Wehby, Dodge, & Greenberg, 1993), was not intended to be a comprehensive measure of classroom management but general procedures that influence on-task behaviors and could impact effectiveness of intervention. Prior studies showed good internal consistency, standard alpha coefficient of .94 to .95, and moderate interrater reliability–interclass correlation coefficient of .44 (n = 115; Barber, Maggin, & Wehby, 2009; Wehby et al., 1993).

Both groups averaged use of about 50% (of the total possible points) of good management practices in place during baseline. The comparison group remained at similar levels in Baseline 2 (52%), whereas the CW-FIT implementation improved 84%.

Observer training and interrater agreement. Observers were trained to collect on-task and fidelity/management ratings. Observers completed supervised classroom observations, and were tested for interobserver agreement by the project coordinator until they reached a minimum level of interrater agreement of 90% on two separate observations. During data collections periods, observers (one to two per class) would go into the classrooms at the designated "problem time" as specified by the teacher during training. Every participating classroom was observed by more than one observer to maximize the data collection opportunities in relation to observer availability. Over the course of the 4-year study, 2,014 group and individual observations were conducted in experimental group classes; and 1,143 group and individual observations were conducted in the comparison group classes.

Interrater agreement. Interrater agreement data were collected for on-task data, praise/points, reprimands, and fidelity data. A second observer was present during the same 20-min observation, during which the primary observer provided a low verbal cue to look and record each group's on-task score of + or – at each 30-s interval (i.e., "Team 1" . . . "Team 2" . . . "Team 3" . . .). Frequencies of praise and reprimands were recorded during the entire observation by each observer. Percent agreement was computed using a point by point system and dividing the number of observer agreements by the total number of agreements and disagreements, multiplied by 100. The number of interrater agreement checks ranged from 0 to 9 for each teacher, with interobserver coding for a total of 10% of data sessions. Interrater agreement for on-task behavior averaged 91.6% (range = 61.7%–100%), praise/points averaged 85.7% (range = 0%–100%), and reprimands averaged 85.2% (range = 0%–100%). The 0% agreement occurred very infrequently and was due to low frequencies (i.e., an occurrence of one by first observer and 0 by second observer would calculate to 0% agreement).

Interobserver agreement for fidelity was collected for 17% of the experimental group fidelity checks (n = 320, averaging 3.9 reliability sessions per teacher) and averaged

99% in baseline and 96% during intervention, and 99% in both baseline periods for the comparison group (n = 146, averaging 2.4 reliability sessions per teacher). Reliability on the coaches' ratings of fidelity constituted 24% of these fidelity checks and was similar, with a mean of 98% and a range of 83% to 100%. For another 22% of the fidelity checks the coaches were the reliability observer for the researcher, with a mean of 97.6%. Interrater agreement on the use of general classroom management averaged 87% in baseline and 92% during intervention for the experimental group and 86.4% and 86.1% in baseline periods for the comparison group.

Consumer satisfaction. Consumer satisfaction questionnaires were completed by 66 of the 86 CW-FIT teachers. Each questionnaire included seven questions regarding acceptability of the procedural components, ease of implementation, and perceptions regarding effectiveness. Questions were scored using a 4-point Likert-type scale, 1 = *high acceptability* to 4 = *lowest rating*. Two open-ended questions (i.e., what he or she liked, did not like, suggestions for improvement) were included. Students satisfaction surveys consisted of two yes/no questions: "Do you like playing the CW-FIT Game?" and "Do you think other kids should get to play the CW-FIT game in their classrooms?" as well as open-ended questions including "What do you like about CW-FIT?" and "Is there anything you don't like about CW-FIT?"

Procedures

Experimental conditions consisted of baseline for the experimental and comparison groups consisting of business as usual procedures, CW-FIT for the experimental group late fall to early spring, and Baseline 2 for the comparison group in winter and early spring (during the same time period as CW-FIT implementation for the experimental group).

Baseline. Baseline consisted of "business as usual" in the classroom period selected for intervention or for comparison purposes, and occurred over a 2- to 3-week period. The curriculum content, general instruction routines, and materials remained the same across all experimental phases. Teachers followed their usual classroom management procedures which commonly included posted classroom rules, reminders about the rules, and reprimands for infractions. Many teachers used a response cost warning system with colored cards in pocket folders for each student. Repeated rule infractions resulted in the students moving their cards to a different color with consequences for each card change (5 min from recess). The colored cards systems remained in place during intervention. No teachers were observed or reported use of token systems or group contingency programs during the selected period. Comparison teachers continued using the "business as usual" throughout the study. Observations were conducted but feedback was not provided on performance other than general comments "That was an interesting lesson." Or "The class did well today." Data were shared at the end of the study. If comparison group teachers expressed concerns, they were encouraged to report behaviors to the student assistance teams, following the usual school procedures.

CW-FIT. The CW-FIT intervention is a behavioral intervention designed to teach appropriate skills, and reinforce students' use of the skills by using a game format (group contingency). The CW-FIT intervention was implemented 3 to 4 times a week beginning in mid to late October and continuing through March of the same school year for participating teachers/classes. Teachers, however, opted to continue the intervention beyond the data collection period in March to complete the school year. The group contingency was designed to address attention (teacher and peer), and escape as commonly reported functions of problem behavior (Ervin et al., 2001). That is, the intervention procedures required teachers to attend to appropriate student behaviors frequently (every 2–5 min); and taught students to request attention or help with lessons using appropriate behavior (i.e., raising hands) and to ignore inappropriate peer behavior. Although common functions of inappropriate behaviors were incorporated into CW-FIT, it is not a function-based intervention per se, in that a functional analysis was not conducted prior to implementation.

The CW-FIT incorporates best practices for *teaching prosocial behaviors* as published in prior curricula and studies (e.g., Mitchem et al., 2001; *Tough Kid Social Skills*, Sheridan, 2010; *Utah's B.E.S.T. Project*, Reavis, Jenson, Kukic, & Morgan, 1988; *Skillstreaming*, McGinnis, 2010); and promoted in SWPBS (Horner & Sugai, 2005; Simonsen et al., 2008). Three target skills were taught in class-wide lessons during the initial 3 to 5 sessions: (a) gaining the teacher's attention, (b) following directions, and (c) ignoring inappropriate behaviors. In subsequent sessions, the teacher would (a) provide brief pre-corrects of skills at the start of the lesson and (b) provide incidental teaching of the skills throughout the lesson. The *group contingency component* of CW-FIT consisted of a game format with class teams of 2 to 5 students (typically rows of students) and the use of a token economy. During the CW-FIT intervention period, the teacher set a timer to beep every 2 to 3 min on a variable schedule. At the beep, the teacher would award a point on the team chart to each team with ALL members engaged in appropriate behaviors. At the end of the class period, rewards were given to all students on each team who met the stated goal. Teachers used tangible rewards (pencils, small tablets), and special activities as incentives (listening to music). Teachers were encouraged to provide

differential reinforcement in the form of frequent, specific praise for appropriate on-task behaviors and use of the skills when awarding team points, and to individuals and groups throughout the lesson. Teachers were encouraged to give minimal attention to inappropriate behaviors.

Training and implementation. Implementation of the intervention consisted of a 2-hr training workshop by project staff in the CW-FIT procedures, modeling of the procedures for 2 to 3 sessions, and weekly feedback from building coaches and researchers. Coaches attended training and provided modeling and feedback to teachers based on fidelity data, giving a verbal report on their use of praise, reprimands, CW-FIT procedures, and the class on-task data. Coaches also assisted in data collection and shared on-task data with teachers on a biweekly basis. Coaches met with researchers on a biweekly basis to review data and receive guidance on consulting with teachers.

Statistical Methods

Descriptive statistics (means, ranges, standard deviations) were reviewed to note intervention and comparison class differences, and to note differences across conditions. General linear mixed model (GLMM) analyses were used to examine differences on the three dependent variables between groups (treatment and control) and across intervention phases. All observation data points for the dependent variables were included in the analysis. Analyses were conducted to determine within- and between-group differences for the on-task class-wide data, and for praise and reprimands across teachers. Because the intervention was conducted at the teacher level, the teacher was the unit of analysis, with time on-task, praise, and reprimands being assessed at the classroom level. Because each teacher/classroom was assessed at multiple time points both prior to, and after the implementation of the intervention, multilevel modeling was used to accommodate the dependence due to repeated observations. SAS PROC MIXED with maximum likelihood estimation was used to analyze models in which on-task, praise, and reprimands were the product of phase (baseline or intervention) and condition (CW-FIT or control) and the interaction of the two. Both phase and condition were considered to be fixed, whereas teacher/classroom was considered random. The same model was run separately for each of the dependent variables.

Group Experimental Design and Randomization Procedures

A *randomized experimental control group design* was the primary design for the study. A block randomization process matched on grade levels within each school during each year of the study was used. The procedure included several steps. First, the researchers presented the study to the entire staff, requesting volunteers to participate. A minimum of eight teachers in each building was required to enroll the school in the study. Teachers were informed that they would not know the group placement (experimental or comparison), until after they agreed to participate. Second, volunteer teachers were sorted by grade level: K-2 and 3 to 5. Then teachers within each grade level were randomly assigned to participate in either the experimental or comparison groups (one drawing for the K-2 teachers and one drawing for the 3–5 teachers). In Year 1, 23 were assigned to the CW-FIT group, 21 to the comparison; in Year 2, 26 and 25; in Year 3, 23 and 21; and in Year 4, 14 and 6, respectively. Across the 4 years of the study, a total of 86 teachers participated in CW-FIT and 73 were in the comparison group.

Because schools only participated for 1 year, teachers could only be assigned to either the experimental or comparison group. All participants' data were combined across the 4 years to serve as one study, with one experimental group and one comparison group for analysis purposes. Because teachers within each school were randomly assigned to group, several procedures were used to prevent contamination of conditions. Teachers signed a consent form agreeing to not implement the intervention procedures if they were selected to be in the comparison group. In addition, fidelity was monitored in comparison and intervention classes to monitor use of procedures (see measures and results). Teachers in the comparison group were offered CW-FIT intervention training (the same training as provided to intervention teachers) in the spring of the school year following data collection.

Results

What Are Class-Wide Effects for On-Task Behavior?

Results are presented first for on-task behavior of the classes and then for teacher behaviors. Overall, the CW-FIT intervention classes showed higher increases in levels of on-task behavior over time than the comparison group classes. On-task data were higher during CW-FIT conditions during each of the 4 years of the study with average scores presented in Table 1. Data analysis indicated there is a significant phase by condition interaction $F(1, 1804) = 406.77$, $p < .0001$. Class-wide on-task behavior during CW-FIT increased from 51.95% in baseline to 82.99%; this was a significant increase $F(1,1765) = 1{,}600.12$, $p < .0001$ (see Table 2). The comparison group classes also increased their on-task behavior from 50.18% to 56.31% $F(1, 1827) = 40.73$, $p < .001$. While both class conditions saw increases in on-task behavior, the significant interaction indicates that the difference observed in the treatment classrooms was

Table 1. Means and Standard Deviations for On-Task, Teacher Praise, and Reprimands Across Groups by Year.

Behaviors/ Years	CW-FIT classes/teachers		Comparison classes/teachers	
	Baseline	CW-FIT	Baseline	Baseline 2
% On-task				
Year 1				
M	56.9	81.7	58.6	56.1
SD	16.1	11.2	19.0	14.84
Year 2				
M	50.8	82.4	50.2	57.7
SD	15.2	9.7	15.0	17.8
Year 3				
M	51.9	83.9	47.4	56.3
SD	12.8	8.8	14.0	16.3
Year 4				
M	49.5	77.9	36.6	44.8
SD	16.1	14.0	14.6	17.8
Praise/points frequency				
Year 1				
M	4.6	36.4	3.4	2.4
SD	4.6	17.2	4.4	3.1
Year 2				
M	4.1	41.2	5.0	6.7
SD	5.8	27.7	5.2	10.3
Year 3				
M	2.6	44.7	3.8	2.7
SD	3.5	24.1	6.0	3.7
Year 4				
M	5.5	37.0	2.8	1.8
SD	7.5	21.7	3.5	2.0
Reprimand frequency				
Year 1				
M	9.3	5.7	8.5	10.8
SD	9.2	5.6	7.3	9.7
Year 2				
M	8.9	6.2	9.1	8.8
SD	7.7	6.6	6.9	7.1
Year 3				
M	7.0	3.7	7.0	9.1
SD	6.4	4.2	5.4	6.9
Year 4				
M	8.6	5.4	11.3	8.6
SD	6.8	5.2	9.1	6.3

Note. CW-FIT = Class-Wide Function-Related Intervention Teams.

significantly different from the change observed in the comparison classrooms. In fact, it was much larger. Figure 1 depicts the averages overall for on-task behavior for the CW-FIT classes and the comparison group classes for the collapsed data across all 4 years.

What Are Effects for Teacher Behaviors?

For praise, there is a phase by condition interaction, $F(1, 1817) = 534.87$, $p < .001$. Within the treatment condition the change from baseline to intervention is greater as can be seen by the LSMEANS which indicated that CW-FIT classroom teachers went from 4 to 40 praises/points (see Figure 2). This was a significant change ($p < .001$), and much greater than the change seen in the comparison classrooms. For the comparison group teachers, the change was small from 4.46 to 4.62 (see Table 2). For reprimands, there is also a significant interaction, $F(1, 1796) = 53.48$, $p < .001$. The treatment classroom teachers decreased reprimands from 7.48 in baseline to 4.45 during CW-FIT

Table 2. Class-Wide On-Task Behavior, Praise, and Reprimands.

Phase	Condition	Estimate	*SE*
On-task behavior			
Baseline	CW-FIT	51.946	1.0984
Baseline	Comp	50.180	1.1941
Intervention	CW-FIT	82.986	0.9361
Baseline 2	Comp	56.311	1.1063
Praise			
Baseline	CW-FIT	3.9986	1.3370
Baseline	Comp	4.4653	1.4498
Intervention	CW-FIT	40.032	1.1272
Baseline 2	Comp	4.6172	1.3358
Reprimands			
Baseline	CW-FIT	7.4829	0.5318
Baseline	Comp	8.4175	0.5761
Intervention	CW-FIT	4.4463	0.4645
Baseline 2	Comp	9.4858	0.5394

Note. CW-FIT = Class-Wide Function-Related Intervention Teams.

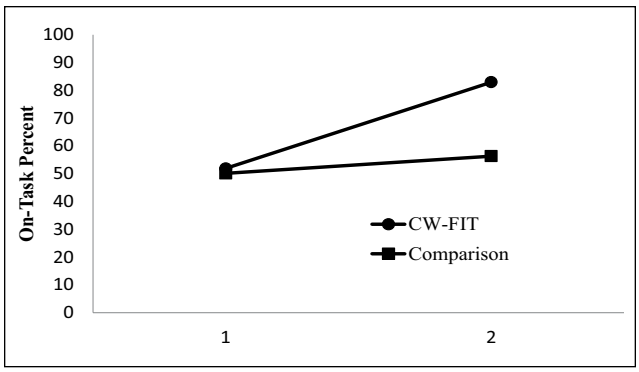

Figure 1. Class-wide effects for on-task behavior.
Note. CW-FIT = Class-Wide Function-Related Intervention Teams.

(see Figure 3). For the comparison group teachers, they went from 8.42 in baseline to 9.49 in the second phase, a significant increase in reprimands $p = .01$.

Consumer Satisfaction

Teachers ($n = 66$) expressed overall satisfaction with the CW-FIT intervention. Mean ratings (1 = *high acceptability*, 4 = *low*) across items averaged 1.4 to 1.88 (*very true* to *mostly true*). Teachers commented that they thought the training, modeling by a consultant in their classroom, and regular feedback were very helpful in promoting their ability to implement the program. Many teachers commented that it helped them improve their use of positive statements to their students, and that the program assisted students in staying focused on lessons. A few suggestions were given for changing the intervention. Seven of the 66 responding

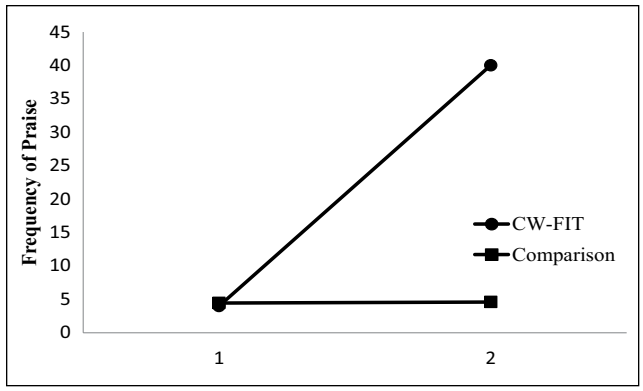

Figure 2. Class-wide effects for teacher praise.
Note. CW-FIT = Class-Wide Function-Related Intervention Teams.

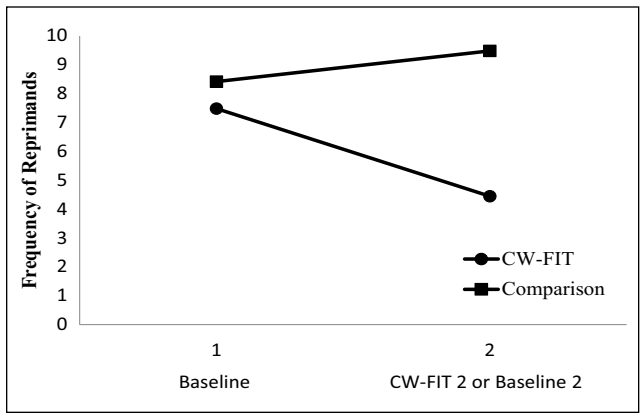

Figure 3. Class-wide effects for teacher reprimands.
Note. CW-FIT = Class-Wide Function-Related Intervention Teams.

teachers (10%) stated that the timer and giving points was distracting to their teaching. Two teachers suggested fewer tangible rewards. Follow-up informal inquiries of CW-FIT teachers indicated that 45% continued to use the intervention. The inquiries were completed 1 to 3 years after their participation in the research study.

Over the 4 years, 1,055 students were asked whether they liked the CW-FIT intervention; 89% said they liked it. Various reasons they provided included opportunities to earn prizes, better learning atmosphere, students were quieter during the lessons, an increase in the enjoyment of lessons, and getting better at self-monitoring their behavior. They reported they did not like it when their team missed a point and reviewing the rules so frequently.

Discussion

The purpose of this research was to conduct a randomized efficacy study of the CW-FIT program and determine outcomes for general education classrooms. Findings indicated that with the group contingency intervention (i.e., direct

teaching of classroom behaviors, teacher attention to appropriate behaviors, use of a point system, and rewards for use of the skills), on-task behavior increased dramatically. This outcome was replicated across 4 years in 17 schools and 86 classrooms. Results also showed that the improvements were significantly greater than changes in on-task behavior in 73 comparison classes that did not receive the intervention. Findings replicate prior CW-FIT studies showing improvements in on-task behavior when teachers implemented the group contingency (Kamps et al., 2011; Wills, Kamps, Hansen, et al., 2010). Findings also support prior studies showing improved student behavior when using group contingency interventions (Maggin et al., 2012; Theodore et al., 2004; Tingstrom et al., 2006), direct instruction of classroom skills (January, Casey, & Paulson, 2011; Mitchem et al., 2001), and differential reinforcement (Lloyd, Eberhardt, & Drake, 1996; Stage & Quiroz, 1997).

The current study was different from the prior Kamps et al. (2011) CW-FIT study in that classrooms were not concurrently participating in a SWPBS intervention. Findings suggest that a school-wide program is not necessary as a pre-requisite for use of the CW-FIT intervention in that the outcomes were similar. In our view, however, the SWPBS program in the prior Kamps et al. study provided several advantages. The SWPBS intervention team promoted maintenance of the CW-FIT intervention, recruitment of new teachers and students who would benefit from intervention, and consistent monitoring of outcomes independent of the researchers (Abbott et al., 2008). In addition, SWPBS promotes use of prosocial skills and appropriate classroom behavior.

The CW-FIT intervention in this study focused on increasing attention to appropriate behavior (Ervin, Miller, & Friman, 1996; Kelshaw-Levering et al., 2000; Thorne & Kamps, 2008) rather than on attention to negative behavior or the use of response cost as indicated in some group contingencies (Davies & Witte, 2000; Maggin et al., 2012). Increased frequencies of attention to appropriate behaviors by teachers participating in CW-FIT, through use of praise and point delivery, were similar to increases in the prior CW-FIT studies (Kamps et al., 2011; Wills, Kamps, Hansen, et al., 2010). This study showed baseline praise and attention to appropriate behavior frequencies of 3 to 4 during 20-min observation periods or a rate of about once every 5 to 7 min. These low rates are similar to the low levels of praise found in other large observation studies (Sutherland, Lewis-Palmer, Stichter, Morgan, 2008; Wills, Kamps, Abbott, Bannister, & Hansen, 2010). Praise and points during CW-FIT intervention, in contrast, averaged 2 per minute or 40 during the 20-min observations (see Table 1). The use of points along with praise was similar to the use of token economy systems (for a review, see Maggin et al., 2011). While the recent systematic evaluation of token economy studies indicated insufficient evidence to be deemed as an evidence-based practice (Maggin et al., 2011); the current study provides favorable evidence supporting token economies as a component of group contingencies with evidence of fidelity, rigorous experimental control, and social validity as recommended by the investigators.

The majority of teachers also were able to use simple, activity rewards (e.g., quiz games, 3-min dance party, no shoes for 1 hr) rather than tangible rewards for teams meeting their point goals. Also similar to prior studies, the rates of reprimands or attention to inappropriate behaviors occurred at a higher frequency than praise during baseline conditions (mean reprimands, range = 8–10 during 20-min or once a minute), for both the intervention and comparison class teachers. CW-FIT implementation reduced the occurrence to 5, or about once every 4 min, and given the increase in praise and points, dramatically improved the ratio of praise to reprimands to a level more in line with recommended best practices (Horner & Sugai, 2005; Sutherland, Wehby, & Yoder, 2002). It is possible that reprimands were used with less frequency due to decreased behavior problems.

A negative finding of the CW-FIT intervention study was the fact that less than 50% of the teachers who participated during the duration of the project continued to implement the protocol and CW-FIT strategies after the completion of the study. Follow-up inquiries were made at eight schools in the spring of the final project year. Twenty of 45 intervention teachers (44%) reported still using the intervention. This was self-report and not based on follow-up data. Though this is lower than desired, the inquiries were from 1 to 3 years following the study. In addition, few studies report maintenance of intervention beyond a few months, so it is unclear whether 44% is a typical or successful level of maintenance for interventions or not. One possible reason is that there was not as much sustained in-school support as there had been when the coaches were available to the buildings. The fairly low sustainability may be an indication of the need for a school-wide support structure such as SWPBS for maintenance of evidence-based practices.

There are several contributions of the study. In addition to the efficacy of the intervention, strengths of the study included the large number of participating classrooms and direct observations across experimental conditions. A randomized control group design, recommended as high-quality methodology in educational research, was used to demonstrate effects. The majority of reported randomized control trial studies of group contingencies have used the Good Behavior Game in which points are marked "against" teams for rule infractions (Kellam et al., 2014; Poduska et al., 2008). The current study adds to the group contingency literature by providing a randomized large sample study of another group contingency, CW-FIT, in which points are "awarded" to teams for appropriate behaviors.

A related unique feature was the collection of fidelity data across all conditions and for intervention and comparison classrooms; a weakness in the group contingency literature (Maggin et al., 2012). The high levels of fidelity quality and the large number of observations (more than 3,150 total; see "Measures" section) show the effectiveness of teacher training and confirm a relationship between the independent variable (CW-FIT procedures) and the dependent variable (class-wide on-task behavior). Interrater agreement percentages on fidelity across all conditions and groups were high (means of 85%–99%). The "general classroom management" checklist also provided interesting information from the study. The instruments' items reflect what many would consider good management practices (e.g., students are compliant and on-task during instruction, lessons are structured, transitions are short, specific and frequent praise is provided). Observations over multiple classrooms and years indicated that approximately 50% of good management practices were routinely in place. The CW-FIT intervention improved scores to 84% of the effective management practices in place, with little change in the comparison classes. It would have provided interesting data to conduct follow-up assessments of management skills to see whether teachers maintained the use of skills.

Fidelity ratings documented that even though intervention and comparison teachers/classes were located within the same schools, little change was noted in the comparison classes during Baseline 2 observations, suggesting limited contamination. This suggests the potential utility of this randomization strategy in future, large-scale studies.

The consumer satisfaction data across the 4 years of CW-FIT intervention provide strong evidence of social validity. Teachers found the intervention helpful for improving students' on-task behavior and for increasing their praise and positive interactions with students. One supportive finding for the teachers' acceptability is that the majority of teachers in the comparison classes in the study voluntarily attended training in the spring and received coaching.

Limitations

Although findings of the study support CW-FIT as a successful classroom group contingency for urban settings, there are several limitations to the study. First, each building had a part-time intervention coach to support teachers' use of the CW-FIT program. Researchers were also available to assist the coaches in consulting with teachers, as needed, yet no data were collected per se on "consulting time," so the impact of the coaching and feedback in addition to use of the CW-FIT procedures may have contributed to changes in teachers' behaviors. The typical rule was to schedule one session per week to observe and give feedback. Teachers with lower levels of procedural fidelity then received additional coaching. Future scalability research must investigate the use of the intervention without outside funding and university support, that is, with support typically available within districts for behavior interventions, including the use of school-wide systems to support teachers' use of behavioral interventions. Fidelity data suggest the feasibility of the CW-FIT in elementary schools using typical supports, but this needs to be demonstrated empirically. Further study is warranted to test for generalization effects of group contingency programs. In addition, the findings are based on 20 min of direct observation 1 to 2 times per week. This was during a time teachers determined to have the most problem behaviors; however, effects across the day were not observed. A low rate of interobserver agreement was collected (10% of observations), and observers were not blind to the experimental conditions. The current study describes effects for the CW-FIT as a multicomponent intervention. Additional study would need to determine which components were most influential for changes.

An additional limitation is that measures were not collected on student academic performance during CW-FIT conditions. It is likely that an additional component to the intervention would be necessary to improve students' academic performance.

Conclusion

In summary, the CW-FIT group contingency including direct instruction of appropriate classroom behavior improved class-wide on-task behavior. Findings suggest the use of CW-FIT as an effective intervention to address widespread concerns with school discipline problems. The use of a game format including teams and points for appropriate behaviors was viewed as enjoyable by students. Teachers valued training, viewed it as effective, acceptable, and easy to implement. We recommend the intervention in urban elementary school classrooms, and in classrooms in need of additional tools to improve classroom management of students' behaviors. CW-FIT may also be beneficial if used for a second period of the day, for example, a morning and afternoon block to help manage behaviors; or used during critical academic blocks, for example, reading and math instruction. The use of this and other group contingency interventions are recommended as efficient, relatively easy to implement, and low-cost procedures to improve on-task behaviors.

Acknowledgment

We gratefully acknowledge the participating teachers and students for their efforts.

Authors' Note

Opinions expressed herein are those of the authors and do not necessarily reflect the position of the funding agency.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research was funded by the Institute of Education Sciences, Department of Education (R324A07181).

References

Abbott, M., Wills, H., Kamps, D., Greenwood, C., Dawson-Bannister, H., Kaufman, J., Veerkamps, M., & Fillingin, D. (2008). The process of implementing a reading and behavior three-tier model: A case study in a Midwest elementary school. In C. Greenwood, T. Kratowill, M. Clemens, & R. Horner (Eds.), *Elementary school-wide prevention models: Real models and real lessons learned* (pp. 215–265). New York, NY: Guilford.

Barber, B., Maggin, D., & Wehby, J. (2009, April). *Improving the reliability and validity of classroom atmosphere assessment: The Classroom Atmosphere Rating Scale–Revised*. Presentation at AERA Annual Convention, San Diego, CA.

Chafouleas, S. M., Volpe, R. J., Gresham, F. M., & Cook, C. R. (2010). School-based behavioral assessment within problem-solving models: Current status and future directions. *School Psychology Review*, *39*, 343–349.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Davies, S., & Witte, R. (2000). Self-management and peer-monitoring within a group contingency to decrease uncontrolled verbalizations of children with attention-deficit/hyperactivity disorder. *Psychology in the Schools*, *37*, 135–147.

Embry, D. E. (2002). The Good Behavior Game: A best practice candidate as a universal behavioral vaccine. *Clinical Child and Family Psychology Review*, *5*, 273–297.

Epstein, M., Atkins, M., Cullinan, D., Kutash, K., & Weaver, R. (2008). *Reducing behavior problems in the elementary school classroom: An IES practice guide*. Washington, DC: Institute of Education Sciences.

Ervin, R., Miller, P., & Friman, P. (1996). Feed the hungry bee: Using positive peer reports to improve social interactions and acceptance of a socially rejected girl in residential care. *Journal of Applied Behavior Analysis*, *29*, 251–253.

Ervin, R. A., Radford, P. M., Bertsch, K., Piper, A. L., Ehrhardt, K. E., & Poling, A. (2001). A descriptive analysis and critique of the empirical literature on school-based functional assessment. *School Psychology Review*, *30*, 193–210.

Horner, R. H., & Sugai, G. (2005). School-wide positive behavior support: An alternative approach to discipline in schools. In L. Bambara & L. Kern (Eds.), *Positive behavior support* (pp. 359–390). New York, NY: Guilford.

January, A. M., Casey, R. J., & Paulson, D. (2011). A meta-analysis of classroom-wide interventions to build social skills: Do they work? *School Psychology Review*, *40*, 242–256.

Kamps, D., Wills, H., Heitzman-Powell, L., Laylin, J., Szoke, C., Hobohm, T., & Culey, A. (2011). Class-Wide Function-Related Intervention Teams: Effects of group contingency programs in urban classrooms. *Journal of Positive Behavior Interventions*, *13*, 154–167.

Kellam, S., Wang, W., Mackenzie, A., Brown, C., Ompad, D., Or, F., & Windham, A. (2014). The impact of the Good Behavior Game, a universal classroom-based preventive intervention in first and second grades, on high-risk sexual behaviors and drug abuse and dependence disorders into young adulthood. *Prevention Science*, *15*(1, Suppl.), 6–18.

Kelshaw-Levering, K., Sterling-Turner, H. E., Henry, J. R., & Skinner, C. H. (2000). Randomized interdependent group contingencies: Group reinforcement with a twist. *Psychology in the Schools*, *37*, 523–533.

Lea, T., Bray, M., Kehle, T., & DioGuardi, R. (2004). Contemporary review of group-oriented contingencies for disruptive behavior. *Journal of Applied School Psychology*, *20*, 79–101.

Leflot, G., vanLier, P., Onghena, P., & Colpin, H. (2013). The role of children's on-task behavior in the prevention of aggressive behavior development and peer rejection: A randomized controlled study of the Good Behavior Game in Belgian elementary classrooms. *Journal of School Psychology*, *51*, 187–199.

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

Lloyd, J. W., Eberhardt, M. J., & Drake, G. P. (1996). Group versus individual reinforcement contingencies within the context of group study conditions. *Journal of Applied Behavior Analysis*, *29*, 189–200.

Maggin, D., Chafouleas, S., Goddard, K., & Johnson, A. (2011). A systematic evaluation of token economies as a classroom management tool for students with challenging behavior. *Journal of School Psychology*, *49*, 529–554.

Maggin, D., Johnson, A., Chafouleas, S., Ruberto, L., & Berggren, M. (2012). A systematic evidence review of school-based group contingency interventions for students with challenging behavior. *Journal of School Psychology*, *50*, 625–654.

McGinnis, E. (2010). *Skillstreaming the elementary school child: A guide for teaching prosocial skills*. Champaign, IL: Research Press.

Mitchem, K. J., Young, K. R., West, R. P., & Benyo, J. (2001). CWPASM: A classwide peer-assisted self-management program for general education classrooms. *Education and Treatment of Children*, *24*, 111–140.

Poduska, J., Kellam, S., Wang, W., Brown, C., Ialongo, N., & Toyinbo, P. (2008). Impact of the Good Behavior Game, a universal classroom-based behavior intervention, on young adult service use for problems with emotions, behavior, or drugs or alcohol. *Drug and Alcohol Dependence*, *95*, S29–S44.

Reavis, K., Jenson, W., Kukic, S., & Morgan, D. (1988). *Utah's Best Project: Behavioral and educational strategies for teachers* (Technical Assistance Manuals). Salt Lake City: Utah State Office of Education.

Reinke, W. M., Stormont, M., Herman, K. C., Puri, R., & Goel, N. (2011). Supporting children's mental health in schools:

Teacher perceptions of needs, roles, and barriers. *School Psychology Quarterly*, *26*(1), 1–13.

Sheridan, S. (2010). *The tough kid social skills book*. Champaign, IL: Research Press.

Simonsen, B., Fairbanks, S., Briesch, A., Myers, D., & Sugai, G. (2008). Evidence-based practices in classroom management: Considerations for research to practice. *Education & Treatment for Children*, *31*, 351–380.

Skinner, C. H., Cashwell, T. H., & Skinner, A. L. (2000). Increasing tootling: The effects of a peer-monitored group contingency program on students' reports of peers' prosocial behaviors. *Psychology in the Schools*, *37*, 263–270.

Stage, S. A., & Quiroz, D. R. (1997). A meta-analysis of interventions to decrease disruptive classroom behavior in public education settings. *School Psychology Review*, *26*, 333–368.

Sutherland, K., Lewis-Palmer, T., Stichter, J., & Morgan, P. (2008). Examining the influence of teacher behavior and classroom context on the behavioral and academic outcomes for students with emotional or behavioral disorders. *The Journal of Special Education*, *41*, 223–233.

Sutherland, K., Wehby, J., & Yoder, P. (2002). Examination of the relationship between teacher praise and opportunities for students with EBD to respond to academic requests. *Journal of Emotional and Behavioral Disorders*, *10*, 5–13.

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

Thorne, S., & Kamps, D. (2008). The effects of a group contingency intervention on academic engagement and problem behavior of at-risk students. *Behavior Analysis in Practice*, *1*, 12–18.

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

Wehby, J. H., Dodge, K. A., & Greenberg, M. (1993). *Classroom Atmosphere Rating Scale* (Unpublished technical manual). Nashville, TN: Vanderbilt University.

Wills, H. P., Kamps, D., Abbott, M., Bannister, H., & Hansen, B. (2010). Classroom observations and effects of reading interventions for students at risk for emotional and behavioral disorders. *Behavioral Disorders*, *35*, 103–119.

Wills, H. P., Kamps, D., Hansen, B. D., Conklin, C., Bellinger, S., Neaderhiser, J., & Nsubuga, B. (2010). The Class-Wide Function-Based Intervention Team (CW-FIT) Program. *Preventing School Failure*, *54*, 164–171.