

2012

Evaluation of the check in/check out (CICO) intervention for students with internalizing behavior problems

Katherine K. Hunter

Louisiana State University and Agricultural and Mechanical College, kathunter110@gmail.com

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_theses



Part of the [Psychology Commons](#)

Recommended Citation

Hunter, Katherine K., "Evaluation of the check in/ check out (CICO) intervention for students with internalizing behavior problems" (2012). *LSU Master's Theses*. 1098.

https://digitalcommons.lsu.edu/gradschool_theses/1098

This Thesis is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Master's Theses by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.

EVALUATION OF THE CHECK IN/CHECK OUT (CICO) INTERVENTION FOR
STUDENTS WITH INTERNALIZING BEHAVIOR PROBLEMS

A Thesis

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Master of Arts

in

The Department of Psychology

by
Katherine K. Hunter
B.A., Tufts University, 2008
May 2012

TABLE OF CONTENTS

LIST OF TABLES.....	iv
LIST OF FIGURES.....	v
ABSTRACT.....	vi
INTRODUCTION.....	1
METHOD.....	11
RESULTS.....	25
DISCUSSION.....	40
CONCLUSIONS.....	47
REFERENCES.....	48
APPENDIX A: STUDENT INTERNALIZING BEHAVIOR SCREENER.....	53
APPENDIX B: STUDENT RISK SCREENING SCALE.....	54
APPENDIX C: DIRECT BEHAVIOR RATING, TEACHER FORM.....	54
APPENDIX D: DIRECT BEHAVIOR RATING, STUDENT FORM.....	56
APPENDIX E: DAILY PROGRESS REPORT.....	57
APPENDIX F: INTERVENTION INTEGRITY FORM, TEACHER.....	58

APPENDIX G: INTERVENTION INTEGRITY FORM, MENTOR.....59

APPENDIX H: INTERVENTION RATING PROFILE.....60

APPENDIX I: INSTITUTIONAL REVIEW BOARD APPROVAL.....61

VITA.....62

LIST OF TABLES

1. Treatment Integrity.....	21
2. Direct Behavior Ratings for All Participants.....	27
3. Direct Behavior Ratings for Subject 1.....	29
4. Direct Behavior Ratings for Subject 2.....	31
5. Direct Behavior Ratings for Subject 3.....	33
6. Direct Behavior Ratings for Subject 4.....	34
7. Correlations Among the Outcome Measures.....	35
8. Student Rating of Feelings.....	36
9. Pre- and Post-Scores on the SIBS.....	37
10. Pre- and Post-Scores on the SSIS:RS.....	38

LIST OF FIGURES

1. Graphs of the Daily Progress Report Data for Each Participant by Session.....	26
--	----

ABSTRACT

Internalizing behaviors are directed inward at the child and are often overlooked in classrooms compared to outwardly-directed externalizing behaviors. When internalizing behaviors are identified, Cognitive Behavioral Interventions (CBIs) are the go-to targeted interventions. However, CBIs are time-consuming and require considerable training to implement. An efficient and time-effective targeted intervention for internalizing behaviors is needed. Check In/Check Out (CICO) has been shown to be an effective targeted intervention for children with externalizing behavior problems. Therefore, the purpose of this study is to evaluate CICO as a targeted intervention for kids with internalizing behavior problems. Four elementary school students with internalizing behavior problems received the CICO intervention. A mixed single-subject design was used to evaluate the intervention's effectiveness. Results suggest that ratings of prosocial replacement behaviors increased during the CICO phase and that overall, participants internalizing behaviors decreased.

Keywords: targeted intervention, check in/check out, internalizing behaviors, school-age

INTRODUCTION

Tiered systems of delivery are the current standard for the implementation of services in schools. A tiered system can be applied to a myriad of domains within a school, including academic instruction, behavior management, social skills instruction, and mental health services. A sign of how integral tiered systems have become in school settings is the reauthorization of the Individuals with Disabilities Education Improvement Act (IDEIA) in 2004, which allowed the use of Response to Intervention, a tiered system, to justify placement of children into special education (Walker & Shinn, 2010).

The tiered model as originally presented by Walker, Horner, Sugai, and Bullis (1996) is a conceptual framework that involves three progressively intensive tiers of service delivery that are coordinated in order to reach all students. In this model, all students are exposed to a universal, school-wide system and for those students who fail to respond to the universal system, additional services are provided. Specifically, at the universal level, also referred to as Tier I, a program is implemented school-wide. This level serves as primary prevention, aimed to prevent the development of adverse outcomes, and thus should be effective for approximately 80% of students. For those students who are still at risk following the universal intervention, targeted interventions, referred to as Tier II, can be applied. These secondary prevention interventions are aimed at those students at risk for adverse outcomes who need additional supports but may not necessarily need intensive interventions. Such interventions should work for a majority of those students, or approximately 15% of all students. For the remaining students still at risk, tertiary prevention, or Tier III, is implemented, which consists of intensive interventions individualized to students' unique needs (Walker, Horner, Sugai, & Bullis, 1996; Walker & Shinn, 2010).

Within this model, all services delivered are evidence-based and implemented with integrity to ensure that lack of student response is not due to the quality of the instruction being received. In addition, the progress of all students is continuously monitored so students can move between tiers as their needs change. No single tier will be effective or necessary for every student but with the integrative use of multiple tiers, all students in schools should be provided effective services that are appropriately matched to their needs (Walker et al., 1996; Walker & Shinn, 2010).

This study focuses on the use of targeted interventions within a tiered model to deal with maladaptive cognitive, social, and emotional behaviors. Such behaviors fall into two categories: externalizing behaviors and internalizing behaviors (Achenbach & Edelbrock, 1978; Gresham & Kern, 2004; Merrell & Gueldner, 2010). Externalizing behaviors are behaviors that are directed outward and are considered to be undercontrolled (Gresham & Kern, 2004). These behaviors include aggressive behavior, conduct problems, disruptive behavior, hyperactivity-impulsivity, opposition/defiance, and acting out. Such behaviors are typically characteristic of Oppositional Defiant Disorder, Conduct Disorder, and Attention Deficit Hyperactive Disorder (Merrell & Gueldner, 2010). Targeted interventions for externalizing behaviors include Check In/Check Out (CICO), additional incentives for positive behavior, increased teacher prompts for correct behaviors, school-home notes, peer mentors, or social skills instruction groups (Sprick & Borgmeier, 2010; Todd, Campbell, Meyer, & Horner, 2008).

The second class of maladaptive behaviors, and the focus of this study, is internalizing behaviors. Internalizing behaviors are behaviors that are directed inward, at the person experiencing them, and are considered to be overcontrolled (Gresham & Kern, 2004; Merrell, 2008). These behaviors occur when individuals try to control their internal emotions or

cognitions to a maladaptive extent (Merrell & Gueldner, 2010). Examples of internalizing behaviors include social withdrawal, lack of participation both academically and socially, somatic complaints, poor self-esteem, negative thoughts about the self, depression, anxiety, and problems that are self-directed. Internalizing behaviors are characteristic of anxiety disorders and depressive disorders, thereby treatment of such behaviors typically falls under the umbrella of mental health services (Merrell, 2008; Merrell & Gueldner, 2010). Individuals with internalizing disorders often have impaired problem-solving abilities, pessimistic cognitive styles, distorted perceptions, and a lack of self-efficacy, as well as poor coping skills (Greenberg, Domitrovich, & Bumbarger, 2001).

By their very nature, internalizing behaviors are more difficult to detect than externalizing behaviors. Since internalizing behaviors are directed inward at the individual they are often overlooked and left untreated in a classroom setting as compared to externalizing behaviors. This is not surprising given that externalizing behaviors are disruptive to daily classroom functioning and are quickly and are thus immediately apparent to teachers. Internalizing behaviors, however, are often not disruptive to an entire classroom and therefore tend to go unnoticed (Gresham & Kern, 2004). As a result of this pattern, children with internalizing behaviors often become “invisible” in a classroom (Merrell & Gueldner, 2010).

The lack of identification and treatment of internalizing disorders is well-documented (Bradshaw, Buckley, & Ialongo, 2008; Greenberg et al., 2001; Gresham & Kern, 2004; Merrell, 2008). Gresham and Kern (2004) discuss that children with internalizing behaviors are not often referred for intervention as their behaviors do not clash with teacher’s behavioral expectations and do not go beyond teachers’ levels of tolerance. In a longitudinal study of 678 children, ranging from grades 1-9, Bradshaw, Buckley, and Ialongo (2008) found that children with

externalizing symptoms were significantly more likely to receive services in schools than children with internalizing symptoms. They coined this phenomenon the “squeaky wheel” phenomenon as those children who are “squeaking”, i.e. making noise, are the ones who get the attention. In addition, there is a considerable lack of research devoted to internalizing disorders compared to research on disruptive behaviors (Greenberg et al., 2001).

In 2003, the Depression and Bipolar Support Alliance reported that high rates of children with depression and bipolar disorder were untreated, that effective treatments were not widely applied, that schools were an underutilized setting for identification and treatment, and that teachers and administration had low awareness of mental health issues (Coyle et al., 2003). Further, it has been estimated that although one in five children needs mental health services, only 30% of these children receive those services (Cheney, Flower, & Templeton, 2008).

The tendency to overlook and to underserve internalizing disorders in children is especially troubling considering that internalizing symptoms in adolescents have been linked to a number of maladaptive outcomes. Such outcomes include future and more severe internalizing disorders, academic problems, externalizing behavior problems, peer rejection, low self-esteem, suicidal behavior, lack of employment opportunities, poor physical health and substance abuse disorders (Compton, Burns, Egger, & Robertson, 2002; Reinherz et al., 2006; Sourander & Helstela, 2005; Vasa & Pine, 2006).

Internalizing problems tend to increase in adolescence, with the early adolescent period, ages 10-14, being especially critical for development of future internalizing problems (Roeser, van der Wolf, & Strobel, 2001). Merrell and Gueldner (2010) report that for school-age children there is a point prevalence of internalizing symptoms of 4-6% and that these problems accumulate over time, which may partially account for the larger prevalence of internalizing

symptoms during adolescence. Even as early as the age of 8, internalizing problems have been shown to predict higher intensity internalizing problems later in adolescence, and into adulthood (Sourander & Helstela, 2005). It is estimated that approximately 10-20% of children will have an anxiety disorder throughout their development and that those children who are untreated are especially likely to have an anxiety disorder that persists into adulthood (Vasa & Pine, 2006). Childhood or adolescent depressive episodes are often followed by the continuation of depression into adulthood as well (Reinherz et al., 2006). Additionally, Merrell (2008) suggests that many children experience internalizing problems but do not fit specific diagnostic requirements, suggesting that the prevalence could be higher than estimated. This pervasiveness and stability of internalizing problems highlights the need to identify and treat internalizing disorders early, which contrasts with the current trend of insufficient identification and services.

Despite the tendency of internalizing disorders to go overlooked and untreated in schools, there are some interventions for internalizing behaviors that have been proven to be effective. Targeted interventions for internalizing disorders include small group counseling and skills training, modeling, systematic desensitization, exposure-based interventions, antecedent- and consequence-based interventions and cognitive-behavioral interventions (CBIs) (Gresham & Kern, 2004; Merrell & Gueldner, 2010).

Currently, CBIs are the most common form of targeted intervention for internalizing behaviors, specifically depression and anxiety (Compton et al., 2002). CBIs focus on the areas of cognitions and behaviors (Maag & Swearer, 2005) and typically involve small groups of students, or an individual student, working with a counselor or social worker in which cognitive and behavioral skills are taught, modeled, and practiced (Compton et al., 2002). After students are taught how to identify somatic and cognitive aspects of depression and/or anxiety, they are

instructed in techniques to address these components, using self-monitoring as they attempt to change their ways of thinking and engage in more positive, constructive activities. If a CBI is targeting depression, the goal is to increase students' knowledge and awareness of depressive thoughts, as well as to change maladaptive cognitions/thought patterns (Merrell & Gueldner, 2010). If the child is suffering primarily from anxiety, the central goal of a CBI is to help change the child's thought patterns and reactions surrounding anxiety-inducing experiences (Vasa & Pine, 2006).

Stark, Reynolds, and Kaslow (1987) documented improved self-report and clinical scale depressive symptoms for 9-12 year old children after receiving a cognitive-behavioral focused school-based intervention. Dadds, Spence, Holland, Barrett, and Laurens (1997) found lower rates of anxiety disorder following a cognitive-behavioral school-based treatment of 7-14 year olds when compared to a group that did not receive treatment. Despite these two examples, most of the studies looking at the effects of cognitive-behavior therapy on anxiety and depression are efficacy trials that lack generalizability to a school setting (Cartwright-Hatton, Roberts, Chitsabesan, Fothergill, & Harrington, 2004; Maag & Swearer, 2005).

Although CBIs are known to be effective, certain aspects of this targeted intervention make it difficult to provide in a school environment. This is a concern considering that schools have become the primary place where children receive mental health services (Kutash, Duchnowski, & Lynn, 2006). CBI programs are time-consuming, and may take up to 16-30 lessons of 1-2 hours each (Merrell & Gueldner, 2010). Besides being time-consuming, correct implementation of a CBI as designed for Tier II requires the service of a professional, and is beyond the scope of services that can be provided by a classroom teacher (Maag & Swearer, 2005). This professional has to be well-versed in the program, which may necessitate intensive

training. Two of the vital aspects of a targeted intervention are that they are efficient and cost effective, meaning that it can be used for multiple students without taking up significant staff time (Hawken & Hess, 2006; Hawken, MacLeod, & Rawlings, 2007). CBIs do not meet this requirement as they require a generous amount of staff time and additional personnel requirements, making implementation of a CBI not always feasible in a school setting.

In order for an intervention to be used in a school setting it should be both efficacious (i.e. works under ideal, tightly controlled conditions) and effective (i.e. works under less than ideal, loosely controlled conditions). Additionally, in order for an intervention to be effective in a natural setting it should be considered acceptable by teachers and school personnel, fit in easily to already established routines, and require minimal time, effort, and additional training (Walker & Shinn, 2010). To address some of the aforementioned hindrances of CBIs and to appropriately serve children with internalizing behaviors, an evidence-based intervention that is quick and easy to implement is needed. Therefore, the purpose of the current study is to assess the effectiveness of a potential intervention that could serve this purpose. The intervention chosen for this analysis is Check In/Check Out (CICO).

Check In/Check Out is an evidence-based targeted intervention for externalizing behavior problems (Crone, Hawken, & Horner, 2010). Origins for CICO can be seen in the concept of pre-correction, in which teachers remind students of appropriate behaviors before problem behaviors have a chance to occur, thereby correcting the problem before it starts. Examples of pre-corrections include verbal reminders, role-play of appropriate behaviors, or demonstrations of correct behaviors (Colvin, Sugai, Good, & Lee, 1997). Pre-correction typically is accompanied by active supervision, with someone monitoring student behavior and reminding students of appropriate behavior as situations occur. Pre-correction generally has been used to

decrease externalizing problem behaviors and has been shown to be effective for general classroom behavior, at recess, and during times of transition (Colvin et al., 1997; DePry & Sugai, 2002; Lewis, Colvin, & Sugai, 2000).

Check In/Check Out is similar to pre-correction in that it involves reminding students of target behaviors before they occur. However, with CICO active supervision is replaced with feedback on an interval schedule, with an added contingency for appropriate behavior. The Behavior Education Program (BEP) is a CICO program that uses the concept of pre-correction. The goals of a CICO program, such as the BEP, are to provide prompts for correct behavior before behavior has a chance to occur, to increase adult feedback contingent on appropriate behavior, to help provide daily structure for students throughout the day, and to give feedback to parents on their child's behavior (Filter, et al., 2007).

Check In/Check Out has been shown to be an effective targeted intervention for problem behavior across age groups (Hawken & Hess, 2006; Hawken & Horner, 2003; Hawken et al., 2007). A study by Todd, Campbell, Meyer, and Horner (2008) found that implementation of CICO reduced office discipline referrals as well as occurrence of problem behavior for four boys in Kindergarten through third grade. In another study, Hawken and Horner (2003), found that implementation of the BEP reduced frequency of problem behaviors and increased academic engaged time for four students in the sixth grade.

CICO consists of five steps. During the first step, the participating child checks in with an adult mentor in the morning who reviews the child's performance from the day before, reminds them of their behavioral goals, and gives them a daily progress report (DPR). The DPR specifies the target behaviors and is broken down into various time slots over the course of the day. The child earns points for performing target behaviors correctly within each period and has

a daily point goal. Throughout the day, the child gives the DPR to their teacher at the beginning of each designated time interval (step two). At the end of the time period the teacher fills out the DPR and gives the child feedback and praise for good behavior (step three). At the end of the day (step four), the child turns in the DPR to the mentor who then calculates the total points earned that day and gives the child a reward if he or she reaches his or her goal. For the final step, the DPR is brought home, signed by the child's parents, and returned to the mentor the next morning (Crone et al., 2010). The DPRs can be analyzed frequently to assess if CICO is working and the intervention can be modified if it is not (Hawken & Hess, 2006). Examples of such modification include adjusting goals, redefining target behaviors, or changing the reward.

CICO has been shown to be acceptable to teachers and students (Hawken & Horner, 2003, Todd et al., 2008), which is important for an intervention to be effective in a school. A study by Filter et al. (2007) found that CICO could be implemented with high fidelity by school personnel with a single 4-hour training session, showing that limited training and resources are required for effective implementation. In addition, multiple studies have reported that school personnel find CICO to be both effective and efficient (Filter et al, 2007; Hawken et al., 2007; Todd et al. 2008), which are two essential elements of a targeted intervention (Hawken & Hess, 2006). Two other important aspects of CICO as a targeted intervention are that it can be started very quickly once a student is referred and it requires little effort from parents and teachers (Crone et al., 2010; Hawken & Hess, 2006). CICO meets the requirements set out by Walker and Shinn (2010) for an effective school-based intervention as it has been shown to be considered acceptable by teachers and school personnel, fit in easily to already established routines, and require minimal time, effort, and additional training.

Although CICO has not been used for internalizing behaviors, there is some evidence that it would work for such behaviors. A study by Cheney et al. (2009) found that following a randomized control trial of a CICO program called Check, Connect, Expect, those who received the intervention improved on the internalizing measure of the *Teacher Report Form* (TRF; Achenbach & Rescorla, 2001), even though internalizing behaviors were not specifically targeted in the intervention. These results suggest that a CICO program could work to ameliorate internalizing symptoms. The authors go on to posit that using a DPR for a child with internalizing behaviors could help improve those behaviors as it would increase interactions between the teacher and that student and encourage social interactions with peers. In addition, as protective factors against the development of psychopathology include cognitive skills, attachments to adults encouraging prosocial behaviors, and social support (Greenberg et al., 2001). CICO would provide a child the support of an adult mentor who encourages prosocial behaviors and helps the child develop cognitive skills to better cope with maladaptive feelings.

Conceptually, CICO appears to be an appropriate intervention aimed at improving internalizing behaviors. Just as CBIs focus on maladaptive cognitions and behavior, CICO can be constructed to focus on cognition and behavior as well. CICO programs, such as the BEP, are based on the behavioral principles that students who are at-risk profit from clear expectations, continuous feedback, a consistent routine, positive reinforcement when goals are met, and a meaningful relationship with an adult (Crone et al., 2010). With the check-in and check-out aspect, the adult mentor could use that time to work on cognitions as well, thereby making CICO compatible with a CBI. Therefore, this study directly assessed the effectiveness of CICO as a school-based intervention to reduce internalizing symptoms in preadolescent students.

METHOD

Participants

Four students in fourth grade were selected to participate in this study. These students attended a public elementary school in East Baton Rouge Parish in Louisiana. Participants were chosen through teacher nomination using a multiple-gating procedure on the basis of possessing significant levels of internalizing behaviors. Students who also exhibited significant levels of co-occurring externalizing behaviors were excluded from this study as CICO has already been shown to be effective for such students. The goal of the current study was to explore the effectiveness of CICO solely as an intervention for internalizing behavior. The first four students who met these criteria were included in this study.

Materials and Measures

Student Internalizing Behavior Screener. The *Student Internalizing Behavior Screener* (SIBS; Cook, in press) is a screening instrument for students in Grades 1-5 that uses teacher rankings to identify students at-risk for developing internalizing behavior problems. The SIBS has seven items (e.g. nervous or fearful, spends time alone) that have been identified as critical behavioral indicators of internalizing behavior problems. Teachers rank all students in the class on how often they display the seven items using a 4-point Likert scale, ranging from 0 for *Never* to 3 for *Frequently*. Total scores on the SIBS can range from 0-21, with higher scores representing greater occurrence of internalizing behavior problems. The SIBS has been shown to have an internal consistency over fall and winter screenings of .81 and .79. In addition, it shows temporal stability with a test-retest coefficient of .74. The SIBS correlates strongly with the Achenbach TRF Internalizing scale, with a correlation of .82. A cut-off score of 8 separates students who are at-risk from those who are not at-risk for developing internalizing behavior

problems, identifying 86% of true positives and only 1% of false negatives. Refer to Appendix A to view the SIBS.

Student Risk Screening Scale. The *Student Risk Screening Scale* (SRSS; Drummond, 1994) is a universal screening tool used to identify students in elementary school, middle school, and high school as at-risk for developing an externalizing disorder. The SRSS has seven items (e.g. steals, has low achievement) that have been shown to be critical indicators of future externalizing behavior problems. The teacher ranks all students in the class on how often they display the seven items on a 4-point Likert scale, with 0 for *Never* to 3 for *Frequently*. Total scores can range from 0-21 with higher scores representing a greater occurrence of externalizing behaviors. Scores from 9-21 indicate the student is at high-risk, scores from 4-8 indicate moderate-risk, and scores from 0-3 indicate low-risk. The SRSS has a high internal consistency ($\alpha = .81$) as well as strong test-retest reliability ($r = .73$). Additionally, the SRSS has both a sensitivity and specificity of .95. Refer to Appendix B to view the SRSS.

Social Skills Improvement System: Rating Scales- Teacher Form. Prior to implementation and at completion of the intervention the teacher form of the SSIS: RS was collected to assess change in social skills as a result of the intervention. The *Social Skills Improvement System: Rating Scales* (SSIS: RS; Gresham & Elliott, 2008) is a “multi-rater assessment of the perceived frequency and importance of a student’s social behaviors.” The measure is a standardized questionnaire that uses a Likert-type scale to assess frequency (*never, sometimes, often, very often*) and importance (*not important, important, critical*) of social skills, problem behaviors, and academic competence. For this study, two scales from the SSIS: RS were analyzed. The Internalizing Scale was used as an estimate of internalizing behaviors before and after intervention implementation. Performance on this scale is measured using raw scores to

classify participants as *below average*, *average*, or *above average*. Norms for raw scores based on gender and age are also provided for the purpose of comparison. Scores on the Social Skills Scale were examined pre- and post-intervention to provide a general measure of social skills. Results on this scale are analyzed using standard scores, with a mean of 100 and a standard deviation of 15. Scores are grouped as *well-below average*, *below-average*, *average*, *above-average*, or *well-above average*.

Direct Behavior Rating. A Direct Behavior Rating (DBR) is a behavioral assessment that is a hybrid of systematic direct observation and behavior rating scales. A DBR is an evaluative rating of behavior that is completed at the time and place when behavior occurs (Christ, Riley-Tillman, & Chafouleas, 2009). DBRs consist of a sample of target behaviors to be rated by a specified observer. They have been shown to correlate strongly with direct observation for ratings of both on-task behavior ($r = .81$) and disruptive behavior ($r = .87$) (Riley-Tillman, Chafouleas, Sassu, Chanese, & Glazer, 2008).

In the present study, DBRs were completed both by the teacher and the student. The items on the DBR were rated on a Likert-scale. The teacher questions consisted of target replacement behaviors that were rated for occurrence on a 9-point Likert scale, with 1 for *Never/Seldom* and 9 for *Usually/Always*. The child DBR consisted of the same behaviors as the teacher DBR but with an additional item measuring a subjective rating of feeling on a scale from 1-10. Refer to Appendix C for a sample teacher DBR and to Appendix D for a sample student DBR.

Daily Progress Report and Reward. The Daily Progress Report (DPR) is an integral part of the CICO intervention as defined by Crone, Hawken, and Horner (2010). The DPR is individualized for each child and is a tool used to evaluate his or her performance of the target

replacement behaviors. The replacement behaviors are listed on the DPR with a slot for the teacher to evaluate target behavior performance in multiple designated time periods throughout the day. Time periods for recording performance are created that fit in naturally with the student's schedule (e.g. at the end of reading, math, etc.). The teacher rates the student's performance of each behavior on a 3-point Likert scale, with 0 for *Needs Improvement*, 1 for *OK*, and 2 for *Good*. For each time period the total number of points earned is calculated and if the child reaches his or her daily point goal, he or she receives a reward at the end of the day. Additionally, there is a spot on the DPR for teacher comments and signature. Teacher ratings on DPRs have been shown to be related to problem behaviors as rated by observers in the classroom (Todd et al., 2008). Refer to Appendix E for a sample DPR.

If the student meets his or her daily point goal, he or she receives a reward at the end of the day. This reward is individualized to the student and is agreed upon by the mentor and the child at the start of the intervention or at the beginning of each day. Rewards do not need to be large and can consist of tangible items (e.g. pencils, silly bands), edibles (e.g. candy, chips), or activities (e.g. 5 minutes of computer time or drawing time). Parents and teachers are consulted before choosing rewards to ensure that the rewards are acceptable both at school and to the parent.

Treatment Integrity Forms: CI/CO/Teacher. Treatment integrity forms were created for the purpose of this study to measure adherence to the steps of the CICO intervention. Forms were completed by the party responsible for completing the intervention and consisted of a checklist of important tasks (e.g. did the mentor go over daily goals with the participant?). Mentor forms included a Check-In integrity form and a Check-Out integrity form. Teacher integrity forms measured teacher responsibilities, such as completing the DPR at the end of each

time period and providing feedback for replacement behaviors. The DPR served as a permanent product of the participant's integrity. Refer to Appendix F for a sample teacher integrity form and to Appendix G for a sample mentor integrity form.

Intervention Rating Profile-15. The *Intervention Rating Profile-15* (IRP-15; Witt & Elliot, 1985) is a 15-item questionnaire completed by the teacher to assess the acceptability of an intervention. All 15 items are answered using a 6-point Likert scale, with 1 for *Strongly Disagree* and 6 for *Strongly Agree*. Refer to Appendix H to view the IRP-15.

Procedure

Student Screening and Selection. In order to identify students with internalizing behaviors, a multiple-gating procedure was used. To pass through the first gate, students had to score an 8 or above on the SIBS as rated by their teachers. Teachers were asked to nominate students who exhibited the listed behaviors on the SIBS and to complete the screener for each child nominated. Any child who scored at or above the cut-off score of 8 progressed to the second gate, in which the teacher filled out the SRSS. To pass through the second gate a child could not score at or above the cut-off score of 9. This gate was used to ensure that the participants did not display comorbid externalizing problems as this study aimed to evaluate the use of CICO for children who primarily exhibit internalizing behavior problems. The first four students who passed through both gates were selected to participate in this study.

Patrick was an 11 year-old African American male in the fourth grade. He participated in the intervention in the Spring of 2011. Chris was a 10 year-old Caucasian male in the fourth grade who participated in the intervention in the Spring of 2011. Caroline was a 9 year-old Caucasian female in the fourth grade who received the intervention in the Fall of 2011. The final participant, Jeff, was a 9 year-old Caucasian male in the fourth grade who received the

intervention in the Fall of 2011. All students were in a regular education classroom and had no reported diagnoses.

Pre-baseline. Following selection to participate in the study, the top rated behaviors on the SIBS were operationally defined for each student. In order to do this, multiple steps were taken. First, a brief teacher interview was conducted. The purpose of this interview was to determine what behaviors the teacher observed to justify a high rating of that behavior on the scale. Frequency and setting of these behaviors was determined as well. Additionally, a brief interview with the child was conducted to determine what the behaviors looked and felt like to them, when they occurred, and how frequently they occurred. Based on the teacher and student interviews, the behaviors were operationally defined. In addition to target problem behaviors, replacement behaviors were also identified. These behaviors were prosocial behaviors designed to reduce the target problem behaviors. The replacement behaviors were the focus of CICO, the DPRs, and the student and teacher DBRs. All definitions of target behaviors and replacement behaviors were idiosyncratic to the participant. The chosen replacement behaviors for each participant are described in the “Dependent Measures” section of this paper.

Baseline. Before implementing the intervention, a baseline phase occurred for each participant. For the first participant, baseline lasted until a stable pattern of behavior was exhibited on teacher-completed DPRs. As standard to a multiple baseline design, each of the following participants had a baseline based on the performance of the participant before, which was determined by when the previous participant started to show effects of the intervention. All baseline phases had to be of a sufficient length for the participant to exhibit stable patterns of behavior.

During baseline, the teacher and the student completed DBRs. Replacement behaviors were measured on the DBRs as operationally defined in the pre-baseline phase. Additionally during baseline, teachers filled out DPRs for the replacement behaviors that were the target of the CICO intervention. The replacement behaviors on the DPRs were the same behaviors listed on the DBRs and were operationally defined in the same way. During baseline, the student did not see the DPRs and did not receive feedback based on performance of the behaviors listed on the DPRs.

CICO Intervention. The daily CICO procedure followed the steps of the CICO intervention as laid out by Crone et al. (2010). The general guidelines of these steps were followed but were adjusted to target internalizing behaviors.

The first step of the CICO intervention involved the student checking in with an adult mentor at the beginning of the school day. Each student was assigned two adult mentors who shared the check in and check out responsibilities equally. The adult mentors were school psychology graduate students who practiced at the school the participants attended. The student and the adult reviewed daily goals, which were listed on the DPR and consisted of performing the replacement behaviors identified in the pre-baseline phase as mentioned above. The participant and the mentor reviewed the daily point goal for the day and discussed what the reward would be for earning the target number of points. Each day's goal was determined by averaging the number of points earned by the student in the previous three days of the intervention or baseline. Additionally during check in, if the participant had any problems performing the behaviors the previous day, the mentor and the participant discussed the problems and attempted to problem-solve ways to make the upcoming day more successful. For each student this step was individualized to their behaviors and needs. If necessary this step included

working on identifying maladaptive, negative thinking patterns and replacing them with positive and productive thinking and/or problem-solving strategies, such as typically would be done when implementing a CBI.

The second and third steps of the CICO intervention were connected, as the student carried the DPR with them throughout the day and received teacher feedback at the end of every designated time period specified on the DPR. In order to receive feedback, the child was responsible for raising his or her hand to get the teacher's attention and having the teacher rate the appropriate time interval on the DPR. The DPR had positive replacement behaviors listed and the student was scored based on performance of those behaviors during the indicated time period. When scoring behavior, the teacher also gave reinforcement through verbal praise for displaying target replacement behaviors and encouraged the student to continue to display appropriate behaviors.

The final step of the CICO intervention involved the student checking out with the mentor at the end of the day. The student gave the mentor the DPR and together they calculated total points earned. If the participant met their daily goal, a daily reward and verbal praise was given by the mentor. If the participant did not meet the daily goal, no reward was given and the mentor helped the participant problem-solve as to why the participant was unable to perform the replacement behaviors and brainstorm ways to perform the behaviors the next day. A summary sheet of the day's performance was completed by the mentor and sent home to the participant's parents to be signed and returned the next day. Additionally, the student and teacher filled out a DBR at the end of every day, rating performance of the prosocial replacement behaviors.

Dependent Measures. The DPRs from the CICO procedure served as the primary dependent measure of the child's functioning in class and teacher perception of behavior

throughout the day. The DPRs only focused on the replacement behaviors, not the internalizing behaviors. Teachers and students also completed DBRs for the target replacement behaviors daily. The behaviors on the teacher and student DBRs were the same behaviors and were operationally defined in the same way.

Patrick's target behaviors were: *responds to directions within 5 seconds of being given, pays attention to the lesson and focuses on completing classwork, and makes eye contact with adults when having a conversation with them.* These behaviors were chosen for Patrick as he was slow to respond to directions, lethargic, and withdrawn, hardly making eye contact with others when speaking to them.

Chris's target behaviors were: *makes eye contact with adults when speaking to them, communicates appropriately with an appropriate tone of voice, only writes one thought in his "thought journal" per subject, and is on-task, requiring no more than three redirections.* Chris' teacher already had a "thought journal" system in place for him where he could write down his troubling thoughts and receive written responses from her a couple of times per day. However, it had gotten to the point where Chris was writing in his journal throughout most of instructional time, which is why his journal was targeted in this intervention. Additionally, Chris had difficulties making appropriate eye contact and communicating with an appropriate volume and tone of voice, thus these behaviors were targeted in his intervention.

Caroline's target behaviors were: *performs her group role responsibilities, speaks clearly and loudly enough to be heard, makes eye contact when interacting with others, and is on-task, participating appropriately in the class activity.* The main internalizing behaviors exhibited by Caroline were withdrawal and lack of interaction with others. Her target behaviors were designed to increase appropriate interactions. Her class had a "group role responsibility"

assigned to them each day, which required them to interact with members of their group and perform certain tasks. Caroline would not always perform her responsibility, thus it was chosen as a target of the intervention.

Jeff's target behaviors were: *completes written assignments neatly and in cursive, responds appropriately when something does not go his way (no whining or talking back), transitions quickly between tasks with materials ready, and is on-task, requiring no more than three redirections*. For Jeff, completing written work was something that was anxiety-provoking as he did not like to write, which is why it was targeted in the intervention. For all of the participants an on-task behavior was targeted as inattention often accompanies internalizing behaviors (Herman & Ostrander, 2007). Additionally, it was hoped that class participation would increase by targeting on-task behavior.

Additionally, the SIBS and the SSIS: RS (Gresham & Elliott, 2008) were completed by teachers before and after the intervention, acting as measures of change of student behavior from pre- to post-intervention. On the SIBS, the scores pre-intervention as rated by teachers were as follows: Patrick was rated a 9, Chris was rated a 16, Caroline was rated a 17, and Jeff was rated a 14. On the SSIS: RS completed before the intervention, Patrick obtained a raw score of 9 on the Internalizing scale, placing him in the *average* range. He had a standard score of 78 on the Social Skills scale, which was in the *below average* range. Chris' pre-scores on the SSIS: RS were 19 for the Internalizing scale, falling in the *above average* range and 83 on the Social Skills scale, falling in the *below average* range. Caroline's scores on the pre-intervention SSIS: RS were 19 for the Internalizing scale, which was in the *above average* range and 57 on the Social Skills scale, which classified her as *well-below average*. On the pre-intervention SSIS: RS as rated by

his teacher, Jeff scored an 11 on the Internalizing scale, which was *above average* and a 64 on the Social Skills scale, which was *well-below average*.

Treatment Integrity. Treatment integrity was collected daily for the check-in procedure, the check-out procedure, and teacher obligations. Integrity was collected through a self-report checklist for the mentor responsible for checking in and checking out, as well as for the teacher. Once a week, an observer also recorded procedural integrity for each of the three components to account for any possible self-report biases. Average total treatment integrity for check-in was 86.62%, ranging from 0%-100%. For check-out, integrity averaged 97.37%, ranging from 20%-100%. Average teacher integrity was 92.21%, ranging from 0%-100%. Days where integrity was 0% were all days when the treatment integrity form was not completed. Total study integrity averaged to 92.07%. Inter-observer agreement was collected for 16.43% of sessions and averaged 96.46%. Refer to Table 1 for integrity data by participant.

Table 1. Treatment Integrity. Treatment integrity is presented by participant for the mentor check-in duties, mentor check-out duties, and teacher duties. The mean is presented with the range in parentheses. The IOA for each participant is presented in the last column, with percentage agreement in parentheses below. IOA across all participants for each of the three responsibilities is presented as well. There was 92.07% integrity across all participants.

	Mentor Check-In	Mentor Check-Out	Teacher	IOA Total (Percent Agreement)
Patrick	88.24% (0%-100%)	96.47% (80%-100%)	97.06% (50%-100%)	13.72% (93.57%)
Chris	90.00% (0%-100%)	98.00% (80%-100%)	90.00% (0%-100%)	20.00% (100%)
Caroline	72.92% (0%-100%)	100.00%	89.58% (0%-100%)	17.00% (95.83%)
Jeff	95.31% (75%-100%)	95.00% (20%-100%)	92.19% (0%-100%)	15.00% (96.43%)
Total	86.62% (0%-100%)	97.37% (20%-100%)	92.21% (0%-100%)	16.43% (96.46%)

Acceptability. Acceptability of CICO as an intervention for internalizing behaviors was measured at the beginning and end of the intervention by the teacher. The teacher filled out an IRP-15 before the intervention was implemented and after it was terminated for each child in their classroom participating in the study. Acceptability was important to assess as CICO has not previously been used as an intervention targeting internalizing behaviors. Before intervention implementation, acceptability ratings averaged 5.60 across teachers. After the intervention, acceptability ratings averaged 5.48. Though this represents a small decrease, the acceptability is still high, suggesting that this intervention was considered acceptable by teachers as a target for internalizing behaviors.

Experimental Design and Data Analysis

A mixed reversal/multiple-baseline design was used. An ABAB reversal design was employed for the first participant in order to evaluate whether treatment effects remained or decreased following removal of the intervention. For this participant, after stability in baseline was achieved as indicated by teacher DPRs, the CICO intervention was implemented. Upon achieving stability in the data during the intervention phase, the intervention was removed and another baseline condition was measured. Upon achieving stability in this phase, the intervention was reinstated.

A non-concurrent AB multiple-baseline design was used for the last three participants. For participant two, following stability in baseline, the CICO intervention was implemented. Once the intervention phase was stable for the second participant, the intervention was implemented for the third and fourth participant.

Visual analysis of the DPRs was employed as the primary data-analytic tool to inspect changes in behavior between baseline and treatment phases (Kazdin, 1982). Additionally,

standardized mean difference (Cohen's *d*; Cohen, 1988) and Percent of All Non-overlapping Data (PAND; Parker et al., 2007) supplemented visual analysis.

In order to calculate a standardized mean difference, all of the baseline scores were averaged to obtain a mean baseline score and all of the treatment scores were averaged to obtain a mean treatment score. The mean baseline score was then subtracted from the mean treatment score and this number was divided by the standard deviation of baseline. The standard deviation of baseline was used as baseline scores represent typical levels of responding. According to Cohen's interpretations, a standardized mean difference score of .20 or less represents a small effect size, a score of .50 represents a medium effect size and a score of .80 or greater represents a large effect size (Cohen, 1988).

As described by Parker, Hagan-Burke, and Vannest (2007), PAND is a nonparametric estimate of effect size that reflects the amount of data that does not overlap between phases. PAND utilizes all data points, as opposed to just a single baseline point as in Percent of Non-Overlapping Data (PND), a similar nonparametric measure. In order to calculate PAND, the number of overlapping data points in baseline when compared to treatment is identified across all participants. This number is then divided by the total number of points and multiplied by 100 in order to provide a percentage of overlapping data. The percentage of overlapping data is subtracted from 100 in order to provide the percentage of all non-overlapping data, or PAND statistic. Parker, Hagan-Burke, and Vannest (2007) did not provide criteria for how to interpret the PAND statistic. However, Scruggs and Mastropieri (1998) suggest that when interpreting PND, greater than 90% should be considered very effective, between 70% and 90% considered effective, between 50% and 70% considered questionable and below 50% considered ineffective.

The authors of the current study used these criteria to interpret the PAND result as PND and PAND are comparable methods of estimating effect in single-subject research.

To further supplement evaluation of the effects of the intervention, teacher- and student-completed DBR data was evaluated using standardized mean difference calculations. Post-scores from the SIBS and SSIS: RS were examined to determine if participants were still at-risk or were functioning within the normal range of internalizing behaviors and social skills after participating in the intervention.

RESULTS

Daily Progress Reports (DPRs)

The data from the DPRs are presented as percentage of total possible points, as the number of possible points differed across participants (refer to Figure 1). Additionally, mean scores of baseline and treatment DPRs, as well as the standardized mean difference statistic for each participant and across participants are presented in Table 2. Visual analysis of Patrick's data (see Figure 1) suggests the intervention was effective in reducing internalizing behaviors through the increase of the pro-social replacement behaviors. A decreasing score in baseline was followed by an immediate increase during treatment. When treatment was removed, the DPR score immediately decreased, and continued to do so until the re-introduction of treatment, when there was once again, an immediate increase. DPR levels rose slower in the second treatment phase than in the first but they ended on a high level, suggesting an effect. As presented in Table 2, during baseline, Patrick's DPRs had a mean score of 48.33%, which increased to 80.78% in treatment. The standardized mean difference for the scores is 2.15, which further suggests that the intervention was effective in decreasing internalizing behaviors through the increase of replacement behaviors for Patrick.

Visual analysis of Chris' data (refer to Figure 1) also suggests that the intervention was effective. The scores on Chris' DPRs exhibited stability in baseline, being low throughout the phase, and then showed an immediate increase following the implementation of treatment. With the exception of the low score for session 8, Chris' DPR scores during treatment displayed an upward trend. Table 2 shows that the mean DPR score during baseline for Chris was 35.43% and this score increased to 76.29% during treatment. The standardized mean difference was 6.53, which supports the effectiveness of the intervention.

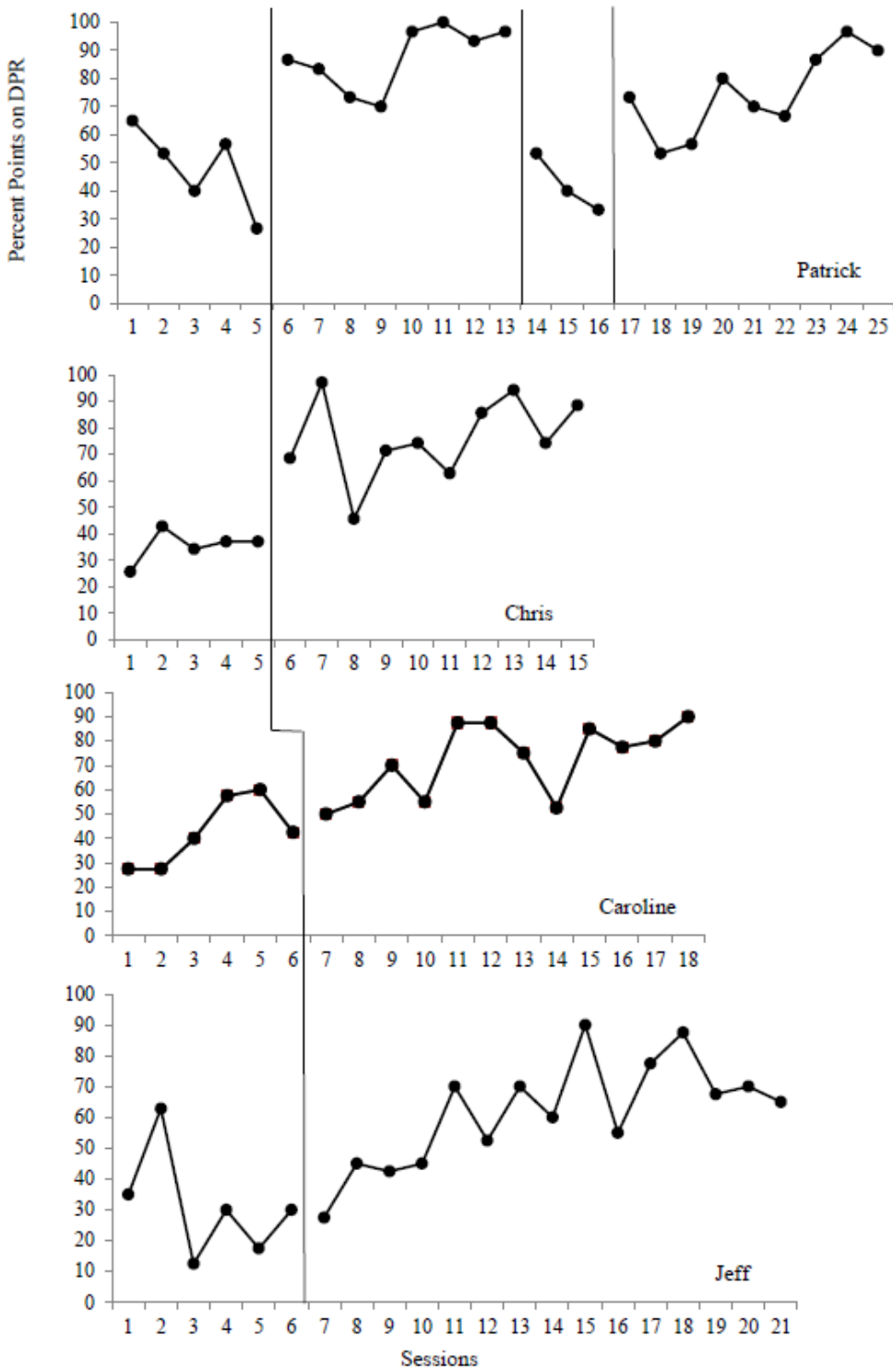


Figure 1. Graphs of the Daily Progress Report Data for Each Participant by Session.

Table 2. Direct Behavior Ratings for All Participants. Mean DPR scores for each participant and combined across all are shown in baseline and treatment. The change and standardized mean difference are presented as well. Standardized mean difference can be interpreted based on Cohen’s d interpretation guidelines (Cohen, 1988): <.20 is a small effect, .50 is a medium effect, and >.80 is a large effect.

Total DPR Percentage	Baseline Mean	Treatment Mean	Change	Standardized Mean Difference (d)*
Patrick	48.33%	80.78%	32.45%	2.15
Chris	35.43%	76.29%	40.86%	6.53
Caroline	42.50%	72.08%	29.58%	2.11
Jeff	31.31%	61.67%	30.36%	1.72
Combined**	39.39%	72.71%	33.32%	4.42

*Cohen’s d: (treatment mean-baseline mean)/baseline standard deviation (Cohen, 1988)

**Calculated using the mean of means and standard deviation of means across participants (Shadish, Rindskopf, & Hedges, 2008)

Visual analysis of Caroline’s data (Figure 1) indicates the intervention was effective in increasing her scores on the DPR. Caroline’s exhibited an increase to almost 60% in baseline in the absence of treatment and an immediate effect was not exhibited upon instating the intervention. However, her scores displayed a large increase in Session 11 that was maintained for the remainder of the intervention, with the exception of Session 14, which was a day where teacher integrity was low. As shown in Table 2, the mean DPR score for Caroline during baseline was 42.50% and the mean DPR score during treatment was 72.08%. The standardized mean difference was 2.11. Taken all together, Caroline’s results suggest that the intervention was effective in decreasing her internalizing behaviors as shown by an increase in her pro-social behaviors.

Jeff’s data are not as clear as the data of the preceding participants (refer to Figure 1). Jeff’s baseline scores were variable; however the last four points suggested a pattern of consistently low DPR scores heading into treatment. Upon treatment implementation, the effect

was not immediate. The DPR scores showed an upward trend in treatment but the scores were variable and did not suggest a clear pattern. Jeff's treatment ended on session 21 due to teacher request. Table 2 displays that Jeff's DPR score had a baseline mean of 31.31% and a treatment mean of 61.67% with a standardized mean difference of 1.72. Although this score is less than that of the other participants, it still suggests that the treatment was effective for Jeff. Although variable, his score did increase during treatment, suggesting that his internalizing behaviors were reduced via the increase in prosocial replacement behaviors.

The combined results from all participants are also presented in Table 2. The combined baseline mean for the DPRs was 39.39% and the combined treatment mean was 72.71%, with a standardized mean difference score of 4.42. Additionally, a PAND score was calculated for the total study. There were 10 overlapping data points out of 79 total data points, resulting in a PAND score of 87.34%, which falls in the effective range based on the criteria of Scruggs and Mastropieri (1998). The integration of all of the DPR results suggests that the intervention was effective across participants.

Direct Behavior Ratings (DBRs)

Teacher- and Student-Completed DBRs. Each participant's DBR results are presented by item for both teacher and student ratings in Figures 4-7. DBR items could be rated on a scale from 1-9, with higher scores representing a greater level of prosocial replacement behavior performance. The results from Patrick's DBRs are presented in Table 3. On the teacher-completed DBR, the first item, *responded to direction and began work within 5 seconds of being told to*, had a baseline mean of 4.80, a treatment mean of 6.00, and a standardized mean difference score of 0.67. On the student-completed DBR, this item had a baseline mean of 8.00, a treatment mean of 8.41, and a standardized mean difference score of 0.58. Although the student

Table 3. Direct Behavior Ratings for Subject 1. Patrick’s mean DBR scores for each item shown for baseline and treatment for both the teacher and student ratings. The change and standardized mean difference are presented as well. Standardized mean difference can be interpreted based on Cohen’s d interpretation guidelines (Cohen, 1988): <.20 is a small effect, .50 is a medium effect, and >.80 is a large effect.

Patrick: DBR Items (1-9)	Teacher				Student			
	Baseline Mean	Treatment Mean	Change	Standardized Mean Difference (d)*	Baseline Mean	Treatment Mean	Change	Standardized Mean Difference (d)*
Responded to directions and began work within 5 seconds of being told to	4.80	6.00	1.20	0.67	8.00	8.41	0.41	0.58
Paid attention in class and focused on work	2.80	5.18	2.38	2.83	9.00	7.76	-1.24	**
Looked at adults when having a conversation with them	4.00	6.89	2.89	4.07	8.00	8.59	0.59	0.72

*Cohen’s d: (treatment mean-baseline mean)/baseline standard deviations (Cohen, 1988)

**Could not be computed due to having a baseline standard deviation of 0.00

DBR scores were higher, the standardized mean differences were comparable, both suggesting a medium effect size for this item. The second item, *paid attention in class and focused on work*, had a teacher-completed baseline mean of 2.80, treatment mean of 5.18, and standardized mean difference of 2.83. On the student-completed DBR, this item had a baseline mean of 9.00 and a treatment mean of 7.76. A standardized mean difference could not be completed for this item as the baseline standard deviation was zero. However, the student-completed DBR suggests that the intervention had a negative effect on paying attention in class, which is contradicted by the teacher-completed DBR, which suggests a high positive effect. The final item, *looked at adults*

when having a conversation with them, had a teacher-completed baseline mean of 4.00, treatment mean of 6.89, and standardized mean difference of 4.07. The student-completed DBR for this item had a baseline mean of 8.00, a treatment mean of 8.59, and standardized mean difference of 0.72. The teacher DBR suggests a higher effect of the intervention on eye contact than does the child DBR, but both suggest a positive effect.

The results from Chris' teacher and student DBRs are presented in Table 4. For the first item, *looked at adults when having a conversation with them*, the teacher-completed DBR had a baseline mean of 4.40, a treatment mean of 6.00, and a standardized mean difference of 1.05. The child DBR for this item had a baseline mean of 8.40, a treatment mean of 8.20, and a standardized mean difference of -0.22. For this item, the teacher DBR suggests a high positive effect, while the child DBR suggests a low negative effect. The second item, *communicated appropriately and used an appropriate voice*, had a teacher-completed baseline mean of 4.40, treatment mean of 5.90, and standardized mean difference of 0.72. The student DBR for this item had a baseline mean of 6.40, treatment mean of 7.90, and standardized mean difference of 0.99. In this case, both DBRs suggest the intervention was effective in increasing appropriate communication, with the child DBR suggesting a slightly stronger effect. The third item, *only wrote one thought in though journal per subject*, had a teacher-completed baseline mean of 4.00, treatment mean of 7.80, and standardized mean difference of 1.27. For this item, the child DBR had a baseline mean of 5.60, a treatment mean of 8.50, and a standardized mean difference of 1.32. The DBRs for this item both suggest a strong intervention effect. The last item was *paid attention in class and completed classwork*. The baseline mean for the teacher-completed DBRs for this item was 2.20, the treatment mean was 4.20, and the standardized mean difference was

Table 4. Direct Behavior Ratings for Subject 2. Chris’s mean DBR scores for each item shown for baseline and treatment for both the teacher and student ratings. The change and standardized mean difference are presented as well. Standardized mean difference can be interpreted based on Cohen’s d interpretation guidelines (Cohen, 1988): <.20 is a small effect, .50 is a medium effect, and >.80 is a large effect.

Chris: DBR Items (1-9)	Teacher				Student			
	Baseline Mean	Treatment Mean	Change	Standardized Mean Difference (d)*	Baseline Mean	Treatment Mean	Change	Standardized Mean Difference (d)*
Looked at adults when having a conversation with them	4.40	6.00	1.60	1.05	8.40	8.20	-0.20	-0.22
Communicated appropriately and used an appropriate voice	4.40	5.90	1.50	0.72	6.40	7.90	1.50	0.99
Only wrote one thought in thought journal per subject	4.00	7.80	3.80	1.27	5.60	8.50	2.90	1.32
Paid attention in class and completed classwork	2.20	4.20	2.00	2.38	8.25	8.30	0.05	0.10

*Cohen’s d: (treatment mean-baseline mean)/baseline standard deviations (Cohen, 1988)

2.38. On the student-completed DBR, the baseline mean was 8.25, the treatment mean was 8.30, and the standardized mean difference was 0.10. On this item, the teacher DBR suggests a strong positive effect, while the student DBR suggests a small positive effect.

Caroline’s DBR scores are listed in Table 5. For the first item, *performed group role responsibilities*, the baseline mean for the teacher DBR was 4.33, the treatment mean was 4.91, and the standardized mean difference was 1.41. The child DBR for this item had both a baseline and treatment mean of 8.83, thus have a standardized mean difference of 0.00. For this item, the teacher DBR suggests the intervention had a strong effect, while the child DBR suggests it had

no effect. The second item, *spoke clearly and loudly enough to be heard when participating in class or talking to others*, had a baseline mean of 3.00, a treatment mean of 4.09, and a standardized mean difference of 2.66 for the teacher-completed DBR. For the student-completed DBR, this item had a baseline mean of 7.83, a treatment mean of 8.50, and a standardized mean difference of 1.63. Both DBRs suggest that the intervention had an effect on increasing the clarity and loudness of Caroline's speech. For the third item, *made eye contact with others when interacting with them*, the teacher-completed DBR had a baseline mean of 1.83, a treatment mean of 4.09, and a standardized mean difference of 1.39. The child-completed DBR for this item had a baseline mean of 8.33, a treatment mean of 8.75, and a standardized mean difference of 0.26. Both DBRs suggest an increase in eye contact during treatment, but the teacher DBR suggests a larger effect than does the child DBR. For the teacher-completed DBR, the final item, *paid attention in class and completed classwork*, had a baseline mean of 3.33, a treatment mean of 5.18, and a standardized mean difference of 2.20. For the child-completed DBR, the baseline mean was 8.50, the treatment mean was 8.83, and the standardized mean difference was 0.39. Similarly to the third item, for the fourth item both DBRs suggest a positive effect of treatment, with the teacher DBR suggesting a larger effect than the child DBR.

The results of Jeff's DBR are presented in Table 6. For the teacher-completed DBR, the first item, *did all written work neatly and tried best to write in cursive*, had a baseline mean of 2.17, a treatment mean of 3.40, and a standardized mean difference of 0.84. The child DBR had a baseline mean of 6.20, a treatment mean of 6.67, and a standardized mean difference of 0.56 for this item. The teacher and student DBRs are comparable for this item, both suggesting a medium-to-large positive effect. The second item, *responded appropriately and used an appropriate voice when something went wrong*, had a baseline mean of 2.67, a treatment mean of

Table 5. Direct Behavior Ratings for Subject 3. Caroline’s mean DBR scores for each item shown for baseline and treatment for both the teacher and student ratings. The change and standardized mean difference are presented as well. Standardized mean difference can be interpreted based on Cohen’s d interpretation guidelines (Cohen, 1988): <.20 is a small effect, .50 is a medium effect, and >.80 is a large effect.

Caroline: DBR Items (1-9)	Teacher				Student			
	Baseline Mean	Treatment Mean	Change	Standardized Mean Difference (d)*	Baseline Mean	Treatment Mean	Change	Standardized Mean Difference (d)*
Performed group role responsibilities	4.33	4.91	0.58	1.41	8.83	8.83	0.00	0.00
Spoke clearly and loudly enough to be heard when participating in class or talking to others	3.00	4.09	1.09	1.09	7.83	8.50	0.67	1.63
Made eye contact with others when interacting with them	1.83	4.09	2.26	2.26	8.33	8.75	0.42	0.26
Paid attention in class and completed classwork	3.33	5.18	1.85	1.85	8.50	8.83	0.33	0.39

*Cohen’s d: (treatment mean-baseline mean)/baseline standard deviations (Cohen, 1988)

5.56, and a standardized mean difference of 2.81 for the teacher-completed DBR. For the student-completed DBR, the baseline mean was 5.80, the treatment mean was 7.47, and the standardized mean difference was 1.02. Both DBRs suggest a positive effect of the intervention on this item. For the teacher-completed DBR, the third item, *was ready for the next task quickly and had materials ready*, had a baseline mean of 2.83, a treatment mean of 4.81, and a standardized mean difference of 1.35. For the student-completed DBR, the baseline mean was 7.00, the treatment mean was 7.20, and the standardized mean difference was 0.10. The teacher DBR suggests a positive effect of the intervention on ability to transition, while the child DBR suggests a small, close to null, effect. The last item, *paid attention in class and completed*

Table 6. Direct Behavior Ratings for Subject 4. Jeff’s mean DBR scores for each item shown for baseline and treatment for both the teacher and student ratings. The change and standardized mean difference are presented as well. Standardized mean difference can be interpreted based on Cohen’s d interpretation guidelines (Cohen, 1988): <.20 is a small effect, .50 is a medium effect, and >.80 is a large effect.

Jeff: DBR Items (1-9)	Teacher				Student			
	Baseline Mean	Treatment Mean	Change	Standardized Mean Difference (d)*	Baseline Mean	Treatment Mean	Change	Standardized Mean Difference (d)*
Did all written work neatly and tried best to write in cursive	2.17	3.40	1.23	0.84	6.20	6.67	0.47	0.56
Responded appropriately and used an appropriate voice when something went wrong	2.67	5.56	2.89	2.81	5.80	7.47	1.67	1.02
Was ready for the next task quickly and had materials ready	2.83	4.81	1.98	1.35	7.00	7.20	0.20	0.10
Paid attention in class and completed classwork	2.50	4.19	1.69	1.22	6.60	7.73	1.13	0.56

*Cohen’s d: (treatment mean-baseline mean)/baseline standard deviations (Cohen, 1988)

classwork, had a teacher-completed baseline mean of 2.50, treatment mean of 4.19, and standardized mean difference of 1.22. The child DBR for this item had a baseline mean of 6.60, a treatment mean of 7.73, and a standardized mean difference of 0.56. Both DBRs suggest that the intervention had a positive effect on this item, with the teacher DBR suggesting a higher positive effect than the child DBR.

Correlations. For the majority of the DBR items, the teacher-completed DBRs suggested a higher positive effect than did the child-completed DBRs. As shown in Table 7, the correlations between the teacher- and student-completed DBRs ranged from an r^2 of .02 to .11. According to Cohen (1988), an r^2 between .01 and .09 is considered small, between .09 and 0.25 is considered medium, and greater than .25 is considered large. The student-completed DBRs and teacher-completed DPRs also had small to medium correlations, with r^2 estimates ranging from .02 to .15. These correlations indicate that the teachers and students did not have a high agreement as to the degree to which the replacement behaviors were being performed, both in baseline and in intervention. The two teacher-completed measures, the DPR and the DBR correlated highly, ranging from an r^2 of .30 to .59. This high correlation suggests that multiple teacher ratings throughout the day correspond well to an overall daily impression given by one single rating at the end of the day.

Rating of Feeling. Table 8 depicts the results of the feelings ratings for each participant and averaged across participants. Patrick rated his feeling a mean of 7.60 during baseline and 7.12 during treatment, with a standardized mean difference of -0.23. According to Cohen’s guidelines, this score suggests that the treatment had a small negative effect on Patrick’s rating of feeling. For Chris, his feeling rating had a baseline mean of 7.20, a treatment mean of 4.00, with

Table 7. Correlations Among the Outcome Measures. Correlations between the DPRs and DBRs for each participant are presented. R is followed by r^2 . Interpretation of r^2 can be based on Cohen’s guidelines (1988): .01-.09 = small, .09-.25 = medium, and $>.25$ = large.

Correlation Table	DPRs and Teacher DBRs r (r^2)	DPRs and Student DBRs r (r^2)	Teacher and Student DBRs r (r^2)
Patrick	.77 (.59)	.14 (.02)	.23 (.05)
Chris	.55 (.30)	.21 (.04)	.15 (.02)
Caroline	.74 (.55)	.39 (.15)	.30 (.09)
Jeff	.75 (.56)	.34 (.12)	.33 (.11)

a standardized mean difference of -0.91, suggesting that the treatment had a large negative effect on his rating of feeling. Caroline had a mean feeling rating of 8.50 during baseline and 10.00 during treatment. The standardized mean difference for these scores is 2.11, suggesting that the treatment had a large positive effect on the level of feeling as rated by Caroline. The final participant, Jeff, had mean baseline rating of feeling of 7.20, a mean treatment rating of 7.07, with a standardized mean difference of -0.15. This score indicates that the treatment had a small negative effect on Jeff’s rating of feeling. Combined across all the participants, the baseline mean rating of feeling was 7.63 with a treatment mean rating of 7.05. The standardized mean difference for these ratings is -0.95, implying that the treatment had a large negative effect overall on the rating of feeling by the participants.

Post Scores on the Student Internalizing Behavior Screener (SIBS)

Table 9 depicts the pre- and post-scores of the participants on the SIBS. According to the developers of the screener, scores equal to 8 or above indicate that the participant is at-risk for

Table 8. Student Rating of Feelings. Mean rating of feeling for each participant is shown for baseline and treatment. The change and standardized mean difference are presented as well. Standardized mean difference can be interpreted based on Cohen’s d interpretation guidelines (Cohen, 1988): <.20 is a small effect, .50 is a medium effect, and >.80 is a large effect.

Feelings Rating (1-10)	Baseline Mean	Treatment Mean	Change	Standardized Mean Difference (d)*
Patrick	7.60	7.12	-0.48	-0.23
Chris	7.20	4.00	-3.20	-0.91
Caroline	8.50	10.00	1.50	2.11
Jeff	7.20	7.07	-0.13	-0.15
Combined**	7.63	7.05	-0.58	-0.95

*Cohen’s d: (treatment mean-baseline mean)/baseline standard deviations (Cohen, 1988)

**Calculated using the mean of means and standard deviation of means across participants (Shadish, Rindskopf, & Hedges, 2008)

Table 9. Pre- and Post-Scores on the SIBS. Pre- and post-scores on the SIBS are presented for each participant. Scores are out of a possible 21 points. Scores above 8 represent students at-risk for internalizing behavior problems (Cook, in press).

Student Internalizing Behavior Screener (SIBS)	Pre	Post	Change
Patrick	9	9	0
Chris	16	10	6
Caroline	17	6	9
Jeff	14	8	6

the development of internalizing behavior problems. Patrick showed no change on the screener from pre- to post-intervention, as he scored a 9 both times. Chris scored a 16 on the screener pre-intervention and a 10 on the screener post-intervention. While this score still qualifies him as being at-risk following the intervention, it does represent a decrease from his pre-treatment score. Caroline score on the screener went from a 17 pre-intervention to a 6 post-intervention. This change put her no longer at-risk for the development of internalizing behaviors according to this screener following treatment. The final participant, Jeff had a pre-intervention score of 14 and a post-intervention score of 8. While this score still puts him at-risk for the development of internalizing behaviors based on the cut-off of the screener, it represents a decreased score and is very close to being no longer at-risk. Overall, the participants showed a decrease in their score on the SIBS following treatment; however, most participants still remained at-risk.

Post Scores on the Social Skills Improvement System: Rating Scales (SSIS: RS)

Pre- and post-scores on the SSIS: RS Internalizing Scale and Social Skills Scale are presented in Table 10. Scores on the Internalizing Scale are presented as raw scores with an accompanying indicator, as well as mean and standard deviation for the norm sample of the as standard scores, along with their accompanying indicators.

Table 10. Pre- and Post-Scores on the SSIS:RS. Pre- and post-scores on the SSIS: RS are presented for each participant. For the Internalizing Scale, the scores are presented as raw scores that can be grouped into either below average, average, or above average. A decrease in score represents an improvement and decrease in internalizing behaviors as rated by the teacher. The norm sample means and standard deviations are presented as well for the corresponding age and gender for each participant. For the Social Skills Scale, scores are presented as standard scores. The standard scores have a mean of 100 and a standard deviation of 15. Scores below 70 are considered well-below average, 70-85 below average, 85-115 average, 115-130 above average, and above 130 well-above average (Gresham & Elliot, 2008).

Social Skills Improvement System: Rating Scale (SSIS:RS)	Internalizing Scale raw score (indicator)			Social Skills Scale standard score (indicator)	
	Pre	Post	Norm Sample mean (sd)	Pre	Post
Patrick	9 (Average)	0 (Average)	5.2 (4.3)	78 (Below Average)	89 (Average)
Chris	19 (Above Average)	13 (Above Average)	3.6 (3.4)	83 (Below Average)	97 (Average)
Caroline	19 (Above Average)	4 (Average)	3.2 (2.6)	57 (Well-below average)	100 (Average)
Jeff	11 (Above Average)	8 (Above Average)	2.7 (2.1)	64 (Well-below average)	75 (Below Average)

Patrick's raw score on the Internalizing Scale pre-intervention was a 9 and his score post-intervention was a 0. Both scores put him in the *average* range compared to the normative sample, which has a mean of 5.2 and a standard deviation of 4.3. His Social Skills Scale standard score went from a 78 pre-intervention to an 89 post-intervention. This change represents switching from the *below average* range to the *average* range in social skills as rated by his teacher.

Before the intervention, Chris scored a 19 on the Internalizing Scale, placing him in the *above average* range. Upon completion of the intervention, Chris obtained a score of 13, which

still placed him in the *above average* range when compared to the normative sample, which has a mean of 3.6 and standard deviation of 3.4. His score on the Social Skills Scale was an 83 before the intervention and a 97 after the intervention. This change is almost one standard deviation and represents a change from the *below average* range to the *average* range, showing an improvement in social skills as rated by his teacher.

Caroline scored a 19 on the Internalizing Scale as rated by her teacher pre-intervention and a 4 post-intervention. This change moved her from being in the *above average* range before the intervention to the *average* range after the intervention, as the normative sample has a mean of 3.2 and a standard deviation of 2.6. On the Social Skills Scale, Caroline obtained a score of 57 before the intervention as rated by her teacher, placing her in the *well-below average* range. After the intervention, she obtained a score of 100 on this scale, representing over a standard deviation increase in social skills as rated by her teacher and movement into the *average* range.

On the Internalizing Scale, Jeff scored an 11 pre-intervention as rated by his teacher and an 8 post-intervention. These scores both fall in the *above-average* range when compared to the normative sample, which has a mean of 2.7 and a standard deviation of 2.1. On the Social Skills Scale, Jeff's teacher rated him in the *well-below average* range pre-intervention, scoring a 64. After the intervention he scored a 75, placing him in the *below average* range. While this is still below the average range of social skills as compared to the normative sample, it represents an improvement from his score pre-intervention.

DISCUSSION

The purpose of this study was to assess the effectiveness of CICO as a targeted intervention to address internalizing behavior problems in schools. The intervention's effectiveness in reducing internalizing behavior problems through the increase of prosocial replacement behaviors was assessed through the use of a DPR, teacher and student DBRs, a student rating of feeling and pre- and post-scores on the SIBS and selected scales of the SSIS: RS. A discussion of the results of this research follows.

Cheney et al. (2009) suggested that using CICO could be an effective method for decreasing internalizing behaviors in students. The results from the visual analysis and the effect size estimates of the DPRs in this study support Cheney's speculation. Based on a comparison of baseline and treatment DPR ratings, the students' prosocial replacement behaviors increased consistently as a result of treatment. This effect was greater for the first two participants, Patrick and Chris. There are few possible reasons for this differential effect, including that Patrick and Chris were older than the other two participants, possibly increasing their ability to understand and challenge their cognitions. Also, Patrick and Chris participated in the intervention in the second half of the school year. As the "invisibility" of students with internalizing behavior problems has been well-documented (Gresham & Kern, 2004; Merrell & Gueldner, 2010), it is possible that these participants were more accurately screened for participation in the study compared to the final two participants, who were nominated only a few weeks into the school year. In fact, the authors posit that Jeff may not have been effectively screened for this intervention, possibly representing a case where symptoms of anxiety were confounded with symptoms of inattention. Despite the more apparent effect for the first two participants, the intervention was still moderately effective for the last two participants and overall, shows

promise to be used for increasing prosocial replacement behaviors in children displaying internalizing behaviors in schools.

Results of the DBRs were variable, as shown by varying levels of standardized mean difference estimates; however, most of the estimates suggested a positive effect. Overall, the teacher DBRs suggested a higher level of effect than did the student DBRs. This difference was due partially to the fact that the students tended to rate themselves significantly higher during baseline than did the teachers. When analyzing the correlations, the teacher and student DBRs had a small to medium-correlation (refer to Table 6), showing that often the teacher and student did not agree on the level of performance of the prosocial replacement behaviors. This lack of agreement between teacher and student ratings is documented extensively in the literature (Achenbach, McConaughy, & Howell, 1987; De Los Reyes & Kazdin, 2005). Although it cannot be said with complete certainty which measure is more accurate, based on established deficiencies with child self-ratings, the teacher DBRs have been considered the criterion measure in this study, suggesting that the intervention was effective.

Additional support for the effectiveness of the intervention comes from the pre- and post-scores on the SIBS and SSIS: RS. When comparing pre- and post-scores on the SIBS, it is apparent that for three of the four participants, there was a large drop in their rating after the intervention compared to their rating before the intervention (refer to Table 9). However, one participant, Patrick, did not show a change in the rating after participating in the intervention and based on the suggested cut-off for the screener, three of the participants remained at-risk for internalizing behavior problems post-intervention. The pre- and post-scores on the SSIS: RS show a clearer effect of the intervention (see Table 10). When comparing pre- and post-scores on the Internalizing Scale, all participants showed a significant drop in their score, showing that

they exhibited less internalizing behaviors according to teacher rating after completion of the intervention. Patrick's score dropped to a raw score of zero, suggesting no display of internalizing symptoms and Caroline went from being in the *above average* range to *average* following the intervention. Chris and Jeff remained in the *above average* range, but both dropped over a standard deviation based on the normative sample for their age and gender. All of the participants' ratings on the Social Skills Scale improved following the intervention, with each participant moving up a range, showing clinically significant improvement. Three of the four participants ended the intervention rated in the *average* range for social skills based on their teachers' responses. These results show not only that the intervention was effective in decreasing internalizing behaviors, but that it had an effect on improving overall social skills as shown by the teachers ratings on the SSIS: RS.

Effect size estimates of the child rating of feeling potentially indicate that the CICO intervention did not have a positive effect on students' mood (refer to Table 8). Caroline is the only participant who showed an improvement in her rating of mood during treatment compared to baseline. Patrick and Jeff displayed a small decrease in mood during treatment when compared to baseline, while Chris showed a large decrease. The combination of all participants' ratings shows a significantly large decrease in rating of feeling from baseline compared to treatment; however, this effect is primarily driven by Chris' ratings. It cannot be said for certain what aspect of the intervention drove this effect. It is possible that the students actually felt worse during treatment than they did during baseline. However, given the unreliability of student self-ratings when compared to adult ratings and observations made by the mentors, this seems unlikely. It is possible that the intervention made the participants more aware of their feelings or that after becoming comfortable with the mentor, the participants were more likely to rate their

feelings honestly, sometimes providing a lower rating. The lowered rating in treatment compared to baseline also could have been driven by attention-seeking behavior, particularly with Chris. Based on mentor report, Chris was more likely to openly display upset feelings compared to the other participants and was given attention from the mentor when he expressed these feelings. While the mentor was following the intervention protocol, it is possible that Chris began to display upset feelings and give lower ratings of feeling as a way to spend more time with the mentor, thereby getting more of the mentor's attention. The reason behind the lowered rating can only be speculated at this point and warrants further investigation in future research.

Another important finding of this study is that the CICO intervention was considered acceptable both pre- and post-implementation by teachers for the reduction of internalizing behaviors. This is encouraging given that the intervention has not been used previously to target this type of behavior. Additionally, the intervention was able to be implemented with high integrity. Considering the effects of the intervention were positive, the intervention was considered acceptable, and the intervention was able to be implemented with integrity, this study suggests that CICO could be a useful targeted intervention for reducing internalizing behaviors in schools.

Limitations and Future Research

Although the results suggest that the intervention was generally effective, some limitations of the study are worth discussing and investigating further. Due to the nature of the intervention, limited time was able to be spent on targeting cognitions during the check-in and check-out procedures. This limitation was partially due to the fact that time did not always allow for extended focus on cognitions. The morning is a busy time in schools and the mentor had to check-in with the participants after they ate breakfast and before they went to class, which often

left only a couple of minutes to devote to check-in. Additionally, not all of the participants were readily aware of their cognitions. The first two participants, Patrick and Chris, were able to discuss their cognitions without much difficulty but this was not the case with the final two participants, Caroline and Jeff. This difference could be due to the fact that Patrick and Chris were older than Caroline and Jeff. It would be worthwhile to study the effectiveness of CICO for the reduction of internalizing behaviors in a slightly older age group. It also would be useful to implement the intervention with a set time for check-in, pulling the students out of class for a limited amount of time, such as 10 minutes, in the beginning of the day in order to devote more time to the discussion and challenging of negative or faulty cognitions. Additionally, creating and testing a protocol that non-psychologist mentors could implement would be helpful and would make the intervention more easily integrated into a school setting with any adult faculty member able to serve as the mentor.

Related to the previous limitation, various components of the intervention were not assessed independently. Thus, it is unclear whether the relationship and time spent with the mentor drove the effect of the intervention or the continuous positive feedback throughout the day from the teacher. It is possible that the combination of the two primary components is what makes the intervention effective. A conclusion cannot be drawn without further research, which would be helpful in providing insight into the important aspects of the intervention. If both components are not necessary for the intervention to be effective, it is possible that the intervention could be done using fewer resources.

Another limitation of this study was that the screening measures and outcome measures consisted solely of rating scales. While widely used, rating scales are subject to bias and differing levels of teacher tolerance (Elliot, Busse & Gresham, 1993). It may be possible that

some students were not screened effectively for participation in this study due to the use of these ratings scales. Additionally, the DPR and DBR inter-individual and daily intra-individual scores could have been subject to teacher perception and may not always have been purely indicative of performance of the target replacement behaviors. While direct observation would have been helpful in solving this problem, it could not be conducted for this study due to teacher request for Jeff and Caroline and reactivity for Chris. Observations were attempted for Patrick; however, the behaviors were difficult to observe given a limited observation sample, and thus the observation data was not useful. Finding a way to directly measure these behaviors is an area that merits more research. Further, it would be helpful to measure the internalizing behaviors in addition to the prosocial replacement behaviors in order to provide a more complete measure of the effectiveness of the treatment.

Though treatment integrity data was collected, the quality of intervention implementation was not recorded. Based on anecdotal observation of the mentors, different teachers implemented the intervention with varying quality, specifically in regards to the type and quality of positive feedback given for performing the behaviors. Additionally, within teacher variance in quality of implementation was noted depending on the attitude of the teacher or activity taking place in the classroom. The quality of the mentor interactions most likely varied from day to day as well. In the future, providing some sort of measure of quality of intervention implementation would provide useful information as to the quality, and therefore integrity, of the intervention. It may be possible that student performance varies based on quality of intervention implementation but this connection can only be speculated until more research has been conducted.

Another limitation of this study is that generalization and maintenance data were not collected. Without generalization data, it is unclear whether the effects of the intervention spread

to other settings or behaviors. It would provide useful information to have a measure of the prosocial replacement behaviors taken in a different setting and by a different observer.

Maintenance data was unable to be collected for Patrick or Chris given that their intervention lasted until the end of the school year. It was unable to be collected for Caroline and Jeff due to the fact that the researcher continued to work with the participants after the intervention; thus, a maintenance probe would not have provided a true measure of whether the effects of the intervention maintained when it was withdrawn. Generalization and maintenance data should be collected in future studies on this topic.

Another topic that deserves further research is the ratings of feelings. Some possible explanations for the results of the rating were provided previously in the discussion section. More research could help shed light on this finding. Additionally, running this study in a group design would be useful in order to provide a more meaningful estimate of effect size and to provide an indicator of external validity to a larger population. Finally, it would be useful to attempt to determine the function of the internalizing behaviors. If the function could be determined, the intervention could possibly be tweaked to match the function, which may increase its effectiveness.

CONCLUSIONS

The results of this study indicate that CICO is an effective intervention for reducing internalizing behaviors through the increase in prosocial replacement behaviors gained through receiving feedback on these behaviors throughout the day and through the mentor relationship. These components should be studied individually in the future to determine which component is the driving force, or if the combination of the components is an important feature in itself. Check-in/Check-out has promise as an easy-to-implement, acceptable, and effective school-based intervention for the treatment of internalizing behavior problems.

REFERENCES

- Achenbach, T., & Edelbrock, C. (1978). The classification of child psychopathology: A review and analysis of empirical efforts. *Psychological Bulletin*, 85(6), 1275-1301. doi:10.1037/0033-2909.85.6.1275.
- Achenbach, T.M., McConaughy, S.H., & Howell, C.T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, 101(2), 213-232.
- Achenbach, T.M. & Rescorla, L.A. (2001). *Manual for the ASEBA School-Age Forms & Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.
- Bradshaw, C., Buckley, J., & Ialongo, N. (2008). School-based service utilization among urban children with early onset educational and mental health problems: The squeaky wheel phenomenon. *School Psychology Quarterly*, 23(2), 169-186. doi:10.1037/1045-3830.23.2.169.
- Cartwright-Hatton, S., Roberts, C., Chitsabesan, P., Fothergill, C., & Harrington, R. (2004). Systematic review of the efficacy of cognitive behavior therapies for childhood and adolescent anxiety disorders. *British Journal of Clinical Psychology*, 43(4), 421-436. doi:10.1348/0144665042388928.
- Cheney, D., Flower, A., & Templeton, T. (2008). Applying response to intervention metrics in the social domain for students at risk of developing emotional or behavioral disorders. *The Journal of Special Education*, 42(2), 108-126. doi:10.1177/0022466907313349.
- Cheney, D., Stage, S., Hawken, L., Lynass, L., Mielenz, C., & Waugh, M. (2009). A 2-year outcome study of the check, connect, and expect intervention for students at risk for severe behavior problems. *Journal of Emotional and Behavioral Disorders*, 17(4), 226-243. doi:10.1177/1063426609339186.
- Christ, T.J., Riley-Tillman, C.T., & Chafouleas, S.M. (2009). Foundation for the development and use of Direct Behavior Rating (DBR) to assess and evaluate student behavior. *Assessment for Effective Intervention*, 34(4), 201-213. doi:10.1177/1534508409340390.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Erlbaum Associates, Publishers, Hillsdale, NJ.
- Colvin, G., Sugai, G., Good, R., & Lee, Y. (1997). Using active supervision and precorrection to improve transition behaviors in an elementary school. *School Psychology Quarterly*, 12(4), 344-363. doi:10.1037/h0088967.

- Compton, S., Burns, B., Egger, H., & Robertson, E. (2002). Review of the evidence base for treatment of childhood psychopathology: Internalizing disorders. *Journal of Consulting and Clinical Psychology, 70*(6), 1240-1266. doi:10.1037/0022-006X.70.6.1240.
- Cook, C. (in press). *Student Internalizing Behavior Screener*. Baton Rouge, LA: Louisiana State University.
- Coyle, J., Pine, D., Charney, D., Lewis, L., Nemeroff, C., Carlson, G., et al. (2003). Depression and Bipolar Support Alliance Consensus Statement on the Unmet Needs in Diagnosis and Treatment of Mood Disorders in Children and Adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry, 42*(12), 1494-1503.
- Crone, D.A., Hawken, L.S., & Horner, R.H. (2010). *Responding to problem behavior in schools: The Behavior Education Program* (2nd ed.) New York, NY US: Guilford Press.
- Dadds, M., Spence, S., Holland, D., Barrett, P., & Laurens, K. (1997). Prevention and early intervention for anxiety disorders: A controlled trial. *Journal of Consulting and Clinical Psychology, 65*(4), 627-635. doi:10.1037/0022-006X.65.4.627.
- De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin, 131*(4), 483-509. doi:10/1037/0033-2909.131.4.483.
- De Pry, R., & Sugai, G. (2002). The effect of active supervision and pre-correction on minor behavioral incidents in a sixth grade general education classroom. *Journal of Behavioral Education, 11*(4), 255-267. doi:10.1023/A:1021162906622.
- Drummond, T. (1994). *The Student Risk Screening Scale (SRSS)*. Grants Pass, OR: Oregon Social Learning Center.
- Elliot, S.N., Busse, R.T., & Gresham, F.M. (1993). Behavior rating scales: Issues of use and development. *School Psychology Review, 22*(2), 313-321.
- Filter, K., McKenna, M., Benedict, E., Horner, R., Todd, A., & Watson, J. (2007). Check in/Check out: A post-hoc evaluation of an efficient, secondary-level targeted intervention for reducing problem behaviors in schools. *Education & Treatment of Children, 30*(1), 69-84. doi:10.1353/etc.2007.0000.
- Greenberg, M., Domitrovich, C., & Bumbarger, B. (2001). The prevention of mental disorders in school-aged children: Current state of the field. *Prevention & Treatment, 4*(1), doi:10.1037/1522-3736.4.1.41a.
- Gresham, F.M., & Cook, C. (in press). *Brief Behavior Rating Scales*. Baton Rouge, LA: Louisiana State University.

- Gresham, F.M., & Elliott, S.N. (2008). *Social Skills Improvement System: Rating Scales*. Minneapolis: Pearson Assessments.
- Gresham, F.M., & Kern, L. (2004). Internalizing behavior problems in children and adolescents. In R.B. Rutherford, M.M. Quinn, & S.R. Mathur (Eds.), *Handbook of research in emotional and behavioral disorders* (pp. 262-281). New York, NY: Guilford Press.
- Hawken, L., & Hess, R. (2006). School psychologists as leaders in the implementation of a targeted intervention: The Behavior Education Program. *School Psychology Quarterly*, 21(1), 91-111. doi:10.1521/scpq.2006.21.1.91.
- Hawken, L., & Horner, R. (2003). Evaluation of a Targeted Intervention Within a Schoolwide System of Behavior Support. *Journal of Behavioral Education*, 12(3), 225-240. doi:10.1023/A:1025512411930.
- Hawken, L., MacLeod, K., & Rawlings, L. (2007). Effects of the Behavior Education Program on office discipline referrals of elementary school students. *Journal of Positive Behavior Interventions*, 9(2), 94-101. doi:10.1177/10983007070090020601.
- Herman, K.C. & Ostrander, R. (2007). The effects of attention problems on depression: Developmental, academic, and cognitive pathways. *School Psychology Quarterly*, 22(4), 483-510. doi: 10/1037/1045-3830.22.4.483.
- Kazdin, A.E. (1982). *Single-case research designs: Methods for clinical and applied settings*. New York, New York: Oxford University Press.
- Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M & Shadish, W. R. (2010). Singlecase designs technical documentation. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf.
- Kutash, K., Duchnowski, A.J., & Lynn, N. (2006). *School-based mental health: An empirical guide for decision-makers*. Retrieved November 2, 2010, from the University of South Florida, Research and Training Center for Children's Mental Health website: <http://rtckids.fmhi.usf.edu>.
- Lewis, T., Colvin, G., & Sugai, G. (2000). The effects of pre-correction and active supervision on the recess behavior of elementary students. *Education & Treatment of Children*, 23(2), 109-121. Retrieved from PsycINFO database.
- Maag, J., & Swearer, S. (2005). Cognitive-behavioral interventions for depression: Review and implications for school personnel. *Behavioral Disorders*, 30(3), 259-276. Retrieved from PsycINFO database.

- Merrell, K. (2008). *Behavioral, social, and emotional assessment of children and adolescents* (3rd ed.). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers. 746. doi:10.1037/0012-1649.41.5.733.
- Merrell, K.W. & Gueldner, B. A. (2010). Preventive interventions for students with internalizing disorders: Effective strategies for promoting mental health in schools. In M.R. Shinn & H.M. Walker (Eds.), *Interventions for achievement and behavior problems in a three-tier model including RTI* (pp. 799-824). Bethesda, MD: National Association of School Psychologists.
- Parker, R., Hagan-Burke, S., & Vannest, K. (2007). Percentage of all non-overlapping data (PAND): An alternative to PND. *The Journal of Special Education, 40*, 194-204.
- Parsonson, B., & Baer, D. (1978). The analysis and presentation of graphic data. In T. Kratchowill (Ed.) *Single Subject Research* (pp. 101–166). New York: Academic Press.
- Reinherz, H.Z., Tanner, J.L., Paradis, A.D., Beardslee, W.R., Szigethy, E.M., & Bond, A.E. (2006). Depressive disorders. In C. A. Essau (Ed.), *Child and adolescent psychopathology: Theoretical and clinical implications* (pp. 113-139). New York, NY: Routledge/Taylor & Francis Group.
- Riley-Tillman, C.T., Chafouleas, S.M., Sassu, K.A., Chanese, J.A.M., & Glazer, A.D. (2008). Examining the agreement of direct behavior ratings and systematic direct observation data for on-task and disruptive behavior. *Journal of Positive Behavior Interventions, 10*(2), 136-143. doi:10.1177/1098300707312542.
- Roeser, R., van der Wolf, K., & Strobel, K. (2001). On the relation between social–emotional and school functioning during early adolescence: Preliminary findings from Dutch and American samples. *Journal of School Psychology, 39*(2), 111-139. doi:10.1016/S0022-4405(01)00060-7.
- Scruggs, T.F., & Mastropieri, M.A. (1998). Summarizing single-subject research: Issues and applications. *Behavior Modification, 22*(3), 221-242. doi:10.1177/01454455980223001.
- Shadish, W.R., Rindskopf, D.M., & Hedges, L.V. (2008). The state of the science in the meta-analysis of single-case experimental designs. *Evidence Based Communication Assessment and Intervention, 188-196*. doi:10.1080/17489530802581603.
- Sourander, A., & Helstela, L. (2005). Childhood predictors of externalizing and internalizing problems in adolescence: A prospective follow-up study from age 8 to 16. *European Child & Adolescent Psychiatry, 14*(8), 415-423. doi:10.1007/s00787-005-0475-6.
- Sprick, R. & Borgmeier, C. (2010). Behavior prevention and management in three tiers in secondary schools. In M.R. Shinn & H.M. Walker (Eds.), *Interventions for achievement and behavior problems in a three-tier model including RTI* (pp. 435-468). Bethesda, MD: National Association of School Psychologists.

- Stark, K., Reynolds, W., & Kaslow, N. (1987). A comparison of the relative efficacy of self-control therapy and a behavioral problem-solving therapy for depression in children. *Journal of Abnormal Child Psychology: An official publication of the International Society for Research in Child and Adolescent Psychopathology*, *15*(1), 91-113. doi:10.1007/BF00916468.
- Todd, A., Campbell, A., Meyer, G., & Horner, R. (2008). The Effects of a Targeted Intervention to Reduce Problem Behaviors. *Journal of Positive Behavior Interventions*, *10*(1), 46-55. Retrieved from Academic Search Complete database.
- Vasa, R.A., & Pine, D.S. (2006). Anxiety disorders. In C. A. Essau (Ed.), *Child and adolescent psychopathology: Theoretical and clinical implications* (pp. 78-112). New York, NY: Routledge/Taylor & Francis Group.
- Walker, H., Horner, R., Sugai, G., & Bullis, M. (1996). Integrated approaches to preventing antisocial behavior patterns among school-age children and youth. *Journal of Emotional and Behavioral Disorders*, *4*(4), 194-209. doi:10.1177/106342669600400401.
- Walker, H.M. & Shinn, M.R. (2010). Systemic, evidence-based approaches for promoting positive student outcomes with a multitier framework: moving from efficacy to effectiveness. In M.R. Shinn & H.M. Walker (Eds.), *Interventions for achievement and behavior problems in a three-tier model including RTI* (pp. 1-26). Bethesda, MD: National Association of School Psychologists.
- Witt, J.C. & Elliott, S.N. (1985). Acceptability of classroom management strategies. In T.R. Kratochwill (Ed.), *Advances in school psychology (Vol 4, pp.251-288)*. Hillsdale, NJ: Lawrence Erlbaum.

APPENDIXES

APPENDIX A: STUDENT INTERNALIZING BEHAVIOR SCREENER (SIBS)

Student Internalizing Behavior Screener (SIBS)

(Cook, in press)

Directions: Please rate each student on each behavior using the following scale:

0=Never, 1=Rarely, 2=Occasionally, 3=Frequently

For each student, write the number that corresponds to the frequency rating in each cell.

Student Name	Nervous or fearful	Bullied by peers	Spends time alone	Clings to adults	Withdrawn	Seems sad or unhappy	Complains about being sick or hurt

APPENDIX B: STUDENT RISK SCREENING SCALE (SRSS)

Student Risk Screening Scale (SRSS)

(Drummond, 1994)

Directions: Please rate each student on each behavior using the following scale:

0=Never, 1=Rarely, 2=Occasionally, 3=Frequently

Student Name	Stealing	Lying, Cheating, Sneaking	Behavior Problems	Peer Rejection	Low Academic Achievement	Negative Attitude	Aggressive Behaviors

APPENDIX C: DIRECT BEHAVIOR RATING, TEACHER FORM

Direct Behavior Rating

Date:

Student : Caroline

Teacher: Ms. S.

School:

*Directions: Review each of the items below. For each item, rate the degree to which the student showed the behavior or met the behavior goal. Please rate the student's behavior for **today** only.*

1. Caroline performed her group role responsibilities.

1	2	3	4	5	6	7	8	9
Never/Seldom				Sometimes			Usually/Always	

2. Caroline spoke clearly and loud enough to be heard when participating in class/talking to others.

1	2	3	4	5	6	7	8	9
Never/Seldom				Sometimes			Usually/Always	

3. Caroline made eye contact with others when interacting.

1	2	3	4	5	6	7	8	9
Never/Seldom				Sometimes			Usually/Always	

4. Caroline was on-task (paying attention to and participating in classwork/activity).

1	2	3	4	5	6	7	8	9
Never/Seldom				Sometimes			Usually/Always	

APPENDIX D: DIRECT BEHAVIOR RATING, STUDENT FORM

Direct Behavior Rating—Student Form

Date:

Student: **Caroline**

Teacher: Ms. S.

School:

Directions: Read each of the items below. For each item, rate how much you showed that behavior today during school.

1. I performed my group role responsibilities.

1 2 3 4 5 6 7 8 9
 Never/Not A lot Sometimes Usually/Always

2. I spoke clearly and loudly enough to be heard when participating in class or talking to others.

1 2 3 4 5 6 7 8 9
 Never/Not A lot Sometimes Usually/Always

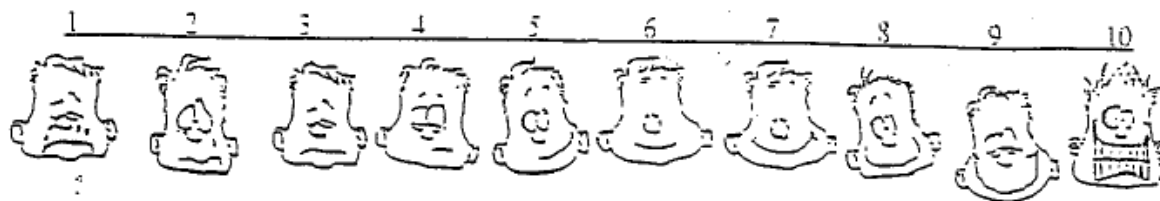
3. I made eye contact with others when interacting with them.

1 2 3 4 5 6 7 8 9
 Never/Not A lot Sometimes Usually/Always

4. I paid attention in class and participated in classwork/class activities.

1 2 3 4 5 6 7 8 9
 Never/Not A lot Sometimes Usually/Always

5. Circle the number that best describes how you feel today.



APPENDIX E: DAILY PROGRESS REPORT

Daily Progress Report

Name: **Caroline**

Date: _____

2 = Good
 1 = OK
 0 = Needs Improvement

Goal: _____
 Points earned: _____

Goal reached? Yes No

Goals	Morning Work	ELA	Math	Science	Social Studies
Perform group role responsibilities	2 1 0	2 1 0	2 1 0	2 1 0	2 1 0
Speak clearly and loud enough to be heard	2 1 0	2 1 0	2 1 0	2 1 0	2 1 0
Make eye contact when interacting with others	2 1 0	2 1 0	2 1 0	2 1 0	2 1 0
Is on-task (doing classwork/activity; and participating appropriately)	2 1 0	2 1 0	2 1 0	2 1 0	2 1 0
Total Points					

Teacher comments:

Teacher signature: _____

APPENDIX F: INTERVENTION INTEGRITY FORM, TEACHER

Teacher:

Date:

- | | | |
|---|-----|----|
| 1. I gave the student feedback at the end of every designated time period. | YES | NO |
| 2. I only rated student performance for <i>that</i> time period. | YES | NO |
| 3. I gave the student positive verbal feedback for performing appropriate behaviors and scoring high in <i>that</i> period. | YES | NO |
| 4. If the student scored low in that period, I gave feedback on how they could improve their behavior. | YES | NO |

APPENDIX G: INTERVENTION INTEGRITY FORM, CHECK IN

Date:

Mentor:

Caroline

- | | | |
|--|-----|----|
| 1. I gave the student the DPR before school started (8:15). | YES | NO |
| 2. I discussed yesterday's performance and how they could improve it if it was low and how it made them feel if it was high (including challenging/discussing negative cognitions around the behaviors). | YES | NO |
| 3. I discussed and practiced the replacement behaviors with the student (practice only if necessary). | YES | NO |
| 4. I set the daily goal with the student. | YES | NO |

APPENDIX H: INTERVENTION RATING PROFILE (IRP-15)

Intervention Rating Profile (IRP-15)

Please rate the intervention along the following dimensions. Please circle the number which best describes your agreement or disagreement with each statement.

	<i>Strongly Disagree</i>	<i>Disagree</i>	<i>Disagree Slightly</i>	<i>Slightly Agree</i>	<i>Agree</i>	<i>Strongly Agree</i>
1. This would be an acceptable intervention for a child's problem behavior.	1	2	3	4	5	6
2. Most teachers would find this intervention appropriate for behavior problems in addition to the one described.	1	2	3	4	5	6
3. This intervention should prove effective in changing a child's problem behavior.	1	2	3	4	5	6
4. I would suggest this intervention to other teachers.	1	2	3	4	5	6
5. The child's behavior is severe enough to warrant use of this intervention.	1	2	3	4	5	6
6. Most teachers would find this intervention suitable for behavior problem described.	1	2	3	4	5	6
7. I would be willing to use this intervention in the classroom setting.	1	2	3	4	5	6
8. This intervention would <i>not</i> result in negative side-effects for the child.	1	2	3	4	5	6
9. This intervention would be appropriate for a variety of children.	1	2	3	4	5	6
10. This intervention is consistent with those I have used in classroom settings.	1	2	3	4	5	6
11. The intervention was a fair way to handle the child's problem behavior.	1	2	3	4	5	6
12. This intervention is reasonable for the problem behavior described.	1	2	3	4	5	6
13. I liked the procedures used in this intervention.	1	2	3	4	5	6
14. This intervention is a good way to handle this child's behavior.	1	2	3	4	5	6
15. Overall, this intervention would be beneficial for a child.	1	2	3	4	5	6

APPENDIX I: INSTITUTIONAL REVIEW BOARD APPROVAL

ACTION ON PROTOCOL APPROVAL REQUEST



Institutional Review Board
Dr. Robert Mathews, Chair
131 David Boyd Hall
Baton Rouge, LA 70803
P: 225.578.8692
F: 225.578.6792
irb@lsu.edu | lsu.edu/irb

TO: Frank Gresham
Psychology
FROM: Robert C. Mathews
Chair, Institutional Review Board
DATE: January 18, 2011
RE: IRB# 3149
TITLE: Evaluation of the Check in/Check out (CICO) Intervention for Students with Internalizing Behavior Problems

New Protocol/Modification/Continuation: New Protocol

Review type: Full Expedited **Review date:** 1/14/2011

Risk Factor: Minimal Uncertain Greater Than Minimal

Approved **Disapproved**

Approval Date: 1/18/2011 **Approval Expiration Date:** 1/17/2012

Re-review frequency: (annual unless otherwise stated)

Number of subjects approved: 4

Protocol Matches Scope of Work in Grant proposal: (if applicable)

By: Robert C. Mathews, Chairman 

PRINCIPAL INVESTIGATOR: PLEASE READ THE FOLLOWING –
Continuing approval is **CONDITIONAL** on:

1. Adherence to the approved protocol, familiarity with, and adherence to the ethical standards of the Belmont Report, and LSU's Assurance of Compliance with DHHS regulations for the protection of human subjects*
2. Prior approval of a change in protocol, including revision of the consent documents or an increase in the number of subjects over that approved.
3. Obtaining renewed approval (or submittal of a termination report), prior to the approval expiration date, upon request by the IRB office (irrespective of when the project actually begins); notification of project termination.
4. Retention of documentation of informed consent and study records for at least 3 years after the study ends.
5. Continuing attention to the physical and psychological well-being and informed consent of the individual participants including notification of new information that might affect consent.
6. A prompt report to the IRB of any adverse event affecting a participant potentially arising from the study.
7. Notification of the IRB of a serious compliance failure.
8. SPECIAL NOTE:

**All investigators and support staff have access to copies of the Belmont Report, LSU's Assurance with DHHS, DHHS (45 CFR 46) and FDA regulations governing use of human subjects, and other relevant documents in print in this office or on our World Wide Web site at <http://www.lsu.edu/irb>*

VITA

Katherine Hunter is a native of Louisiana, hailing from the city of New Orleans. She completed her undergraduate education at Tufts University in the cold, northern city of Medford, Massachusetts. She graduated from Tufts with a Bachelors of Arts in Psychology in May 2008. During her time at Tufts, she was able to study abroad in Florence, Italy for a semester to focus on her minor of art history. After her worldly undergraduate experience in Europe and the northern United States, she returned to the warm and humid state of Louisiana. There, she taught special education at an elementary school in New Orleans for a year, during which she decided she would benefit from spending more time on the learning end of the student-teacher relationship. She chose to further her education at Louisiana State University in Baton Rouge, Louisiana, where she is studying the discipline of school psychology.