Managing Classroom Behavior of Head Start Children Using Response Cost and Token Economy Procedures

Jennifer D. Tiano
Beverly L. Fortson
Cheryl B. McNeil
Lisa A. Humphreys

The efficacy of three behavior management techniques used in a Head Start classroom was examined. The three techniques included: (a) techniques currently used by the teacher, (b) response cost, and (c) the Level System (token economy). The current study used an ABACA single subject withdrawal design with follow-up where all conditions were implemented until stability was reached. Classroom behavior was evaluated by both behavioral observation and teacher report. Children’s and teacher’s behavior were examined. No conclusions could be made concerning the efficacy (i.e., inappropriate behavior) of the techniques. Teachers used more labeled praise statements and fewer critical statements during the Level System condition than all other conditions.

Key words: Head start, token systems, response cost, pre-schoolers

Problematic behaviors exhibited by young children have become a topic of concern. Head Start staff members are reporting an increase in the number of children displaying challenging and disruptive behaviors (Piotrkowski, Collins, Knitzer, & Robinson, 1994). Despite the increase in classroom behavior problems, teachers reported deficiencies in managing these behaviors. In fact, behavior management strategies comprised 3 of the top 5 areas in which Head Start staff requested additional training (Buscemi, Bennett, Thomas, & DeLuca, 1995). Therefore, it is important to provide teachers with effective strategies to manage classroom behaviors.

Comparing Response Cost and Token Economy Procedures

Research has supported the effectiveness of both response cost and token economy procedures in decreasing disruptive behavior in academic settings. However, both procedures employ different strategies to manage classroom behavioral problems. Thus, it is important to examine research comparing the relative effectiveness of these two procedures.

Iwata and Bailey (1974) used a reversal design to compare reward and cost token systems with special education children in elementary school. The children’s behavior was observed every morning for three months during math class. Results indicated that both programs were equally effective in decreasing off-task behavior and the violation of the teacher’s rules. Mean percent off-task behavior returned to a level similar to baseline during the reversal phase. The number of mathematic problems completed more than doubled for the group earning tokens while the cost group only showed a small increase in problems completed. The teacher provided more statements of approval to students who were in the reward-cost group than to children who were in the cost-reward group.

Sullivan and O’Leary (1990) also compared the effectiveness of both response cost and token economy procedures. Children were observed daily for 40 minutes during reading and language and math class. Both programs were highly and equally effective in reducing the amount of off-task behavior. More specifically, average percent on-task behavior in the baseline condition was approximately 60%, but then increased to approximately 85% when both programs were in effect. However, rates of on-task behavior differed between the two programs when the programs were faded from use. When the response cost program was faded, improvements in on-task behavior were maintained for all students participating in this program. However, half of the children who participated in the reward program did not maintain treatment effects when the program was faded from the classroom.

The effectiveness of response cost and token economies in reducing disruptive behavior was examined with children with Attention-Deficit/Hyperactivity Disorder (ADHD; McGoey & DuPaul, 2000). Children were observed in their preschool classroom for 20-minute observation periods occurring at least three times
each week. Both programs were effective in reducing disruptive behavior exhibited in the classroom.

The current study examined the efficacy of three different behavior management strategies used in Head Start classrooms (i.e., techniques currently being used by the teacher, response cost, and the Level System). The techniques used by the teacher and teacher’s aide prior to treatment implementation served as the baseline and withdrawal conditions of the study. The response cost program consisted of a board with four levels (see Figure 1). The first three levels have sunshines and comprised the “sunny zone” of the board. The bottom level had clouds and comprised the “cloudy zone” of the board. A shape is assigned to each child in the classroom. Shapes were moved down contingent upon child inappropriate behavior (e.g., poking neighbor). The teacher gave a verbal warning for inappropriate behavior. If the inappropriate behavior continued, that child’s shape was moved down one level. However, a shape was moved down without a warning for destruction of property and hurting. At specified times throughout the day each child whose shape was in the sunny zone (i.e., any of the three sunny levels) received a reward. The Level System (McNeil & Filcheck, 2001) possesses characteristics of both a token economy and response cost and provides teachers of young children with strategies in the management of behavior problems. The Level System has seven levels: three sunny levels, one neutral level, and three cloudy levels (see Figure 2). Each child in the classroom used the same shape they were assigned in the response cost condition. For appropriate behavior, the teacher provided social reinforcement (i.e., labeled praise) and moved the children’s shapes up one level. Consequences for inappropriate were the same as the response cost program. While the response cost only permitted the downward movement of shapes, shapes in the Level System were moved down for antisocial behavior and up for prosocial behavior.

The Level System is a new program and only a small amount of literature exists supporting its effectiveness. Both child and teacher behaviors were affected by the implementation of the Level System. More specifically, child inappropriate behavior decreased with the Level System (Filcheck, 2003; Filcheck, McNeil, Greco, & Bernard, 2004). The amount of praise provided by the teacher increased, while the amount of teacher criticisms decreased with the use of the Level System (Filcheck). Upon completion of research, some teachers chose to use the Level System (Filcheck, McNeil, Greco, & Bernard), but others did not (Filcheck). Interestingly, the teachers in Filcheck’s study reported high satisfaction ratings for the Level System, but chose not to implement this program at follow-up. These teachers indicated that the children’s behaviors had improved and the program was unnecessary due to the amount of time required for implementation.

The current study provided novel information to the existing literature on response cost and token economy procedures. Most research that has evaluated response cost and token economy programs with young children has been implemented with individual children. The current project implemented both a response cost program and token economy on a whole-class basis. In addition, no research was found comparing the efficacy of these two programs in a Head Start setting.

Two hypotheses were examined in this study. First, it was hypothesized that the target children would exhibit less inappropriate behaviors when the response cost and the Level System were implemented than when the teacher was using the strategies utilized before the study, and the response cost and token economy procedures would be similarly efficacious in reducing problem behaviors (Iwata & Bailey, 1974; McGoey & DuPaul, 2000; Sullivan & O’Leary, 1990). Second, it was hypothesized that the teacher would use more criticisms and fewer praise statements in the response cost and baseline/withdrawal conditions than in the Level System condition. This result was expected because teachers were trained to reduce the number of criticisms used and increase the number of praises they gave children in the Level System condition. However, these skills were not targeted during the response cost procedure. In addition, Iwata and Bailey found that teachers engaged in more approval statements with children participating in the token economy procedure than with children participating in the response cost program.

Method

Setting

Data were collected in one Head Start classroom in southwestern Pennsylvania in the children’s regular classroom with the primary teacher and teacher’s aide. However, data were not collected when the primary
Participants
Participants for the treatment component of this study were 3 children between the ages of 3 and 5 who were enrolled in the Head Start program. Participants were identified by the teacher as exhibiting disruptive behavior based on the teacher’s report on the Conners’ Global Index (CGI, Conners, 1997). All three child participants were four years old. Two children participated in the entire study, but one child (“Damon”) withdrew from Head Start prior to the follow-up condition.

Measures

Treatment Efficacy Measures
Revised Edition of the School Observation Coding System (REDSOCS). The REDSOCS (Jacobs et al., 2000) is a coding system used to rate observed behavior of both students and teachers in the classroom setting. The REDSOCS uses 10-second observe intervals with immediate marking of behaviors. Observations occurred for approximately 40 minutes during structured morning activities. The teacher’s behaviors recorded were unlabeled praise, labeled praise, and criticism (adapted from the Dyadic Parent-Child Interaction Coding System – Second Edition; Eyberg, Bessmer, Newcomb, Edwards, & Robinson, 1994). Each unlabeled and labeled praise and criticism was coded regardless of which child the teacher directed the statements toward. Child inappropriate behavior and teacher behavior were coded to obtain total percentages during that observation. Jacobs et al. report good psychometric properties of the REDSOCS. Interobserver agreement for the Appropriate Behavior and Inappropriate Behavior categories were .85 and .83 respectively.

Treatment Integrity Measure
Treatment Integrity. Treatment integrity, evaluated using a treatment integrity checklist for each treatment condition, was completed daily by one research assistant during both treatment phases. Accurate implementation of the interventions was achieved if a score of 85% or higher was obtained on the treatment integrity checklist. If a score less than 85% was obtained on 2 consecutive days, data for those days were omitted from data analyses and both the teacher and teacher’s aide were retrained in the accurate implementation of the intervention. Retraining was not needed for either treatment program (i.e., response cost and Level System). The teacher and teacher’s aide were aware that these integrity measures were completed, and feedback on their performance was provided daily.

Interobserver agreement. Interobserver agreement data were collected randomly for 25% of the observations in each design phase. Interobserver agreement remained at or above .75 Kappa, and no retraining was required for either research assistant. The research assistants were unaware of the study’s hypotheses.

Procedure

Teacher Training
Baseline data were collected prior to the implementation of the treatment phases. The teacher and teacher’s aide received a 1-hour workshop for each program (i.e., response cost and Level System). Then the teacher and teacher’s aide were coached in the classroom in the use of each program until 85% treatment integrity was achieved.

Experimental Conditions
The current study used a single-subject withdrawal design (i.e., ABACA) with a 6-week follow-up evaluation. The three conditions that were comprised in the design included: behavior management techniques currently employed by the teacher and teacher’s aide (“A”), response cost (“B”), and the Level System (“C”). Each condition continued until data were stable.
Figure 1. Pictorial representation of the response cost program board.
Figure 2. Pictorial representation of the Level System board.
Figure 3. Average percent of intervals scored with inappropriate behavior exhibited by each participant. ■ indicates stimulant medication implementation for Ruby and // indicates summer
Figure 4. Average percentage of intervals scored with labeled praises, unlabeled praises, and criticisms exhibited by the teacher. // indicates summer break.
Results

The efficacies of techniques currently used in the classroom, response cost, and the Level System was examined by visual inspection of behavioral observation data (i.e., REDSOCS). Results demonstrated a decreasing trend in inappropriate behavior throughout the study for each child (Figure 3). As inappropriate behavior did not return to baseline levels during withdrawal conditions, no conclusions about the efficacy of the treatment programs could be made based on these data. “Ruby” began taking stimulant medication concurrently with the implementation of the Level System, further prohibiting conclusions about the efficacy of the program in decreasing inappropriate behavior for this child. In addition, data collection ceased for approximately three months following the response cost phase due to summer break. The decrease in observed child behaviors is consistent with the teacher report of these behaviors. More specifically, Ruby’s scores on the CGI decreased from the disruptive range at pre-treatment (T=>90) to the borderline range at post-treatment (T=59). Mitch’s behavior scores decreased from the borderline range (T=57) to the typical range (T=46). The teacher rated Damon’s behavior in the disruptive range (T=63) prior to the study. No post-treatment scores were obtained for Damon because he withdrew from Head Start.

Teacher data are displayed in Figure 4. Visual inspection of labeled praise data for teachers suggested an increase in the use of labeled praise in both treatment conditions as compared to baseline and withdrawal conditions. Labeled praise was used more often in the Level System condition than in any other condition of the study. In addition, the percentage of intervals that the teacher used these statements did not return to baseline levels in the second withdrawal or follow-up phases. The average percentages of intervals containing labeled praise statements for the conditions were as follows: 1.15 (baseline), 1.99 (response cost), 1.38 (withdrawal 1), 5.31 (Level System), 3.32 (withdrawal 2), and 3.04 (follow-up). Visual inspection of unlabeled praise data suggested that the teacher used more unlabeled praise statements in the latter conditions of the study than the former conditions. The average percent of intervals with unlabeled praises were as follows: 2.34 (baseline), 4.69 (response cost), 9.48 (withdrawal 1), 7.12 (Level System), 7.64 (withdrawal 2), and 8.96 (follow-up). The average percent of intervals containing criticisms appeared to decrease during intervention phases, but increase with the removal of the programs (with the exception of follow-up). The mean percentages were as follows: 6.66 for baseline, 5.28 for response cost, 6.43 for withdrawal 1, 3.5 for the Level System, 4.61 for withdrawal 2, and 2.12 for follow-up. The teacher used the least critical statements in the follow-up condition as compared to all other conditions.

Discussion

The current study examined the efficacy of three behavior management strategies (i.e., current strategies, response cost, and the Level System) used in a Head Start classroom. No conclusions about treatment efficacy could be reached through inspection of student data. Inappropriate behavior continued to decrease throughout the study whether a treatment was implemented or not. These findings could be attributed to the following aspects. First, the teacher’s behavior changed throughout the study. She used a greater number of praise statements and fewer critical statements as the study progressed. Similarly, the teacher in the study by Filcheck, McNeil, Greco, and Bernard (2004) used more praise and fewer criticisms throughout the study. The teacher’s use of more positive interactions with the children may have reinforced child appropriate behavior. This carryover effect (e.g., Parsonson & Baer, 1992) indicates that no functional control was obtained and limits the ability to draw conclusions about treatment efficacy.

Teacher behavior also was examined in the current study. The approximate rate of praise statements per hour used by the teacher changed throughout the study: 4.5 per hour for baseline, 7.5 per hour for response cost, 5.0 per hour during the first withdrawal phase, 19.5 with the Level System, 12.0 in the second withdrawal phase, and 11.0 at follow-up. As expected the teacher used the most labeled praise statements during the Level System. This finding is consistent with previous research that also found increased use of teacher labeled praise during implementation of the Level System (Filcheck, McNeil, Greco, & Bernard, 2004). The teacher was instructed to use these statements during the Level System phase, but not in other phases. Instructing the teacher to attend to appropriate behaviors and provide social rewards for those behaviors (i.e., verbal praise) could promote a more positive atmosphere in the classroom with the use of the
Level System (Filcheck, McNeil, Greco, & Bernard). Specifically, percent of intervals of appropriate child behaviors should increase as those behaviors are receiving reinforcement from the teacher. In addition, the teacher praised more at follow-up than at baseline. Filcheck (2003) also found this increase in teachers’ use of labeled praises. The Level System may have impacted the teacher’s skill by encouraging her to attend to and reinforce appropriate classroom behavior. The labeled praise statements may have resulted in improvements in classroom behavior, thus decreasing the necessity of the time-intensive Level System in managing behavior.

Unlabeled praises and criticisms also were examined to assess teacher behavior. The teacher used more unlabeled praise statements toward the end of the study as compared to the beginning of the study. This finding may be attributed to the targeting of praise statements during the Level System condition. The teacher was instructed to use labeled praises during the Level System, but may have reverted to using unlabeled praise statements following the Level System condition. In other words, using praise may have become part of the teacher’s behavioral repertoire. Without feedback, she, however, used vague, shorter unlabeled praise statements as opposed to the more difficult skill of specific labeled praise. The use of more unlabeled praises throughout the study also could be attributed to the children’s behavior. More specifically, the teacher may have begun praising more as the children’s appropriate behaviors improved. Unfortunately, the study’s methodology does not permit conclusions to be made as to whether the teacher’s use of praise influenced child behavior or child behavior influenced the teacher’s use of praise.

The teacher’s use of criticisms appeared to vary with the implementation of treatment programs. Though teachers used the fewest criticisms in the follow-up phase, criticisms were used less often in the intervention phases as opposed to the no-intervention phases. It appears that functional control was established with regard to criticisms. The frequency of criticisms decreased when the interventions were used, but increased with the removal of these programs. These data suggest that the response cost and Level System programs aided in decreasing the teacher’s use of critical statements, particularly in the Level System. The frequency of criticisms may have been impacted by the students’ behaviors as well. The decreasing trend in child inappropriate behaviors throughout the study may have resulted in fewer opportunities for the teacher to use critical statements. In addition, the teacher delivered a reduced amount of criticisms in the Level System as compared with the baseline, response cost, and withdrawal phases. The low percentage of intervals of criticisms during this phase would be expected as criticisms were targeted in the Level System condition. Further, the teacher was using more labeled and unlabeled praise statements, thus decreasing the amount of time she would have to deliver critical statements. The use of fewer criticisms and more praise statements with the children could help to foster a more positive classroom environment for the children as well as the Head Start staff members.

The teacher in the current study was interviewed and reported both advantages and disadvantages of the two programs. When asked to compare the two programs, she felt that the response cost had a stronger impact than the Level System on the children in the classroom because the children had fewer chances to engage in inappropriate behavior and still earn the reward. On the other hand, the teacher reported that the response cost program was a more negative program than the Level System, as the response cost did not allow the shapes to be moved back up the board. The moving of shapes up the board was a reported advantage of the Level System, but the frequent moving of shapes was time consuming. Furthermore, McGoey and DuPaul (2000) suggested that teachers may choose not to implement token economies because these systems require much time and effort from the teacher. The teacher in the current study further indicated that she was unsure as to whether the children attended to the frequent moving of shapes on the board. In addition, she felt the Level System may not be effective because the children are provided many chances for their shapes to move up the board and earn the reward. Thus, moving shapes down may not be punishing enough to lead to behavior change. When asked about modifications that could improve the programs, the teacher discussed the possibility of using the program only during challenging time periods and with a small group of children of the same age. McGoey and DuPaul also suggested that token economies may be difficult to implement in large classrooms.

Limitations
Participant selection. Child participants were selected for participation from teacher report of disruptive behavior on the CGI. However, the children identified by teacher report as exhibiting the most disruptive behavior in the classroom were not observed as exhibiting high rates of disruptive behavior. Percent of intervals of inappropriate behavior continued to decrease throughout the study for all children. Thus, a floor effect for inappropriate behavior occurred creating difficulty in distinguishing between treatment and no-treatment conditions based on these data. Further change in the dependent variable is difficult to detect when data reach a low limit (Kazdin, 1980). Therefore, the decrease in inappropriate behavior during the utilization of the response cost and withdrawal phases prohibit additional decrease in inappropriate behavior to occur throughout the study. In addition, the CGI may not have assessed the child behaviors with which the teacher was experiencing difficulty. Perhaps a different measure of child behavior or an interview with the teacher to identify target behaviors would improve the consistency between teacher report and direct observations of child data.

Observation Assessment. The observation system utilized for the current study provided 14 minutes of data per observation for each child participant. Thus, a limited amount of behavioral data was obtained for each participant. Therefore, conclusions about the children’s behavior are difficult to ascertain from 14 minutes of behavioral data. In addition, this short observation period may have contributed to the difficulty in establishing stability in child data. The teacher could have indicated the most difficult periods of the day. Observations could be conducted at those times to capture child inappropriate behavior.

Study Confounds. The lack of reversal in child behavior to baseline levels with the removal of the interventions (i.e., during withdrawal conditions) was another limitation. This lack of reversal disallows conclusions to be made about the efficacy of the treatments in decreasing child disruptive behavior. Therefore, one or more variables may have been present that confounded the current data. These confounding variables may have included lengthy conditions, extended break from data collection, decrease in class size, and other interventions. Several study conditions were extended due to the difficulty in obtaining stability in the data. In addition, summer break from school created a long time interval between the response cost and first withdrawal phases (approximately 2 months). These two facets caused the project to extend for a longer period of time (i.e., approximately one year) than originally anticipated, and therefore may have lead to unexpected confounds and further limitations of this study (e.g., child maturation). Children were engaging in different activities throughout these conditions. For example, children were participating in outdoor activities in the warm months and indoor activities in the winter and on rainy days. Thus, these different activities in different seasons may have impacted child classroom behavior. Moreover, the number of students enrolled in this Head Start classroom decreased throughout the study. Initially there were approximately 18 children enrolled in the class. Following summer break, however, the number of students decreased to approximately 10 children due to the discontinuation of bus service. This decrease in the number of children could have lead to less teacher stress, thus increasing her ability to utilize effective behavior management strategies. In addition, the teacher may have had more time to provide individual attention and social reinforcement to the children, thereby rewarding appropriate behavior. Finally, “Ruby” began taking stimulant medication during the Level System phase. Therefore, it is impossible to conclude whether a decrease in inappropriate behavior for “Ruby” was due to the interventions or the medication.

This study also has limitations concerning the population and location. This study was conducted in one classroom with one teacher and three children. Therefore, the small sample size makes it difficult to generalize findings to the general population. Results of this study could be due to dynamics of this particular classroom that may not be representative of most Head Start classrooms. For example, the Head Start classroom in this study was located in rural Pennsylvania, and results may not be generalizable to urban Head Start classrooms. Finally, the participants were exhibiting disruptive behaviors at low rates, and different results may be obtained in classrooms with children exhibiting more severe behavior problems.

Directions for Future Research

Future research should be conducted to determine the efficacy and social validity of these two programs (i.e., response cost and Level System), as well as their impact on teacher skill. These studies should examine the programs in both rural and urban Head Start classrooms to determine if the efficacies of these programs vary based on location. In addition, future research should be conducted utilizing a larger sample
size to increase the ability of results to be generalizable to and representative of the larger population. Data collection should extend over a short period of time to ensure the control for maturation effects and the use of a rapid reversal design may control for school-year effects. Individual data collection points, however, should be conducted for longer than 14 minutes per child to allow a more representative sample of each child’s behavior. Furthermore, future research should utilize baseline observation data in addition to teacher report in determining child participants. Thus, floor effects in disruptive behavior could be avoided as children observed to be exhibiting high rates of disruptive behaviors would be chosen for participation.

Future research could examine the efficacy of teacher labeled praise compared to the Level System. This research design may allow one to determine if improvement in classroom behavior can be attributed to the entire Level System program or certain components of that program. For example, decreased inappropriate classroom behavior may be obtained only with the use of teacher labeled praise. Future research also should be conducted to determine if these response cost and Level System programs impact child behavior. If child inappropriate behavior is found to decrease with these programs, perhaps the response cost and token economy programs could be used as a short-term training tool in schools and faded out of the classroom following behavior change. Short-term implementation of the program may teach skills to both teachers and students that can be maintained following termination of the program. Future research, however, will be necessary to conclude whether the response cost and token economy programs can make an impact on behavior problems in Head Start classrooms.

References


Acknowledgements

We would like to thank Southwestern Pennsylvania Head Start Association for their assistance with this project.

Author Contact Information:

Jennifer Tiano is a graduate student in the Department of Psychology at West Virginia University. Please address correspondence to Jennifer Tiano at jtiano@mix.wvu.edu.

Beverly Fortson is a predoctoral intern at the Medical University of South Carolina. Cheryl McNeil is an associate professor of psychology at West Virginia University. Lisa Humphreys is a graduate student at Indiana University.

Advertising in the Journal of Early and Intensive Behavior Intervention

Advertising is available in JEIBI. All advertising must be paid for in advance. Make your check payable to Joseph Cautilli. The ad copy should be in our hands at least 3 weeks prior to publication. Copy should be in MS Word or Word Perfect format and advertiser should include graphics or logos with ad copy.

The prices for advertising in one issue are as follows:

1/4 Page: $50.00    1/2 Page: $100.00 vertical or horizontal    Full Page: $200.00

If you wish to run the same ad in both issues for the year, you are eligible for the following discount:

1/4 Pg.: $40 - per issue
1/2 Pg.: $75 - per issue -vertical or horizontal
Full Page: $150.00-per issue.

For more information, or place an ad, contact: Joe Cautilli via e-mail at jcautilli@cctckids.com or by phone at (215) 462-6737.