*BEHAVIORAL SCHOOL PSYCHOLOGY GOES OUTDOORS:
THE EFFECT OF ORGANIZED GAMES
ON PLAYGROUND AGGRESSION*

H. ALLEN MURPHY, J. MICHAEL HUTCHISON, AND JON S. BAILEY

FLORIDA STATE UNIVERSITY

This research focuses on the inappropriate, largely aggressive, behaviors of 344 K-2 children assigned to a playground prior to the beginning of the school day. Initially, a system for observing large numbers of children freely roaming over a large, open area was developed. This observational method was then used to determine the effectiveness of providing organized games for reducing potentially dangerous playground behaviors. Using a reversal (ABAB) design, it was found that the games, rope jumping, and foot racing, along with an infrequently used time-out procedure, significantly reduced the frequency of inappropriate incidents. It is suggested that when dealing with large groups, antecedent environmental manipulations may be more practical than providing consequences for the behaviors exhibited by identified individuals.

DESCRIPTORS: recreation activities, aggressive behavior, group control, playground behavior, children

The elementary schools have been a popular and productive setting for applied behavior analysts (Kazdin, 1975) who have directed a great deal of effort toward the problem of managing students in the classroom (e.g., Barrish, Saunders, & Wolf, 1969; Greenwood, Hops, Walker, Guild, Stokes, & Young, 1979; Madson, Becker, & Thomas, 1968; O'Leary, Becker, Evans, & Saudargas, 1969). However, relatively little attention has been paid to the difficulties encountered by schools outside the classroom setting. This is somewhat ironic since many of these (e.g., lunch, recess, assemblies) are particularly problematic due to the children's greater numbers and dispersal.

Technologies have emerged for a few of the problematic situations outside the classroom. In a day-care center, LeLaurin and Risley (1972)

found that assigning staff members to specific locations rather than to specific children improved orderly behavior during transitions from one activity to another. Annoying, and potentially dangerous, behaviors on school buses have been decreased by stopping the bus whenever a disruptive incident occurred (Campbell, Adams, & Ryabik, 1974), or terminating music contingent on disruptiveness or excessive noise (Greene, Bailey, & Barber, 1981; Ritschl, Mongrella, & Presbie, 1972). Safe pedestrian skills have been taught by analyzing the skills into their component steps and training small groups of children at the specific intersections they normally encountered while walking to school (Yeaton & Bailey, 1978).

Behavioral methods have also been applied in school lunchrooms. MacPherson, Candee, and Hohman (1974) trained aides to use standard management techniques, contingent praise, and the assignment of extra academic tasks, to decrease disorderly behavior in lunchrooms used by about 40 students at a time. More recently, Bostow (Note 1; Gordon & Bostow, Note 2) has reduced the difficulty of managing large numbers of children in lunchrooms through the

The authors gratefully acknowledge the assistance of the faculty and staff at Kate Sullivan elementary school. Special thanks are due to Nancy Duden, principal, Linda Alexionic, former physical education teacher, the teachers' aides on whom many demands were placed, and the too-numerous-to-list college students involved in this project. Reprints may be obtained from Al Murphy, Department of Psychology, Florida State University, Tallahassee, Florida 32306.

use of electronic devices designed to monitor and provide feedback on noise levels.

With the exception of Bostow's research, these studies described techniques designed for working within a relatively confined setting with classroom size groups for whom reasonably well-defined activities are expected. Very different circumstances arise on the playground where large numbers of children are allowed to roam a frequently undersupervised open area freely and are instructed merely to "play."

At the beginning of the 1979-1980 school year, the faculty of a Tallahassee elementary school contacted the authors for advice on dealing with the rowdy behavior of students who arrived early and stayed on the playground until the beginning of the school day. During this year an observational system was developed and several management strategies were compared with three teachers' aides carrying out the procedures. Here, it was found that the greatest improvement in the children's behavior resulted when organized games were provided as an alternative to the activities in which they normally engaged when left to their own devices.

The current research systematically replicates the previous year's pilot work at the same elementary school. A reversal (ABAB) design was used to determine the effectiveness of organized games for reducing unwanted behaviors on the playground.

METHOD

Children and Setting

Prior to the 8:45 a.m. opening of school, the 344 kindergarten, first-, and second-grade children were assigned to a one-acre playground located on the side of the school building. The playground was partially fenced-in and contained seven structures for climbing, a balance beam, and a sliding board.

Three teachers' aides were assigned to oversee the children, with one arriving at 8:05 a.m. and the remaining two at 8:35. In order to maintain

the same child-adult ratio as was available in the previous year when two additional aides were available at 8:15, one or two college students assisted in the supervision of the playground activities. Children came to the playground as early as 7:30 and their numbers increased to about 180 by 8:35, and 220 by 8:40. On cold or rainy days the pupils went to the cafeteria or their homerooms.

Dependent Variables

Based on school reports and informal observations, the primary measure was comprised of the following three classes of problem behaviors: (a) *aggression*: striking, slapping, tripping, kicking, pushing or pulling others; "karate" moves ending within 1 foot of another person; doing anything that ends with another child falling to the ground; (b) *property abuse*: taking another person's property without permission; throwing school books, lunches, or anyone else's property; throwing any object at passing or parked cars; digging holes in the ground with one's feet or hands; breaking pencils or pens; (c) *rule violations*: resisting or talking back to an aide; climbing more than 1 foot off the ground on a playground structure; suspending one's self on a playground structure in any position that resulted in the head not being 180 degrees above the feet. The final two playground structure items were included at the request of school personnel who felt that there was not enough supervision to ensure safe usage of the structures during the preschool period.

In addition, ratings of the children's behavior were obtained from the three aides assigned to the playground on a daily basis. The rating instrument covered six child problem behaviors (hitting other children, talking back to you, arguing, playing dangerously, fighting, and complaining) each of which was evaluated on a 4-point scale ("not at all," "just a little," "pretty much," or "very much") in a fashion similar to that used in an accepted teacher rating scale (Conners, 1969). Further, a 4-point rating of a

work-related category ("How would you rate the difficulty of your job today?") was included. These ratings were intended as measures of social validity and worker satisfaction.

The Playground Observation System

After numerous attempts during the pilot study, an acceptable observation strategy was found. The playground was divided into three roughly equivalent "pie slices" based on the existing geography, (e.g., building corners or edges of playground equipment) which were the responsibility of separate observers. These slices were then halved (again, defined by other permanent structures), and each half was monitored for alternate 15-sec periods. Thus, an observer attended to only one-sixth of the playground at a time, and only half the playground was observed at any given moment.

Three observers stood in the middle of the playground facing their areas. A tape recording instructed them to start watching the left-hand portion of their slice at which time they began recording incidents with the aid of hand counters. After 15 sec, the tape cued a "switch" to the remaining portion of the observers' area. This continued for 2 min when a "stop" signaled that the cumulative frequency of incidents observed was to be entered on the data sheets. The entire process occurred for 10 iterations (i.e., observe left for 15 sec; observe right for 15 sec; and back left again; recording the totals every 2 min) from 8:20 to 8:40 a.m.

A particular inappropriate incident (e.g., kicking) directed at one child by another individual was only counted once per 15-sec interval. However, more than one incident was scored if one child inflicted several types of aggression on another (e.g., one child hitting and kicking another resulted in two incidents being counted). In either the case of two children assaulting a third individual, or one child assaulting two peers, two incidents were scored. The 15-sec intervals were arbitrarily considered to be independent; thus, if two children were observed to

be wrestling with one another for two intervals, four incidents were recorded.

Observer Training

Seven undergraduate college students were recruited from psychology classes to participate as observers. They were given a 2-page handout which defined the behaviors to be observed and described the observational system. Typically, the first day of training on the playground itself consisted of reviewing the observation system and definitions, citing examples of ongoing incidents, and having the students point out incidents that were occurring and explaining how these should be scored. On later training, the sessions began with a brief review, a 10- to 14-min period of observation, and feedback on their performance for that day. Asking for clarification was encouraged at all times except during the practice observation period. However, if a student expressed confusion during the practice period, the observations were occasionally interrupted to answer any questions that had arisen and then restarted.

Generally, about five days of training were sufficient for observers to reach criterion agreement. The criteria were two consecutive practice periods at 80% agreement with the first author or another experienced observer.

Experimental Conditions

Baseline. For the first 12 days, inappropriate incidents were recorded and ratings taken without any intervention. On the last day of baseline a 1-page description of the organized games and time-out contingency was read and explained to the children by their teachers or the first author in the individual classrooms.

Organized games. Rope jumping and foot races were selected as the organized activities. Ten nylon ropes were made available by an aide or college student who supervised the play and provided instruction on such rudimentary skills as forming lines while waiting to jump and turning ropes in unison. A second adult was re-

sponsible for organizing the races. Here, four lanes were constructed by stringing ropes through plastic pipes which were embedded vertically in the ground. The children lined up behind the lanes until it was their turn to race between the ropes to the end pipe.

For particularly unruly behaviors a time-out procedure was in effect. Children were placed on a specially provided bench or a balance beam, whichever was closer, and were required to sit quietly for 2 min as measured by kitchen timers. Notepads were placed at the time-out site for recording the names of those children who were sent there.

Prior to the beginning of the organized games condition, all adults involved with supervising the children read a 2-page handout which emphasized the types of behaviors that were considered to be inappropriate, patrolling and scanning the entire playground, unemotional delivery of discipline, and frequent praise for acceptable behavior (especially for those children who were normally considered to be problems). These principles were discussed during one-hour meetings with one of the authors, and in vivo modeling and feedback were provided for the operation of the activities and the use of time-out.

Baseline. Preintervention conditions were re-established for 4 days.

Organized games. Organized activities and time-out were in effect for the final 6 days before Christmas break.

RESULTS

Reliability

Interobserver checks were made on 37 (67%) of the days across conditions and observation areas. Frequency reliability was computed by dividing the smaller observed frequency by the larger and multiplying by 100. Agreement ranged from 67% to 100% with a mean of 87%.

Inappropriate Incidents

The daily totals of incidents observed are depicted in Figure 1. The average number of incidents during the initial baseline was 212 per 20-min period, whereas introducing the activities more than halved the incidents to a mean of 91. Returning to baseline increased the observed number of incidents to an average of 191, and the final intervention decreased the daily mean to 97.

The time-out procedure was infrequently used in that, of the 13 experimental days, children were sent to the bench or balance beam on only 10 days, and an average of only 2.5 children were placed in time-out per day. On the other hand, alternative disciplinary measures were often used. The most frequent addressed the problem of children running into others when returning from their turns at racing. This was quickly eliminated by requiring offenders to walk slowly back to the finish line before returning to the waiting area for their next turn. Similarly, children who presented problems while jumping rope, e.g., intentionally spinning the rope incorrectly or jumping out of turn, were sent to the back of the line for their rope jumping group.

Aide Ratings

Visual inspection of the aides' daily checklist responses indicated that there was little correspondence between these data and those gathered by direct observation of aggressive incidents. To examine the ratings further, correlations and analyses of variance were carried out. However, since these data are of the time-series variety, the statistical tests can be considered descriptive at best.

Analyses of variance confirmed the lack of discrimination among experimental conditions by the aides' ratings of the children. Of the 21 analyses (seven rating items completed by three aides) only one, "complaining" as rated by Aide 1, approached significance, $F(3, 19) = 3.7$,

$p = .03$. Given the large number of tests computed, this single finding should be considered negligible, especially when compared to the same analysis of observed incidents, $F(3, 24) = 29.2, p < .0001$.

Correlations of the rating data indicate that the aides did not respond differentially to individual items, i.e., items rated by a single aide were highly intercorrelated. Conversely, the ratings among aides were found to be unreliable inasmuch as individual items did not correlate well across aides.

DISCUSSION

Several issues warrant further elaboration; among these is the observation system. Due to

the number of children and size of the areas under observation, it was impractical to monitor separately different classes of behaviors (e.g., hitting children vs. resisting an aide) or the intensities of the responses (e.g., repeatedly striking a person in anger vs. a, perhaps playful, slap). Nonetheless, it was clear that the majority of incidents were of the aggressive type, and that the more serious aggressions were the most affected by the introduction of games.

The potential for biased results in this experiment exists from the need to use the same individuals in the dual roles of observers and change agents. However, the two students who most often acted as aides were not involved in collecting any of the primary data, and the

Fig. 1. Frequency of incidents recorded during the 20-min morning observation periods on the playground.

other students never served both functions on the same day. Further, none of the students were informed of the purpose of the experiment, although it was clear that many of them were able to discern it during the course of the study.

On the other hand, three characteristics of the intervention suggest that the data indicating treatment effectiveness are robust. First, since the number of children and potential for observing inappropriate incidents increased within the daily observation period, later monitoring times would have inflated the number of observed incidents. This was the case during treatment phases, i.e., equipment preparation for the games often led to later initiation of observations than during baseline. Second, in order to participate in the games the children essentially restricted themselves to relatively small areas (e.g., waiting lines) which increased the probability of unintentional physical contacts that could have been recorded as aggressive. Third, since the games were attractive to a large number of children, more children were probably on the playground when the games were available. This was not restricted to the defined subjects, but also included third- through fifth-grade pupils who were assigned to a different area but came to the younger children's playground to participate in the activities.

In the pilot study, time-out was included as a separate treatment but was so infrequently used by the aides that our intention was not to implement any time-out procedures in the present study. However, the aides and school principal requested that some method be made available for the more severe pupil infractions and, again, it appears that time-out was used sparingly. That time-out procedures are effective methods for controlling behavior is not an issue here. Rather, with the size of the population and playground, it is questionable whether any procedure relying completely on the provision of individual consequences would be practical.

Alternatively, the relatively simple environmental manipulation of providing appropriate activities had an obvious impact on the chil-

dren's behavior. Although rope jumping remained essentially the same over the pilot and this experiment, racing underwent a change toward greater environmental control. Replacing the start, finish, and waiting areas denoted by painted milk containers in the pilot research with clearly defined racing lanes in this study was intended to provide smoother movement of children through the racing activity, much like the lanes used by banks, movie theaters, and amusement parks. (The lanes here did not extend into the waiting area. Nonetheless, the well-defined racing lanes were sufficient stimuli for the children to form lines while waiting.) However, a beneficial side effect appeared. During the pilot study children were roughly split into three groups: rope jumpers, racers, and those who attempted to amuse themselves on other parts of the playground where supervision was more difficult. Introducing the racing lanes produced a fourth subgroup, the fans, who stood at the perimeter of the racing lanes and cheered their favorite runners. That watching the races was reinforcing for the fans was evident by a method used for crowd control. Whenever the fans pressed against the ropes no one was allowed to race, and very soon the fans stood a respectful pace behind the ropes and admonished one another to comply.

A further advantage of the use of games lay in their placement. During baseline, the aides usually huddled together in a location next to the side entrance of the school and directed most of their attention to those children who came to them to be "mothered" or to complain about other children. Giving the aides specific supervisory roles required them to disperse and come into contact with a much greater portion of the playground population. In effect, a "zone defense" (LeLaurin & Risley, 1972) was achieved.

The lack of consistency among the ratings provided by the aides and the observational data seems to have been a function of the aides' insensitivity to the children's behavior. Since there was no agreement among the aides, while

much higher relationships were found across the items rated by a single aide, it appears that responses to the various items were a function of some general prevailing condition rather than the children's actual performance on the specific items.

A final issue concerns the distinction between problem children and problematic situations. It could be argued that the majority of incidents was a function of a core of difficult pupils, and in a simultaneous but separate effort some very preliminary data were collected which suggest that the major feature distinguishing aggressive-hyperactive children from their peers on the playground was the incidence of negative peer interactions. Intensive training of these individuals would certainly be desirable and would probably have an impact on the overall playground aggressiveness. However, it was clearly not the case that a small number of children accounted for the majority of aggressive incidents, and that virtually every child assigned to the playground engaged in some form of inappropriate behavior when given nothing better to do.

REFERENCE NOTES

1. Bostow, D. E. *Devices which mediate reinforcement contingencies in school and community settings*. Paper presented at the annual meeting of the Association for Behavior Analysis, Milwaukee, May 1981.
2. Gordon, J., & Bostow, D. E. *Lunch room noise management in an inner city public elementary school*. Paper presented at the annual meeting of the Florida Association for Behavior Analysis, Orlando, September 1981.

REFERENCES

- Barrish, H. H., Saunders, M., & Wolf, M. M. Good behavior game: Effects of individual contingencies

for group consequences on disruptive behavior in a classroom. *Journal of Applied Behavior Analysis*, 1969, 2, 119-124.

- Campbell, D. P., Adams, R. M., & Ryabik, J. E. Group-contingent time-out and behavior on a school bus. *Psychological Reports*, 1974, 34, 883-885.
- Conners, C. K. A teacher rating scale for use in drug studies with children. *American Journal of Psychiatry*, 1969, 126, 884-894.
- Greene, B. F., Bailey, J. S., & Barber, F. An analysis and reduction of disruptive behavior on school buses. *Journal of Applied Behavior Analysis*, 1981, 14, 177-192.
- Greenwood, C. R., Hops, H., Walker, H. M., Guild, J. J., Stokes, J., & Young, K. R. Standardized classroom management program: Social validation and replication studies in Utah and Oregon. *Journal of Applied Behavior Analysis*, 1979, 12, 235-253.
- Kazdin, A. E. Characteristics and trends in applied behavior analysis. *Journal of Applied Behavior Analysis*, 1975, 8, 332.
- LeLaurin, K., & Risley, T. R. The organization of day-care environments: "Zone" versus "man-to-man" staff assignments. *Journal of Applied Behavior Analysis*, 1972, 5, 225-232.
- MacPherson, E. M., Candee, B. L., & Hohman, R. J. A comparison of three methods for eliminating disruptive lunchroom behavior. *Journal of Applied Behavior Analysis*, 1974, 7, 287-298.
- Madsen, C. H., Becker, W. C., & Thomas, D. R. Rules, praise, and ignoring: Elements of elementary classroom control. *Journal of Applied Behavior Analysis*, 1968, 1, 139-150.
- O'Leary, K. D., Becker, W. C., Evans, W. B., & Saundargas, R. A. A token reinforcement program in a public school: A replication and systematic analysis. *Journal of Applied Behavior Analysis*, 1969, 2, 3-13.
- Ritschl, C., Mongrella, J., & Presbie, R. L. Group time-out from rock and roll music and out-of-seat behavior of handicapped children while riding a school bus. *Psychological Reports*, 1972, 31, 967-973.
- Yeaton, W. H., & Bailey, J. S. Teaching pedestrian safety skills to young children: An analysis and one-year followup. *Journal of Applied Behavior Analysis*, 1978, 11, 315-329.

Received March 2, 1982

Final acceptance August 20, 1982