JOURNAL OF APPLIED DEVELOPMENTAL PSYCHOLOGY **14,** 317–345 (1993)

The Short-Term Impact of Two Classroom-Based Preventive Interventions on Aggressive and Shy Behaviors and Poor Achievement

LAWRENCE J. DOLAN, SHEPPARD G. KELLAM,
C. HENDRICKS BROWN,
LISA WERTHAMER-LARSSON, GEORGE W. REBOK,
LAWRENCE S. MAYER,
JOLENE LAUDOLFF, JAYLAN S. TURKKAN
The Johns Hopkins School of Hygiene and Public Health

CARLA FORD
LEONARD WHEELER
Baltimore City Public Schools

Two classroom-based preventive interventions were carried out on an epidemiologically defined, varied population of children in a metropolitan area in the United States. This is a report of the short-term impact and specificity of the two interventions from fall through spring of first grade. The first intervention, the Good Behavior Game, was aimed at reducing aggressive behavior and shy behavior. Aggressive behavior has been shown to be an important developmental antecedent in first grade of later delinquency and heavy drug use, particularly when coupled

We acknowledge the contributions of the city of Baltimore, its families and children, and the administration, faculty, and staff of the Baltimore City Public Schools. In particular, we would like to thank Walter Amprey, Superintendent of Baltimore City Public Schools; Lillian Gonzales and Patsy Blacksheare, Deputy Superintendents; Juanita Lewis, Director; Denise Borders, Chief of Accountability; Robert Solomon, Director, Special Education; Carla Ford, Specialist, Early Childhood Education; Louise Fink, Coordinator, Social Work Services; Dale Parker-Brown, Director, Compensatory Education; Matthew Riley, Director of the Eastern District; and Willie Foster, Director of Middle Schools. We also thank Alice Brogden for manuscript control and production; Fionnuala Regan for editorial preparation; and Pamela Spencer and Maria Corrada-Bravo for their contributions to data analyses. We thank Alan Harris for contributing to the development of the GBG intervention, and Lisa Crockett for her help in developing the Peer Assessment Inventory.

The studies on which this article is based have been supported by the following grants, with supplements from the National Institute on Drug Abuse: National Institute of Mental Health (NIMH) Grant No. P50 MH38725, Epidemiologic Prevention Center for Early Risk Behavior; NIMH Grant No. 1R01 MH42968, Periodic Outcome of Two Preventive Trials; NIMH Grant No. 1R01 MH40859, Statistical Methods for Mental Health Preventive Trials.

Correspondence and requests for reprints should be sent to Lawrence Dolan, Department of Mental Hygiene, School of Hygiene and Public Health, 624 North Broadway, Baltimore, MD 21205.

with shy behavior. The second intervention, Mastery Learning, was designed to improve poor reading achievement, which has been shown to be an antecedent for later depressive symptoms, as well as a correlate of aggressive and shy behaviors. Each of the two interventions had a significant and very specific impact only on its own proximal target(s). In addition to main effects, there were theoretically important variations of impacts among subgroups of children. The Good Behavior Game appeared to have a greater impact in reducing aggressive behavior among the more aggressive children. The nature of the impact of Mastery Learning differed by gender, with female high achievers benefiting more from the intervention than female low achievers, and male low achievers benefiting more than male high achievers. Developmental epidemiologically based preventive trials provide a powerful means of addressing questions about etiology and development, particularly around the issue of the malleability of developmental processes. Important questions that future work could test are whether achievement is improved by improving aggressive or shy behaviors and whether aggressive or shy behaviors are improved by improving achievement. Such investigation would inform our understanding of their etiology.

The study of developmental paths leading to psychopathology or to psychological well-being, and to social maladaptation or to social adaptation is a key area of concern in prevention research. Specification of earlier risk factors can provide well-defined targets for preventive trials. Such targets may reside in the individual, in the environment, or in the interaction between the environment and the individual's behavioral, psychological, and physiological responses. The trials are designed to develop effective preventive programs; but they also serve important theory-building purposes by determining the malleability of the target developmental processes. They indicate whether changing the early risk condition is possible and whether this lessens the risk of problem outcomes later in life (Kellam & Rebok, 1992). Such preventive trials thereby test the functional relationships of the proximal target to the more distal developmental outcome.

The Johns Hopkins Prevention Research Center has carried out two intervention strategies aimed at reducing risk behaviors within a classroom context. The interventions are directed at specific antecedents that have been shown to predict specific later problem outcomes. The Good Behavior Game (GBG) intervention is aimed at reducing aggressive and shy behaviors, which in turn would reduce antisocial behavior and heavy drug use among vulnerable children. The Mastery Learning (ML) intervention is aimed at increasing reading achievement within the classroom which would reduce psychiatric symptoms, particularly depressive symptoms, among vulnerable children.

In these preventive trials, the study populations and their environments are epidemiologically defined and deliberately varied so that the results can be examined as they apply to specific populations, and the impacts of the trials can be studied to determine if they differ for subpopulations and under varying conditions. The trials tested the two interventions in parallel for two purposes: to test the impact and the specificity of the two interventions, and to test whether achievement is improved by improving aggressive or shy behaviors and whether aggressive or shy behaviors are improved by improving achievement. The

second purpose would inform our understanding of the etiology of aggressive behavior, shy behavior, and poor school achievement. This article focuses on the first of these goals.

A community epidemiological and life-course developmental framework has guided the Johns Hopkins prevention research (Kellam, 1990; Kellam & Ensminger, 1980; Kellam & Rebok, 1992; Kellam & Werthamer-Larsson, 1986; Kellam et al., 1991). The integration of community epidemiology and life-course development we have termed developmental epidemiology, and we have found the resultant concepts and methods very useful in mapping early antecedents of later problem outcomes. Developmental epidemiology allows us to study variation in the rates of antecedents in defined populations over time and stages of life, and to study variation in the function of antecedents within subgroups of individuals in the same population.

The concepts of *social adaptation* and *social adaptational status* (SAS) provide a basis for searching for antecedents, particularly in the social task demands and behavioral responses of individuals in specific social fields and stages of life. Social adaptation postulates that an individual is involved in a few major social fields at each stage of life, such as the family, school and classroom, and peer groups in the childhood stage. Natural raters within each field (e.g., the teacher in the classroom, the classmates/peers in the peer group, the parents in the family) define the tasks to be performed and rate formally or informally the adequacy of the individual's performance. SAS is the rating or grade given by the natural rater (Kellam, 1990; Kellam, Branch, Agrawal, & Ensminger, 1975; Kellman & Ensminger, 1980; Kellam & Rebok, 1992).

SAS focuses on the basic interface between individuals and their immediate environment, and is a social status measure. By contrast, *psychological well-being* (PWB) refers to an individual's internal state and includes symptoms such as depression, anxiety, bizarre thinking or feeling, self-esteem, and neuropsychological or physiological status. Its measures are inferred rather than direct and reflect conditions internal to the individual.

The study presented here examines the short-term impact and specificity of the GBG and ML interventions on their proximal targets. Impact is determined on the basis of teacher ratings and peer nominations of student aggressive and shy behaviors, and standardized reading achievement test scores. Each of these are SAS measures reflecting the children's social adaptation to the classroom in the case of the teachers' ratings and test scores, and social adaptation to the peer group in the case of classmate/peer nominations.

The choice of first grade as a context for intervention is consistent with empirical evidence as well as with our developmental epidemiological perspective (Kellan, 1990). Until very recently in many urban areas in the United States, entry to first grade provided the first registry of all children in a defined population after the birth certificate. This is an important epidemiological criterion for choosing first grade because it avoids selection bias in the study population.

From the developmental viewpoint, the transition to first grade entails adaptation to a new social field and a new and potentially difficult set of social, behavioral, and cognitive developmental tasks. Moreover, empirical evidence from epidemiologically defined and other studies is rapidly accumulating on the lasting importance to later SAS and PWB of early success or failure in social adaptation to the early elementary school.

The Woodlawn studies (Ensminger, Kellam, & Rubin, 1983; Kellam, Branch, Agrawal, & Ensminger, 1975; Kellam, Brown, & Fleming, 1982; Kellam, Brown, Rubin, & Ensminger, 1983) demonstrate that as early as the start of elementary school, there are clearly identifiable antecedents leading to specific outcomes of psychopathology, delinquency, and heavy drug use. The findings on early aggressive behavior, particular when combined with shy behavior, are directly relevant to the current preventive interventions. First-grade aggressive behavior by boys was a predictor of increased teenage delinquency and of drug, alcohol, and cigarette use. Similar findings have emerged in many other studies (Conger & Miller, 1966; Kaplan, 1980; Lefkowitz, Eron, Walden, & Huesman, 1977; Mitchell & Rosa, 1981; Robins, 1978; Spivak, Marcus, & Swift, 1986; Tremblay, Masse, Perron, LeBlanc, Schwartzman, & Ledingham, 1992).

In Woodlawn and in other studies, the combination of early shy and aggressive behaviors in first-grade boys was associated with higher levels of delinquency and substance use than aggressive behavior alone. This shy/aggressive combination takes the form of children who are loners but who break rules and fight as well. It is very much like the DSM-III undersocialized conduct disorder (Block, Block, & Keyes, 1988; Ensminger et al., 1983; Farrington, Gallagher, Morley, St. Ledger, & West, 1988; Farrington & Gunn, 1985; Hans, Marcus, Henson, Auerbach, & Mirsky, 1992; Kellam et al., 1983; McCord, 1988; Schwartzman, Ledingham, & Serbin, 1985). Similarly, learning problems have been found to predict psychiatric distress, particularly depressed mood (Kellam, 1990; Kellam et al., 1983; Kohlberg, Ricks, & Snarey, 1984; Shaffer et al., 1979). Together, poor achievement and early aggressive behavior are important antecedents of school dropout as well (Ensminger & Slusarcick, 1992). These results focus attention on the classroom as a crucially significant social field.

Much of the literature on the prediction from early aggressive behavior to heavy substance use and antisocial behavior utilizes teacher ratings of aggressive/disruptive maladaptive behavioral responses or school records of behavior. However, a growing literature on the effect of peer rejection (Asher & Dodge, 1986; Bierman, 1986; Bierman, Smoot, & Aumiller, 1987; Coie & Dodge, 1988; Dodge, 1983), sometimes but not always coupled with aggressive behavior, suggests that maladaptation to the peer group is an important factor contributing to the vulnerability of the child. This suggests that both peer ratings and teacher ratings should be employed for evaluation of short-term impact of the GBG intervention.

Given the importance of early aggressive and shy behaviors in predicting later

problem behaviors, as well as the implications of aggressive and shy behaviors for a productive learning environment, many strategies have been proposed to alter early aggressive behavior and to a lesser degree shy behavior in the classroom. The most popular strategy is some form of behavior modification targeted at students exhibiting aggressive or shy behaviors. If implemented well, these techniques generally have short-term impact (Duke & Meckel, 1984; Hawkins, VonCleve, & Catalano, 1991; Medland & Vitales, 1984; O'Leary & O'Leary, 1977).

Even with positive results, however, few teachers are willing to commit themselves to individualized behavior modification strategies. Major deterrents are a lack of reinforcement options and the amount of planning, record keeping, and management skills required to keep them in place over the long term and to involve multiple students in the classroom. Heavy emphasis on extrinsic reinforcers is problematic in the classroom; yet the availability of natural reinforcers is often limited within the typical classroom context. Another factor associated with the size and duration of impact of behavioral interventions is the level of training teachers receive and their attitude toward the training efforts (Elliott, 1988; Robinson & Swanton, 1980). It is important that teachers become well versed in behavioral principles as well as the specific technology of a particular intervention protocol. perhaps the most problematic issue with individualized strategies is the lack of convincing evidence on the extent to which short-term gains generalize across time, across other aspects of performance (that is, achievement), and across other settings.

The GBG intervention addresses these problems by reducing the burden placed on the teacher for implementation, focusing on risk behaviors in classrooms and not at-risk individuals, structuring plans for generalization, and measuring impact in time and at least over the classroom and classmate/peer contexts. In prior studies, the GBG strategy proved to be an effective change agent by altering levels of out-of-seat and talking-out behaviors in fourth-, fifth-, and sixth-graders (Barrish, Saunders, & Wolfe, 1969; Medland & Stacknik, 1972). In subsequent years there have been numerous replications of the GBG strategy. Harris and Sherman (1973) used the GBG to reduce disruptive talking and out-of-seat behavior of fifth- and sixth-grade students. As before, the GBG was very effective in reducing targeted behaviors, and in this experiment they found a small concomitant improvement in school achievement.

Mastery Leaning is a teaching strategy that appears to improve school achievement, although there continues to be disagreement about its overall effectiveness (Kulik, Kulik, & Bangert-Drowns, 1990; Slavin, 1990). The theory and research underlying this strategy specify that under appropriate instructional conditions, virtually all students will learn most of what they are taught (Block & Burns, 1976; Dolan, 1986; Doyle, 1977; Guskey, 1985; Kulik et al., 1990). Mastery Learning individualizes learning by clearly defining what students should learn and the expected level of competence, by giving them additional

time to learn, and by helping them when and where they have learning difficulties.

The benefits of ML strategies in improving school achievement have been documented for limited time periods (Block & Burns, 1976; Kulik et al., 1990). The likelihood of success is increased if systematic staff development and monitoring of implementation are key components of the intervention process (Guskey, 1985). Mastery Learning appears to be especially successful when introduced in the sequential, skill-based curricula in the early elementary school years.

The GBG and ML are universal interventions in this study because they were applied to all children in the population (Gordon, 1983), not those already at increased risk. The potential of these strategies as public mental-health-oriented preventive interventions focusing on a large epidemiologically defined population over long periods of time has yet to be determined.

This article first reports data on the effectiveness of the GBG strategy for reducing aggressive and shy behavior in first-grade children. We hypothesized that the GBG intervention would reduce aggressive behavior from fall to spring of first grade. Reductions in shy behavior also were predicted because the GBG intervention used in this study incorporates strategies aimed at this target behavior as well. The ML intervention was expected to lead to improved reading achievement from fall to spring.

The data are taken from a cohort of first-graders who participated in the Johns Hopkins Prevention Research Center's first longitudinal preventive intervention trial administered at the classroom level and involving more than 2,000 children over the course of first and second grades. Gender-related effects were examined because gender differences in the levels of and in the interrelationships among aggressive, shy behavior, and achievement patterns were found in the developmental baseline studies (Kellam et al., 1991) and could affect intervention impact over the course of first grade.

METHOD

Design and Sample

The GBG and ML interventions were carried out in the eastern half of Baltimore City. From within the overall area, five varied urban areas were selected. Some areas exhibited many of the characteristics of community decay and poverty that are associated with high risk for problem behavior. Some were more middle-class areas that had community organization and access to resources associated with lower risk.

The three or four most closely matched schools within each area were identified, yielding a total of 19 schools. Matching characteristics included students' achievement levels, family socioeconomic status, and ethnicity. Within each of the five urban areas, the three or four most similar schools were identified for

random assignment to three intervention conditions: one to GBG, one to ML, and the other one or two schools to an external control condition with no experimental intervention activities.

First-grade classrooms within the intervention schools were assigned to the interventions at random. In addition, the school principals agreed to assign randomly the entering first-grade children to classroom membership to obtain balanced distributions for gender, preschool and kindergarten (K-level) experience, K-level conduct grades, and K-level achievement test scores. Therefore, both classrooms and students were randomly assigned to intervention status.

The design for a preventive trial of this type must provide for an estimate of leakage or spillover effects that might happen if all or part of the intervention strategies were adopted in the comparison classrooms. These problems were addressed in the research design by having both internal control classrooms that did not receive any special intervention within the GBG and ML intervention schools as well as external control classrooms within the nonintervention schools. This design also controlled for school-level effects.

The total sample for the first phase of the intervention research consisted of the cohort of children who entered the first grades of the involved schools during the 1985–86 academic year. The sample was 49% male. Sixty-four percent of the sample was African-American, 29% white, and 7% represented other ethnic groups. All students remained in the same design condition for one year and had to be available to receive teacher ratings in the fall of first grade to be included here. A total of 864 students met these criteria for the two interventions. The sample for the GBG condition comprised 182 students from 8 classrooms; the sample for the GBG internal control condition comprised 107 students from 6 classrooms; the sample for the ML condition comprised 207 students from 9 classrooms; the sample for the ML internal control condition comprised 156 students from 7 classrooms; and the sample for the external control condition comprised 212 students from 12 classrooms.

The Good Behavior Game Intervention

The Good Behavior Game is a classroom team-based behavior management strategy that promotes good behavior by rewarding teams that do not exceed maladaptive behavior standards. After baseline measurement of behavior, children were assigned to one of three heterogeneous teams in each classroom and points were given to the team for precisely defined aggressive behavior. The teacher determined team divisions and team membership, making sure that teams equally contained aggressive/disruptive children and social isolates. The goal of the strategy was to encourage students to manage their own and their teammates' behavior through the process of the group activity and through mutual self-interest. Details of the training and the intervention strategy are provided in the *Good Behavior Game Training Manual* (Dolan, Turkkan, Werthamer-Larsson, & Kellam, 1989). The eight teachers who participated received 40 hs of training.

While the GBG was in progress in the classroom, the teacher assigned a checkmark on the blackboard next to the name of a team whenever one of its members displayed one of the specified inappropriate or disruptive behaviors. Disruptive behaviors were verbal disruption, physical disruption, out of seat without permission, and noncompliance. A team could win the game if the total number of team checkmarks did not exceed four at the end of the game period. Thus, all teams could win during a particular game period. Initially, children on the winning teams received a tangible reward (stickers, erasers), and later they engaged in a rewarding activity (extra recess, class privileges). In addition, teams who won the most games during the week were termed the Weekly Winners and received a special reward on Friday.

The second aim of the GBG intervention was to alter not only aggressive behavior but also shy behavior. A tactical problem here was that we did not want to label shy children and isolate them for special intervention, inasmuch as the GBG should work on the basis of group contingencies. One way this problem was approached was to appoint the shy children on each team to the Team Leader position or to other socially integrating roles within the team. The Team Leader was in charge of handing out prizes, putting the star on the scoreboard, and helping the teacher with activities for his or her winning team.

During the first weeks of the intervention, the GBG was played three times each week for a period of 10 min. Over successive weeks duration increased approximately 10 min per game period every 3 weeks, up to a maximum of 3 hrs, although the checkmark criterion for winning the game remained at 4. Initially, game periods were announced and the rewards were delivered immediately after the game. Later, the teacher initiated the game period without announcement and the rewards were delayed until the end of the school day or week. Over time the game was played at different times of the day, during different activities, and even in different locations (such as in the hallway walking to the cafeteria). In this manner, the GBG evolved from a procedure that was highly predictable and visible with a number of immediate reward props to a procedure with an unpredictable occurrence and location with deferred rewards.

The Mastery Learning Intervention

The ML intervention consisted of an extensive and systematically applied enrichment of the reading curriculum. Intervention teachers received 40 hrs of training from a member of the Prevention Center and a member of the Baltimore City Curriculum Office, as well as much-augmented ML instructional strategies and curriculum materials, and the flexibility to adjust the curriculum timing such that a much higher percent of children reached mastery than in the regular curriculum.

Key elements in the ML interventions were clear statements of instructional goals and objectives; communication of high expectations for success; small, sequenced instructional units; use of formative and summative testing; use of a

variety of corrective methods until mastery was reached; presentation of mastery standards for each instructional unit; immediate feedback to students; and clear and updated records of student progress.

Critical aspects of the training program included a group-based approach to mastery, a more flexible corrective process, and material support. The development of a more group-paced approach to mastery meant that students did not proceed to the next learning unit until the majority (80% achieving 80–85% of objectives) of students had fulfilled the learning objectives of the previous unit.

The systematic provision and use of correctives constitutes another key to the effectiveness of ML. Corrective instruction, by its very nature, must be targeted toward particular learners and particular learning difficulties. The corrective process developed for this intervention was flexible in terms of time, grouping strategy, and variety of correctives. A systematic linking of objectives, test items, error prescriptions, and correctives was developed and reinforced through an instructional management model. Further details about the training are contained in *The Mastery Learning Manual* (Dolan, Ford, Newton, & Kellam, 1989).

Measures

Teacher ratings, peer nominations, and a standardized achievement test were used to measure SAS in fall and spring of first grade in both intervention and control classrooms. The use of both teacher and peer measures to examine intervention impact is critical in that the teacher rating is completed by the main intervention agent.

Teacher Observation of Classroom Adaptation-Revised (TOCA-R). The TOCA-R is a measure of each child's adequacy of performance on the core tasks in the classroom as rated by the teacher. It was originally developed and used in the Woodlawn studies (Kellam et al., 1975), and after modification it has been used as a major periodic assessment instrument for the prevention trials (Werthamer-Larsson, Kellam, & Wheeler, 1991). Briefly, it involves a structured interview administered by a trained member of the staff who follows a script precisely, responds in a standardized way to issues the teacher initiates, and records the teacher's ratings of the adequacy of each child's performance on three basic tasks: social participation (the maladaptive form being shy behavior); accepting authority (the maladaptive form being aggressive behavior); and concentration and being ready for work (the maladaptive form being inattention or having concentration problems). These three tasks plus achievement are the four core tasks in the classroom as the teacher views the classroom. The TOCA-R was administered in the fall and spring of first grade. Teachers completed the interview during the school day in a private place. Teacher ratings of the two factors Authority Acceptance (aggressive behavior) and Social Contact (shy behavior) were used to analyze intervention impact. Psychometric properties of the rating

were reported by Werthamer-Larsson et al. (1991). Higher scores indicate more aggressive behavior and more shy behavior.

Peer Assessment Inventory. The Peer Assessment Inventory is a measure based on each classmate's nominations of each other child in the classroom. It is a classroom-administered modified version of the Pupil Evaluation Inventory (PEI; Pekarik, Prinz, Leibert, Weintraub, & Neale, 1976). Ten items were selected from the original PEI on the basis of their relevance to three SAS constructs: aggressive behavior, shy behavior, and likability. Six items from the peer assessments were used to determine the impact of the GBG: three peer ratings of aggressive behavior (*fights, gets into trouble, is mean to others*) that work well psychometrically as a single score and three peer ratings of shy behavior (*is too shy, plays alone, has few friends*) that do not function as well as a single measure. The internal consistency of the aggressive behavior scale as assessed by standardized alpha coefficients is .87, and the internal consistency of the shy behavior scale is .74.

California Achievement Test-Reading. Reading achievement was assessed by the total reading score (standard scores) from the California Achievement Test (CAT, Forms E and F) administered during the fall and spring of the first grade. The CAT was administered to entire classes as part of the scheduled testing program in the Baltimore City Public Schools. Test scores were obtained by the Prevention Center via magnetic data transfer, with both error and reliability checks.

RESULTS

The means and standard deviations are displayed in Table 1 for teacher-rated and peer-rated aggressive behavior and shy behavior in the fall and spring of first grade across the three intervention conditions by gender. Those for reading achievement scores across the three intervention conditions by gender are displayed in Table 2.

Analysis of covariance (ANCOVA) methods were used to test the effects of the interventions on their proximal targets within gender. Boys and girls are treated separately because the temporal regression of spring values on fall values differs significantly in slope for almost all the key variables. Pooling the genders sharply violates the assumptions of analysis of covariance. The interventions were compared separately to each of their controls because the control groups were not considered to be poolable. The differences between the control groups, which in some cases were quite large, were not allowed to confound the comparison of the intervention with each control.

The ANCOVA provides a test of the interventions on the central tendency, controlling for the level of severity of the targets at fall of first grade. The

TABLE 1
Descriptive Statistics by Intervention Condition for Aggressive Behavior and Shy Behavior for First-Grade Boys and Girls

	Good Behavior Game		Internal Control		External Control	
Variable	Boys	Girls	Boys	Girls	Boys	Girls
Teacher-Rated Aggressive Behavior						
Fall, Grade 1						
M	2.46	1.66	2.27	1.78	2.05	1.67
SD	1.37	0.79	1.06	0.72	1.06	0.76
Spring, Grade 1						
M	2.16	1.52	2.12	1.88	2.23	1.66
SD	1.14	0.71	0.68	0.79	1.03	0.71
Teacher-Rated Shy Behavior						
Fall, Grade 1						
M	2.49	2.45	2.62	2.36	2.86	2.71
SD	1.03	1.02	0.78	0.72	1.01	1.05
Spring, Grade 1						
M	2.29	2.21	2.81	2.46	2.55	2.66
SD	1.05	0.98	0.81	0.77	0.93	0.87
Peer-Rated Aggressive Behavior						
Fall, Grade 1						
M	0.28	0.12	0.26	0.17	0.25	0.15
SD	0.20	0.08	0.17	0.10	0.16	0.13
Spring, Grade 1						
M	0.28	0.13	0.32	0.18	0.30	0.17
SD	0.15	0.11	0.15	0.11	0.13	0.10
Peer-Rated Shy Behavior						
Fall, Grade 1						
M_1	0.22	0.15	0.15	0.21	0.18	0.21
M_2	0.20	0.11	0.16	0.19	0.16	0.13
M_3	0.21	0.21	0.22	0.25	0.23	0.22
SD_1	0.17	0.09	0.10	0.10	0.10	0.11
SD_2	0.19	0.08	0.13	0.13	0.12	0.10
SD_3	0.14	0.15	0.11	0.13	0.12	0.12
Spring, Grade 1						
M_1	0.15	0.20	0.17	0.22	0.16	0.20
M_2	0.18	0.15	0.19	0.20	0.19	0.14
M_3	0.19	0.13	0.22	0.22	0.23	0.18
SD_1	0.11	0.10	0.11	0.13	0.11	0.11
SD_2	0.12	0.10	0.11	0.11	0.12	0.11
SD_3	0.15	0.12	0.14	0.12	0.15	0.13

Note. M_1 = mean for Shy 1 (too shy), M_2 = mean for Shy 2 (play alone), M_3 = mean for Shy 3 (few friends); SD_1 = standard deviation for Shy 1 (too shy), SD_2 = standard deviation for Shy 2 (plays alone), SD_3 = standard deviation for Shy 3 (few friends).

TABLE 2
Descriptive Statistics by Intervention Condition for Reading Achievement for First-Grade Boys and Girls

	Mastery Learning		Internal Control		External Control	
Variable	Boys	Girls	Boys	Girls	Boys	Girls
			Reading Achievement			
Fall, Grade 1						
M	256.25	260.12	264.22	265.32	246.07	248.77
SD	35.68	32.16	34.24	39.03	31.92	33.69
Spring, Grade 1						
M	316.93	321.81	310.14	318.35	295.11	302.18
SD	37.72	34.35	41.11	38.61	44.06	36.94

underlying models and related scatterplots also provide an opportunity to examine variation in impact as a function of differences in severity as assessed in fall of first grade. This examination requires assessing differences in slopes, in residual variances, and in residual patterns (Laird, 1983). We will present such plots for core results. These presentations contain the scatterplot of the children in the intervention groups and each of the control groups, with the individual children indicated by a square. Those in the periphery marked by letters instead of squares indicate internal (I) or external (E) control, or GBG (G) or ML (M) intervention. These are the outliers who fell outside the 90% prediction interval from regression analyses. The outliers are of concern because they included children who had sharply higher or sharply lower values in the spring than would have been predicted from their values in the fall.

Results of the Good Behavior Game Intervention

The ANCOVA results of GBG impact on aggressive and shy behaviors in the spring of first grade, controlling on fall aggressive and shy behaviors respectively, are displayed in Table 3 for boys and Table 4 (pp. 330–331) for girls.

Teacher Ratings of Aggressive Behavior. For both boys and girls, the GBG had a significant impact on aggressive behavior as rated by teachers, but the effect was gender-specific. For boys, the GBG group, compared with the external control, had significantly lower mean aggressive behavior (as rated by the teacher in the spring of first grade; $p < .05$; see Figure 1). The corresponding comparison between the GBG group and the internal control group was similar in direction but not significant. For girls, the GBG group, compared with the internal control, had significantly lower mean aggressive behavior (as rated by the teacher in spring of first grade; $p < .05$; see Figure 2). The corresponding

TABLE 3
Analyses of Covariance Results for Good Behavior Game Impact in Spring of First Grade for Boys

Measure/Source	*df*	*SS*	*F*
	Aggressive Behavior		
Teacher Ratings			
Fall aggressive	1	105.40	179.53***
External-GBG	1	2.27	3.87*
Error	194	113.90	
Fall aggressive	1	58.90	109.85***
Internal-GBG	1	.254	.474
Error	134	71.86	
Peer Ratings			
Fall aggressive	1	4.30	244.94***
External-GBG	1	.034	1.96
Error	165	2.89	
Fall aggressive	1	2.59	115.68***
Internal-GBG	1	.166	7.42**
Error	119	2.66	
	Shy Behavior		
Teacher Ratings			
Fall shy	1	97.23	212.44***
External-GBG	1	.535	1.17
Error	194	88.79	
Fall shy	1	51.04	100.82***
Internal-GBG	1	8.03	15.85***
Error	134	67.84	
Peer Ratings			
Shy 1 (too shy)			
Fall shy	1	.003	.215
External-GBG	1	.007	.623
Error	165	1.95	
Fall shy	1	.0004	.032
Internal-GBG	1	.009	.725
Error	119	1.45	
Shy 2 (play alone)			
Fall shy	1	.229	15.73***
External-GBG	1	.004	.253
Error	165	2.40	
Fall shy	1	.145	11.58**
Internal-GBG	1	.007	.582
Error	119	1.49	
Shy 3 (few friends)			
Fall shy	1	.040	1.74
External-GBG	1	.059	1.72
Error	165	3.81	
Fall shy	1	.059	3.11
Internal-GBG	1	.015	.793
Error	119	2.27	

* $p < .05$. ** $p < .01$. *** $p < .001$.

TABLE 4
Analyses of Covariance Results for Good Behavior Game Impact in Spring of First Grade for girls

Measure/Source	*df*	*SS*	*F*
	Aggressive Behavior		
Teacher Ratings			
Fall aggressive	1	50.43	212.67***
External-GBG	1	.47	1.99
Error	188	44.58	
Fall aggressive	1	43.42	155.67***
Internal-GBG	1	1.19	4.25*
Error	145	40.57	
Peer Ratings			
Fall aggressive	1	.70	66.39***
External-GBG	1	.01	1.34
Error	157	1.66	
Fall aggressive	1	.42	35.29***
Internal-GBG	1	.02	1.32
Error	127	1.52	
	Shy Behavior		
Teacher Ratings			
Fall shy	1	85.07	207.38***
External-GBG	1	3.86	9.42**
Error	188	77.12	
Fall shy	1	51.65	112.14***
Internal-GBG	1	3.50	7.60***
Error	145	66.78	
Peer Ratings			
Shy 1 (too shy)			
Fall shy	1	.01	.847
External-GBG	1	.00	.041
Error	157	1.83	
Fall shy	1	.03	1.62
Internal-GBG	1	.01	.70
Error	127	2.29	
Shy 2 (play alone)			
Fall shy	1	.05	3.99*
External-GBG	1	.02	1.86
Error	157	1.77	
Fall shy	1	.10	7.38**
External-GBG	1	.03	2.22
Error	127	1.66	

(continued)

TABLE 4
(Continued)

Measure/Source	*df*	*SS*	*F*
Shy 3 (few friends)			
Fall shy	1	.003	.20
External-GBG	1	.082	5.24*
Error	157	2.45	
Fall shy	1	.029	2.08
Internal-GBG	1	.222	15.94***
Error	127	1.77	

* $p < .05$. ** $p < .01$. *** $p < .001$.

comparison between the GBG group and the external control group was similar in direction but not significant.

For insight into the comparisons given here we examined, in an exploratory mode, the overall and within-group regression models embedded within the ANCOVA methods. We looked for variation in slope, variances, residual patterns, and outliers. For both boys and girls, the overall regression of teacher ratings of spring aggressive behavior on fall aggressive behavior was quite strong; $R^2 = .50$, $p < .0001$ for boys and $R^2 = .55$, $p < .0001$ for girls; and the

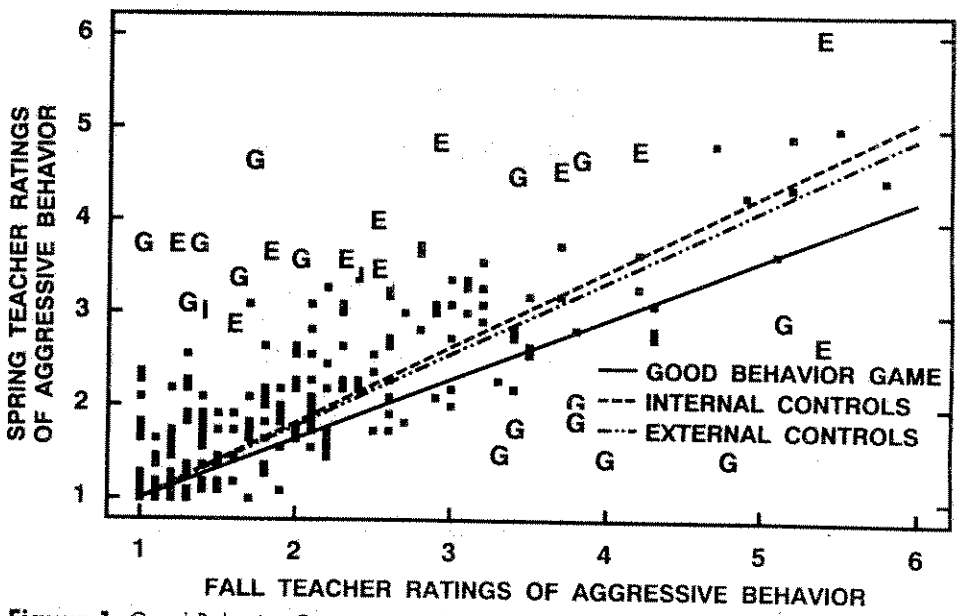

Figure 1. Good Behavior Game impact during first grade on teacher ratings of aggressive behavior for boys ($N = 243$).

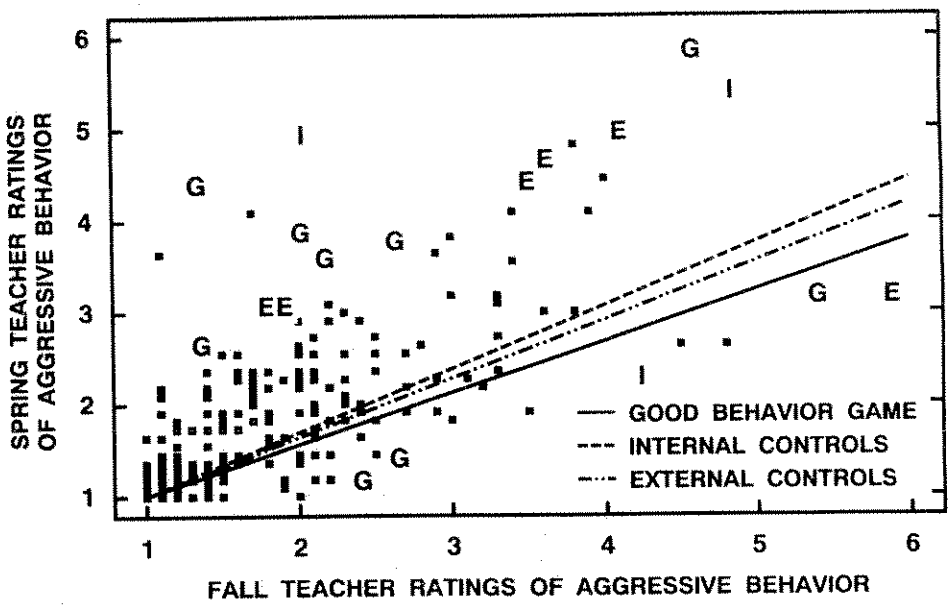

Figure 2. Good Behavior Game impact during first grade on teacher ratings of aggressive behavior for girls (N = 252).

residual distributions were quite normal. For both genders, the slope for the GBG group (b = .56 for boys, .66 for girls) was less than the slope for the external control group (b = .74 for boys, .75 for girls) or the internal control group (b = .72 for boys, .82 for girls); but the slope differences were not significant. The order of slopes suggests that the GBG may have had a greater impact in reducing aggressive behavior in children who began the year at the upper end of the distribution; that is, children with higher teacher ratings of aggressive behavior (see Figures 1 and 2).

The outlier pattern in Figure 1 shows that the majority of boys who were sharply less aggressive in the spring than would be predicted from the fall were in the GBG group. Only a single male lower outlier was from a control group. Some of these outliers may present particularly successful intervention cases and are targeted for further study. No other patterns such as sharp nonlinearity or heteroscedasticity were found in these residuals.

The outliers for girls in Figure 2 do not show much pattern. Both low and high outliers tend to be from both intervention and control groups. The residuals are approximately normal with no evidence of nonlinearity or heteroscedasticity.

A second method of assessing the effect of the intervention is to compare the means of the change scores from fall to spring. For both genders, for teacher ratings of aggressive behavior the mean change score was negative for the GBG groups (−.12 for boys and −.05 for girls) but positive for both control groups. The mean change scores for external control were .18 for boys and .04 for girls;

for internal control they were .06 for boys and .11 for girls. These differences were significant for boys ($p < .03$) but not for girls. In both cases, they were in the direction suggested by the ANCOVA results. Taken together, these results provide evidence that the GBG intervention reduced the level of aggressive behavior as rated by teachers.

Peer Nominations of Aggressive Behavior. Analyses of the peer nominations of aggressive behavior were based on the mean of the proportion of nominations on the three items: *fights a lot, gets into trouble, is mean to others*. Applying ANCOVA methods to the boys and controlling on fall peer nominations of aggressive behavior, the GBG group had lower mean aggressive behavior in the spring (as rated by peers, when compared with the internal control; $p < .01$; see Figure 3). The comparison with the external control group was in the same direction but not significant. For girls, mean level of spring aggressive behavior as rated by peers did not differ between the GBG group and either control group although the mean was lowest for the GBG group (see Table 1). The very low level of peer nominations of female aggressive behavior throughout first grade should be noted.

Again, we examined the regression model embedded within the ANCOVAs to explore the impact of GBG on slopes, variances, residual patterns, and outliers. For boys, the overall regression, ignoring intervention group, of peer nominations of spring aggressive behavior on fall was quite strong, $R^2 = .55$, $p < .0001$, as it was for teacher ratings. For girls, the overall regression,

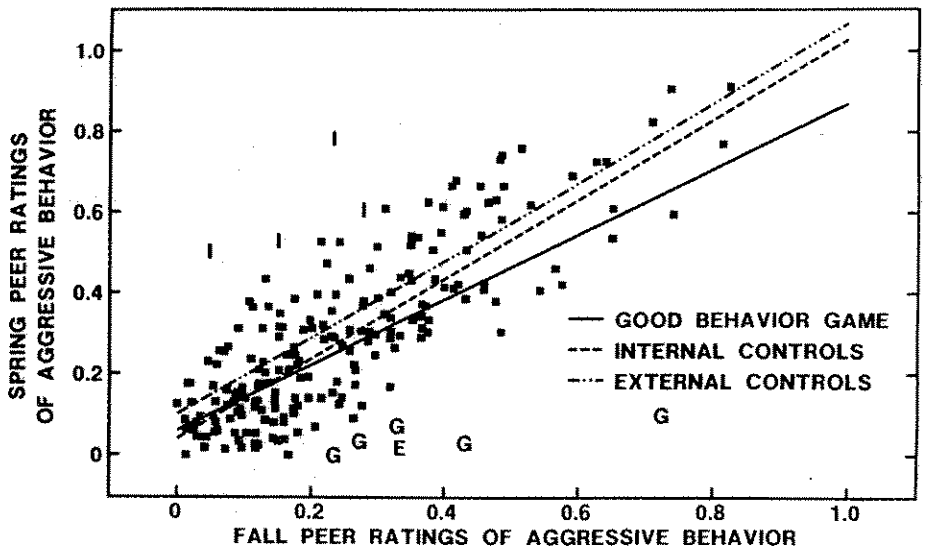

Figure 3. Good Behavior Game impact during first grade on peer ratings of aggressive behavior for boys ($N = 211$).

$R^2 = .28$, $p < .0001$, was considerably weaker for peer nominations than it was for teacher ratings. For both genders, the empirical residual distributions were quite normal.

For both boys and girls, the effects of GBG on the slope parameter for peer nominations of aggressive behavior were similar to the relationship for teacher ratings of aggressive behavior. In both genders, the slope for the GBG group ($b = .82$ for boys; .41 for girls) was less than the slope for the external control group ($b = 1.0$ for boys; .77 for girls) or the internal control group ($b = .97$ for boys; .70 for girls), but the slope differences were not significant. These slopes suggest that for both boys and girls the GBG may have had a greater impact in reducing aggressive behavior, as rated by peers, in children who began the year with higher peer nominations of aggressive behavior.

The outlier pattern in Figure 3 shows that the majority of boys who were much less aggressive in the spring than would have been predicted from the fall were in the GBG group. Only a single male low outlier was from a control group. All of the high outliers were from the internal control group.

For a second look at the effect of the interventions on peer nominations of aggressive behavior, we turn to the difference between spring and fall level of nominations. The mean change scores for peer nominations of aggressive behavior for boys was .02 for the GBG group, .05 for the external control group, and .10 for the internal control group. These differences were significant ($p < .01$) and give additional evidence that the GBG reduced the level of male aggressive behavior as nominated by peers. The mean change scores for girls in peer nominations of aggressive behavior increased by .01 for the intervention and both control groups, indicating no effect of the GBG on the change in female aggressive behavior as rated by peers. Again, the very low level of peer nominations of female aggressive behavior throughout first grade should be noted.

Teacher Ratings of Shy Behavior. As shown by ANCOVA methods, the GBG group, compared to the internal controls, had significantly lower mean shy behavior as rated by teachers in the spring of first grade ($p < .01$ for both genders). For girls, but not for boys, the corresponding comparison between the GBG group and the external control group was significant and in the direction hypothesized ($p < .01$).

For insight into the comparisons given here we examined the overall and within-group regression models embedded within the ANCOVAs. For both girls and boys, the regression of teacher ratings of shy behavior in the spring of first grade on the rating in the fall of first grade was strong, $R^2 = .44$, $p < .001$ for boys and $R^2 = .41$, $p < .001$ for girls). The slopes of the regression lines did not differ significantly for either boys or girls. For boys the slopes for GBG, internal control, and external control were .69, .36, and .66; for girls the slopes were .77, .29, and .52. The order of the slopes is not consistent with the hypothesis that the intervention tends to work at the more severe end of the scale. In contrast to the

pattern for teacher ratings of aggressive behavior, the residuals and outliers for teacher ratings of shy behavior showed little pattern.

Again we supplement the ANCOVAs by assessing the effect of the intervention on the change scores from fall to spring. For the GBG group the mean of teacher ratings of shy behavior in the spring was lower than the mean rating in the fall (M difference = $-.29$ for boys and $-.21$ for girls). In contrast, the mean rating for spring shy behavior was higher for the internal control group (M difference = .21 for boys and .11 for girls). For boys, the external control group mean change was very similar to the mean change for the GBG (M difference = $-.25$), whereas the external control group for girls revealed almost no change in the mean level of shy behavior from fall to spring (M difference = $-.01$). The differences in these change scores among the three groups were highly significant ($p < .001$ for both boys and girls). The pattern in these difference is not clear. All of these results taken together provide evidence that the GBG intervention reduced the level of shy behavior as rated by teachers.

Peer Nominations of Shy Behavior. These ratings were analyzed at the individual item level because the three nominations relating to shy behavior—*too shy, plays alone*, and *has few friends*—were only moderately intercorrelated (r = .22–.52 for the fall; r = .23–.39 for the spring). Applying ANCOVA methods, GBG compared with either control group did not reduce shy behavior of boys as nominated by peers on any of these three measures. For girls, the only significant effect was for *has few friends*. We concluded that GBG did not make significant impact on the peer nominations of shy behavior, conceivably because of the children's lack of recognition of the ideas reflected in the items measuring shy behavior. Analysis of the within-group regressions and the residual patterns produced no results of note.

Results of the Mastery Learning Intervention

Results of the ANCOVA of ML impact on reading achievement in spring of first grade, controlling on fall reading achievement, are displayed in Table 5. There was an affect of ML on reading achievement, but the effect varied by gender. For boys, there was a significant effect of ML on spring achievement compared with both the internal and external control groups (see Figure 4). For girls, the ML versus external control comparison was highly significant, with the comparison of ML with the internal control also being moderately significant ($p < .05$) (see Figure 5).

For insight into these comparisons we explored the overall and within-group regression models embedded within the ANCOVAs. For boys, as for girls, the overall regression of spring reading achievement on fall reading achievement was strong; $R^2 = .41$, $p < .0001$ for boys and $R^2 = .40$, $p < .0001$ for girls. In boys, the slope of the ML group ($b = .69$) was less than the slope for the internal control ($b = .76$) or the external control ($b = .89$). This result provides further

TABLE 5
Analyses of Covariance Results for Mastery Learning Impact in Spring of First Grade

Reading Achievement/Source	*df*	*SS*	*F*
Boys			
Fall reading achievement	1	152,615.4	151.46***
External-ML	1	9,909.5	9.83**
Error	200	201,525.3	
Fall reading achievement	1	114,503.3	123.18***
Internal-ML	1	7,015.1	7.55**
Error	189	175,687.4	
Girls			
Fall reading achievement	1	115,503.9	143.83***
External-ML	1	8,906.6	11.09**
Error	222	178,276.4	
Fall reading achievement	1	120,494.2	161.54***
Internal-ML	1	3,095.3	4.15*
Error	216	161,116.5	

* $p < .05$. ** $p < .01$. *** $p < .001$.

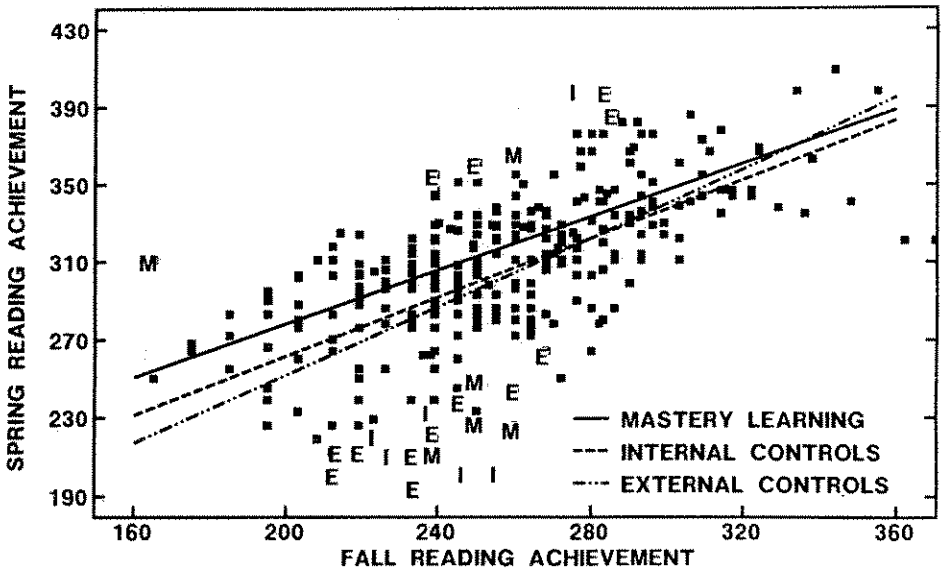

Figure 4. Mastery Learning impact during first grade on reading achievement for boys (N = 297).

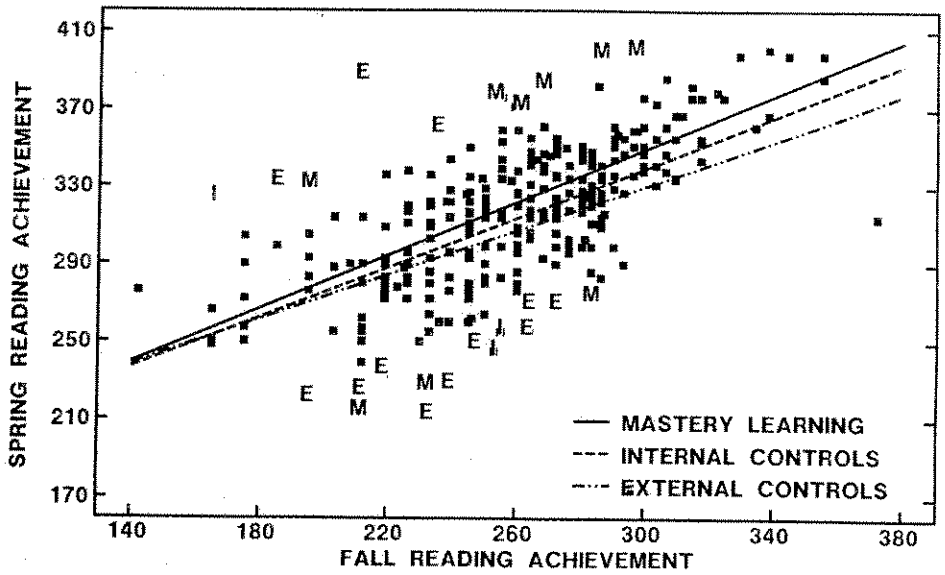

Figure 5. Mastery Learning impact during first grade on reading achievement for girls (N = 312).

evidence that the ML intervention primarily benefited boys at the lower end of the achievement continuum. For girls, the slope for the ML group (b = .69) was greater than the slope for the internal control b = .65) or the external control (b = .58). This result further indicates that the ML intervention worked at the higher end of the achievement continuum for girls.

We next examined the design conditions of the outliers from the common regression (see Figures 4 and 5). As shown in Figure 4, for boys there is no pattern for the upper outliers, whereas the bottom outliers were mainly not in the ML group. This pattern suggests that ML helped boys by keeping them from dropping into the lower achievement groups. For girls, there appears to be a trend that the majority (6–10) of the high outliers were in the ML group, whereas there is no clear outlier pattern among the low outliers. The pattern suggests that ML enhanced achievement among already high-achieving girls by boosting them into the top achievement groups.

To test whether there was a differential impact of the ML intervention as a function of baseline achievement, we ran one-way analyses of variance (ANOVAs) on boys and girls, using the higher (> 250) and lower (< 250) fall achievement scores as a factor. For girls, the ML intervention appeared to work best with high achievers. For high-achieving girls, there were significant differences between the ML and internal control groups, $F(1, 336) = 5.89$, $p < .05$, and between the ML and external control groups, $F(1, 119) = 13.24$, $p < .001$. For the low achievers these comparisons were nonsignificant. For boys, the opposite pattern was found, with the ML intervention significantly benefiting the

low achievers, $F(1, 70) = 5.10$, $p < .05$ and $F(1, 86) = 13.93$, $p < .001$, for the ML internal control and ML external control comparisons, respectively. For boys, there was little evidence that ML benefited high achievers.

Finally, we translated the standard reading achievement scores into reading grade equivalent scores reflecting the actual reading competence of the groups as indexed by years and months of reading growth. At the end of one year of intervention, the ML students were reading at a grade equivalent (GE) equal to 1 year and 9 months of reading competence (GE = 1.9) compared with GE = 1.7 for both the internal and external control conditions (a difference of approximately 2 months).

Specificity of Intervention Impact

The impact of the GBG was predicted to be specific to aggressive and shy behaviors, its own proximal targets, and not significantly affect achievement, the proximal target of ML. Likewise, it was hypothesized that ML would affect achievement but would not produce a significant impact on aggressive or shy behaviors. ANCOVA methods, controlling on baseline measures of the proximal target, were used to test for these specificity effects. A main effect of GBG on achievement or a main effect of ML on aggressive or shy behaviors would be evidence of nonspecificity. The results indicate a high degree of specificity for both the GBG and ML interventions.

The ANCOVA results for ML yielded only one significant effect, and that was only for girls and only in the case of teacher-rated shy behavior. Both the internal and external control conditions differed significantly from the ML conditions ($p < .05$). These results indicate that there was a benefit of the ML intervention only on shy behavior for girls. This analysis tests for attention or nonspecificity and should not be confused with indirect or crossover effects. These effects, which include change in the proximal target by one intervention as a condition for change in the proximal target of the other, will be tested in future analyses.

DISCUSSION

The results for the GBG point to a significant and highly specific effect of that intervention for both boys and girls on aggressive behavior as rated by teachers. For peer nominations of aggressive behavior, there was a gender difference. The GBG had an impact for boys on aggressive behavior as nominated by peers, but it had no effect on peer nominations of aggressive behavior among girls. This result needs further study to assess its validity and meaning. It suggests that first-graders may attribute aggressive behavior or change in aggressive behavior differently among girls than they would among boys. The frequencies of peer-rated aggressive girls were markedly lower even before the GBG. For both boys and girls, GBG had the greatest impact on children who began the school year with more aggressive ratings by teaches. These results support the usefulness of the

GBG as a low-cost intervention that can alter aggressive behavior among more aggressive children.

The GBG had an impact on shy behavior as rated by teachers for both boys and girls. However, its impact on peer nominations of shy behavior was limited to the shy behavior item *has few friends* for girls. The very low correlation between fall and spring peer nominations of shy behavior may have attenuated the ability of the GBG intervention to affect the peer nominations of this variable. Overall, the regression patterns were very similar for boys and girls, with strong regressions (spring first grade on fall first grade) for aggressive behavior as rated by teachers and peers, and for shy behavior as rated by teachers.

The results for the ML intervention indicate that it raised reading achievement test scores for both boys and girls, at least through the end of first grade. In line with previous ML studies (see Kulik et al., 1990), we found evidence that low-achieving male students actually gained more from ML than high-achieving male students. For girls, the opposite appeared to be true, with high achievers benefiting more from ML than low achievers. Thus, ML provided a fairly economical intervention that altered the early achievement trajectory for the majority of students involved in the classroom intervention.

In summary, the GBG was found to be an effective intervention for reducing aggressive behavior among first-grade children, and the ML intervention was found to be an effective intervention for improving reading achievement among first-grade children.

For the analysis of the overall effects of the GBG and ML interventions, we made separate comparisons with the internal and external controls. Internal controls permit us to hold school effects constant in comparing the effects of the intervention, whereas external controls can be used to help determine whether the internal controls show evidence of spillover effects. The possibility of spillover effects would be supported by significant differences between the external control classrooms and the two interventions but not between internal control classrooms and the interventions. Teacher ratings of male aggressive behavior reflected possible spillover from the GBG classrooms to the internal control classrooms. There was no clear evidence of spillover effects for boys or girls for shy behavior or for achievement.

The specificity of the GBG and ML interventions on their proximal targets is important for both theory building and intervention planning. Aggressive behavior and shy behavior can be impacted directly by the GBG and achievement can be targeted by ML. Results indicate that the GBG does not directly influence achievement and the ML does not directly influence aggressive behavior. For girls only, there was weak evidence that ML improved shy behavior as rated by teachers. The relationships between these proximal targets and distal ones (such as depression) are equally important for the baseline developmental model and intervention planning and are the subject of current analyses.

The use of two epidemiologically based interventions in a parallel design was

intended not only to test the specificity of effects of each intervention on its proximal target, but also to examine the crossover effects; that is, the effect of changing the proximal targets of each intervention on the proximal target of the other. Currently we are examining the crossover effects of changing reading achievement produced by the ML intervention on aggressive and shy behaviors and the crossover effects of changing aggressive and shy behaviors produced by the GBG intervention and reading achievement. This approach exposes the developmental direction, if any, in the correlation of aggressive and shy behaviors with achievement, by improving achievement to see if aggressive and/or shy behavior improved, and by improving aggressive and/or shy behavior to see if achievement improved.

Certain limitations should be noted as the results of this study are reviewed. Although teacher ratings are considered an important measure of SAS, they are confounded by the fact that the teacher was the main intervention agent and was knowledgeable about specific outcome targets. Therefore, the internal validity of the teacher ratings is at issue. Furthermore, because teachers in the intervention school knew they were in a contrast condition, they may have wanted to improve the rating within their classroom. Because this study also included peer ratings to measure impact, this concern was substantially reduced.

Peer nominations are also not without potential bias. The peers within the GBG classrooms received the expectation of better behavior by their teacher, and hence may have inferred better behavior among their peers than existed in reality, thereby creating a self-fulfilling expectation of a better behaved class. Alternatively, they may have rated their peers more harshly because the cost of maladaptation increased with GBG. However, other investigators have questioned whether children at age 6 or 7 can draw valid inferences about each other (Younger, Schwartzman, & Ledingham, 1985, 1986). The results of our psychometric analyses support the inference that first-grade children can evaluate each other on constructs requiring attributions such as *is mean to others, fights a lot,* and *gets into trouble*. Peer nominations for these three aggressive behavior items had sufficient alpha and met criteria for validity; the shy behavior items, as has been found in other studies (Younger et al., 1985, 1986), did not comprise a single category. The impact results for peer nominations for shy behaviors should therefore be viewed conservatively.

The results of this study form the initial foundation for analyses that examine variation in specific early-risk behaviors, and variation in response to the interventions. What processes were at work during the interventions that might lead to the results obtained? Furthermore, what are the developmental processes and moderators that follow from this early stage of elementary school? The GBG intervention provided specific reinforcement for behaviors that were not aggressive or disruptive. The hypothesized process of change primarily involved aiming an intervention at the social adaptational process between teacher and children, including getting students to recognize teacher-defined maladaptive

behaviors and to respond to teacher-defined team contingencies and classroom rewards to reduce such behavior. A large part of the change in behavior was likely due to group pressure to change, but the intervention might have led to a greater sense of social awareness and perceptions of social competence. This latter factor might be critical over time, after the removal of the group contingencies in later grades. Current follow-up studies will be able to track this moderating factor of social competency over time and how it relates to various aspects of SAS and PWB.

The process of changing shy behavior was less targeted and received less attention over the course of the GBG intervention. Teachers were trained in the GBG to recognize withdrawn and shy children and to place them in positions of esteem during the game operations. By giving shy children leadership roles, perhaps the intervention provided nonthreatening opportunities for them to interact with others. Shy children also were more likely to be good team members because they were less likely to violate the rules of good behavior that were the focus of the intervention.

An alternative explanation for the positive effects of the GBG intervention on aggressive and shy behaviors is that the children received generally greater attention in the GBG condition compared with the two control conditions. However, the advantage of two preventive interventions being implemented in one design is that one intervention can control for general attention as the explanation of impact of the other intervention. Specifically, ML did not improve aggressive behavior, nor did GBG improve achievement. This result provides evidence that each intervention acted specifically on its proximal target.

One possible mechanism for the impact of ML on reading achievement is that the strategy attempts to reinforce more internal attributions for both success and failure. Students are given more responsibility for their own learning. When students fail at a learning task, they are given the feedback, resources, and time to correct errors before moving to the next unit of instruction. Over time, many of the instructional strategies that involve feedback and correctives, reinforcement, and regulation of time and resources may become more internalized and routinized within the style of learning in the classroom.

Corno and Rohrkemper (1985) characterized these sets of skills as self-regulated learning competencies. The skills include alertness to classroom cues, distinguishing task-relevant information, planning instructional routines, and self-monitoring for effectiveness. Within ML these self-regulated learning skills are consistently stressed to the learner, particularly one who might want to become disengaged from learning because of early failure experiences. One explanation for the achievement gains of the male students in the ML condition is that they were not permitted to become disengaged, the result being greater achievement. We hypothesize that over time the combination of improved achievement, perception of academic competence, and self-regulatory skills in the classroom will lead to the eventual reduction of depressive symptoms in these

students. The timing and level of this impact on depression will be a major component of future analyses.

The practical implications of these two interventions are that specific technologies are now available to improve first-grade reading achievement and reduce levels of aggressive behaviors. These technologies are low-cost universal interventions that could be immediately implemented in classrooms in many school districts.

Currently, we are investigating the long-term effects of the GBG and ML interventions over the elementary years and through the transition into middle school. Over time, the impact of the two interventions may become stronger because the possibility of synergistic effects increases. For example, reducing aggressive behavior and increasing attention to classroom tasks may increasingly make the child more available to the impact of other educational programs and interventions. In the next series of preventive trials, we plan to explore the possibility of synergistic effects by combining the GBG intervention with the ML intervention. As participants in the intervention are followed over time, it will be vital to identity possible moderators of impact that can explain variation in the responses of children to each intervention. Characteristics of the child, classroom, classmates, teachers, and families are all important aspects for studying variation in the children's developmental trajectories with and without the interventions.

Finally, investigating impact outside the classroom, in the home or in other school contexts, will help determine the generalizability of the current findings. Looking at child behavior in the family, peer group, and community, as well as investigating the power of home-based interventions, with and without comparable classroom-based support, represent an important research agenda for the next stage of preventive intervention research.

REFERENCES

Asher, S.R., & Dodge, K.A. (1986). Identifying children who are rejected by their peers. *Developmental Psychology, 22,* 444–449.

Barrish, H.H., Saunders, M., & Wolfe, M.D. (1969). Good Behavior Game. Effects of individual contingencies for group consequences and disruptive behavior in a classroom. *Journal of Applied Behavioral Analysis, 2,* 119–124.

Bierman, K.L. (1986). The relationship between social aggression and peer rejection in middle childhood. In R. Prinz (Ed.), *Advances in behavioral assessment of children and families, Vol. 2* (pp. 151–178). Greenwich, CT: JAI.

Bierman, K.L., Smoot, D.L., & Aumiller, K. (1987, March). *Distinguishing characteristics of aggressive-rejected, aggressive (non-rejected) and rejected (non-aggressive) boys.* Paper presented at the meeting of the Society for Research in Child Development, Baltimore, MD.

Block, J., Block, J.H., & Keyes, S. (1988). Longitudinally foretelling drug usage in adolescence: Early childhood personality and environmental precursors. *Child Development, 59,* 336–355.

Block, J., & Burns, R. (1976). Mastery learning. In L. Shulman (Ed.), *Review of research in education, Vol. 4* (pp. 3–49). Itasca, IL: F.E. Peacock.

Coie, J.D., & Dodge, K.A. (1988). Multiple sources of data on social behavior and social status in the school: A cross-age comparison. *Child Development, 59,* 815–829.

Conger, J.J., & Miller, W.C. (1966). *Personality, social class and delinquency.* New York: Wiley.

Corno, L., & Rohrkemper, M. (1985). The intrinsic motivation to learn in classrooms. In C. Ames & R. Ames (Eds.), *Research on motivation in education,* Vol. 2. Orlando, FL: Academic.

Dodge, K.A. (1983). Behavioral antecedents of peer social status. *Child Development, 54,* 1386–1399.

Dolan, L. (1986). Mastery learning as a preventive strategy. *Outcomes, 5,* 20–27.

Dolan, L., Ford, C., Newton, V., & Kellam, S. (1989). *The Mastery Learning manual.* Baltimore, MD: The Johns Hopkins Prevention Research Center.

Dolan, L., Turkkan, J., Werthamer-Larsson, L., & Kellam, S. (1989). *The Good Behavior Game training manual.* Baltimore, MD: The Johns Hopkins Prevention Research Center.

Doyle, W. (1977). Paradigms for research on teacher effectiveness. In L. Shulman (Ed.), *Review of research in education, Vol. 5* (pp. 163–198). Itasca, IL: F.E. Peacock.

Duke, D.L., & Meckel, A.M. (1984). *A teacher's guide to classroom management.* New York: Random House.

Elliott, S.N. (1988). Acceptability of behavioral treatments used in the classroom. *Professional Psychology: Research and Practice, 19,* 68–80.

Ensminger, M., Kellam, S.G., & Rubin, B.R. (1983). School and family origins of delinquency: Comparisons by sex. In K.T. Van Dusen & S.A. Mednick (Eds.), *Prospective studies of crime and delinquency* (pp. 73–97). Boston: Kluwer-Nijhoff.

Ensminger, M.E., & Slusarcick, A.L. (1992). Paths to high school graduation or dropout: A longitudinal study of a first grade cohort. *Sociology of Education, 65,* 95–113.

Farrington, D.P., Gallagher, B., Morley, L., St. Ledger, R.J., & West, D.J. (1988). Are there successful men from criminogenic backgrounds? *Psychiatry, 51,* 116–130.

Farrington, D.P., & Gunn, J. (Eds.). (1985). *Aggression and dangerousness.* New York: Wiley.

Gordon, R.S. (1983). An operational classification of disease prevention. *Public Health Reports, 98,* 107–109.

Guskey, T. (1985). *Implementing mastery learning.* Belmont, CA: Wadsworth.

Hans, S.L., Marcus, J., Henson, L., Auerbach, J.G., & Mirsky, A.F. (1992). *Interpersonal behavior of children at risk for schizophrenia. Psychiatry: Interpersonal and Biological Processes, 55,* 314–335.

Harris, V.W., & Sherman, J.A. (1973). Use and analysis of the Good Behavior Game to reduce disruptive classroom behavior. *Journal of Applied Behavior Analysis, 6,* 405–417.

Hawkins, J.D., VonCleve, E., & Calatano, F., Jr. (1991). Reducing early childhood aggression: Results of a primary prevention program. *Journal of the American Academy of Child and Adolescent Psychiatry, 30,* 208–217.

Kaplan, H.B. (1980). *Deviant behavior in defense of self.* New York: Academic.

Kellam, S.G. (1990). Developmental epidemiologic framework for family research on depression and aggression. In G.R. Patterson (Ed.), *Depression and aggression in family interaction* (pp. 11–48). Englewood Cliffs, NJ: Erlbaum.

Kellam, S.G., Branch, J.D., Agrawal, K.C., & Ensminger, M.E. (1975). *Mental health and going to school: The Woodlawn program of assessment, early intervention, and evaluation.* Chicago: University of Chicago Press.

Kellam, S.G., Brown, C.H., & Fleming, J.P. (1982). Social adaptation to first grade and teenage drug, alcohol and cigarette use. *The Journal of School Health, 52,* 301–306.

Kellam, S.G., Brown, C.H., Rubin, B.R., & Ensminger, M.D. (1983). Paths leading to teenage psychiatric symptoms and substance use: Developmental epidemiological studies in Woodlawn. In S.B. Guze, F.J. Earls, & J.E. Barrett (Eds.), *Childhood psychopathology and development* (pp. 17–55). New York: Raven Press.

Kellam, S.G., & Ensminger, M.E. (1980). Theory and method in child psychiatric epidemiology. In F. Earls (Ed.), *Studies of children* (pp. 145–180). New York: Prodist.

Kellam, S.G., & Rebok, G.W. (1992). Building developmental and etiological theory through epidemiologically based preventive intervention trials. In J. McCord & R.E. Tremblay (Eds.), *Preventing antisocial behavior: Interventions from birth through adolescence* (pp. 162–195). New York: Guilford Press.

Kellam, S.G., & Werthamer-Larsson, L. (1986). Developmental epidemiology a basis for prevention. In M. Kessler & S.E. Goldston (Eds.), *A decade of progress in primary prevention* (pp. 154–180). Hanover, NH: University Press of New England.

Kellam, S.G., Werthamer-Larsson, L., Dolan, L., Brown, C.H., Mayer, L., Rebok, G., Anthony, J., Laudolff, J., Edelsohn, G., & Wheeler, L. (1991). Developmental epidemiologically-based preventive trials: Baseline modeling of early target behaviors and depressive symptoms. *American Journal of Community Psychology, 19*, 563–584.

Kohlberg, L., Ricks, D., & Snarey, J. (1984). Childhood development as a predictor of adaptation in adulthood. *Genetic Psychology Monographs, 110*, 91–172.

Kulik, C.C., Kulik, J.A., & Bangert-Drowns, R.L. (1990). Effectiveness of Mastery Learning programs: A meta-analysis. *Review of Educational Research, 60*, 265–299.

Laird, N. (1983). Further comparative analyses of pretest–posttest research designs. *The American Statistician, 37*, 329–330.

Lefkowitz, M.M., Eron, L.D., Walden, L.O., & Huesman, L.R. (1977). *Growing up to be violent*. New York: Pergamon.

McCord, J. (1988). Parental behavior in the cycle of aggression. *Psychiatry, 51*, 14–23.

Medland, M., & Vitales, M. (1984). *Management of classrooms*. New York: Holt, Rinehart & Winston.

Medland, M.B., & Stacknik, T.J. (1972). Good Behavior Game: A replication and systematic analysis. *Journal of Applied Behavioral Analysis, 5*, 45–51.

Mitchell, S., & Rosa, P. (1981). Boyhood behavior problems as precursors of criminality: A fifteen-year follow-up study. *Journal of Child Psychology, 22*, 19–33.

O'Leary, S.G., & O'Leary, K.D. (1977). Ethical issues of behavior modification research in schools. *Psychology in Schools, 14*, 299–307.

Pekarik, E., Prinz, R., Leibert, C., Weintraub, S., & Neale, J. (1976). The Pupil Evaluation Inventory: A sociometric technique for assessing children's social behavior. *Journal of Abnormal Child Psychology, 4*, 83–97.

Robins, L.N. (1978). Sturdy childhood predictors of adult antisocial behavior: Replications from longitudinal studies. *Psychological Medicine, 8*, 611–622.

Robinson, V., & Swanton, C. (1980). The generalization of behavioral teacher training. *Review of Educational Research, 50*, 486–498.

Schwartzman, A.E., Ledingham, J.E., & Serbin, L.A. (1985). Identification of children at-risk for adult schizophrenia: A longitudinal study. *International Review of Applied Psychology, 34*, 363–380.

Shaffer, D., Stokman, C., O'Connor, P.A., Shafer, S., Barmack, J.E., Hess, S., & Spaulten, D. (1979, November). *Early soft neurological signs and later psychopathological development*. Paper presented at the meeting of the Society of Life History Research in Psychopathology and Society for the Study of Social Biology, New York.

Slavin, R.E. (1990). Mastery Learning re-reconsidered. *Review of Educational Research, 60*, 300–302.

Spivak, G., Marcus, J., & Swift, M. (1986). Early classroom behaviors and later misconduct. *Developmental Psychology, 22*, 124–131.

Tremblay, R.E., Masse, B., Perron, D., LeBlanc, M., Schwartzman, A.E., & Ledingham, J.E. (1992). Early disruptive behavior, poor school achievement, delinquent behavior, and delinquent personality: Longitudinal analyses. *Journal of Consulting and Clinical Psychology, 60*, 64–72.

Werthamer-Larsson, L., Kellam, S.G., & Wheeler, L. (1991). Effect of classroom environment on shy behavior, aggressive behavior, and concentration problems. *American Journal of Community Psychology, 19*, 585–602.

Younger, A.J., Schwartzman, A. E., & Ledingham, J.E. (1985). Grade-related changes in children's perceptions of aggression and withdrawal in their peers. *Developmental Psychology, 21*, 70–75.

Younger, A.J., Schwartzman, A.E., & Ledingham, J.E. (1986). Age-related differences in children's perceptions of social deviance: Changes in behavior or in perspective? *Developmental Psychology, 22*, 531–542.