Published in final edited form as:
Drug Alcohol Depend. 2008 June 1; 95(Suppl 1): S29–S44. doi:10.1016/j.drugalcdep.2007.10.009.

Impact of the Good Behavior Game, a Universal Classroom–Based Behavior Intervention, on Young Adult Service Use for Problems with Emotions, Behavior, or Drugs or Alcohol

Jeanne Poduska,
American Institutes for Research, 921 E. Fort Avenue, Suite 225, Baltimore, MD 21230

Sheppard Kellam,
American Institutes for Research, 921 E. Fort Avenue, Suite 225, Baltimore, MD 21230

Wei Wang,
Department of Epidemiology and Biostatistics, College of Public Health, University of South Florida, 13201 Bruce B Downs Blvd, Tampa, FL 33612

C. Hendricks Brown,
Department of Epidemiology and Biostatistics, College of Public Health, University of South Florida, 13201 Bruce B Downs Blvd, Tampa, FL 33612

Nicholas Ialongo, and
Johns Hopkins University, Bloomberg School of Public Health, 624 N. Broadway, 8th Fl., Baltimore, MD

Peter Toyinbo
Department of Epidemiology and Biostatistics, College of Public Health, University of South Florida, 13201 Bruce B Downs Blvd, Tampa, FL 33612

Abstract

Background—The Good Behavior Game (GBG) is a classroom behavior management strategy focused on socializing children to the role of student and aimed at reducing early aggressive, disruptive behavior, a confirmed antecedent to service use. The GBG was tested in a randomized field trial in 19 elementary schools in two cohorts of children as they attended first and second grades. This article reports on the impact of the GBG on service use through young adulthood.

Methods—Three or four schools in each of five urban areas were matched and randomly assigned to one of three conditions: 1) GBG, 2) an intervention aimed at academic achievement, or 3) the standard program of the school system. Children were assigned to classrooms to ensure balance, and teachers and classrooms were randomly assigned to intervention conditions.

Results—This study provides evidence of a positive impact of a universal preventive intervention on later service use by males, although not by females, for problems with emotions, behavior, or drugs or alcohol. For both cohorts, males in GBG classrooms who had been rated as highly aggressive, disruptive by their teachers in the fall of first grade had a lower rate of school-based service use than their counterparts in control classrooms.

Phone: 410-347-8553, Email: jpoduska@air.org.

Replication—The design employed two cohorts of students. Although both first- and second-grade teachers received less training and support with the second cohorts of students than with the first cohort, the impact of GBG was similar across both cohorts.

Keywords

Good Behavior Game; developmental epidemiology; universal prevention; classroom behavior management; aggressive behavior; mental health service use

1. Introduction

Over the past decade, the public health concept of an integrated system of interventions and services for mental health has entered the parlance of practitioners and policymakers in many fields, including education. This concept is based on an epidemiological understanding of population risk in which preventive interventions followed by treatment and maintenance are delivered to distinct populations and sub-populations on the basis of levels of risk (Gordon, 1983; Mrazek and Haggerty, 1994). Mrazek and Haggerty (1994) defined five levels of intervention in public mental health: three levels of preventive interventions, treatment intervention, and maintenance intervention. *Universal preventive interventions* are delivered to an entire population without focusing on any specific group in the population and often aim at strengthening some aspect of the environment. An example in education would be training teachers in an effective classroom behavior management strategy thereby strengthening the classroom environment and socializing students to behave in a positive manner. In a truly integrated service system, *selective interventions* backed up by *indicated interventions* are delivered to the segment of the population that is at increased risk relative to the population average and does not respond to interventions earlier in the continuum. It follows that in a population served by an integrated system of interventions, we would expect that influencing a risk factor or a set of risk factors through a preventive intervention would lower the proportion of the population needing to be served through more intensive levels of interventions and services for problems attributable to that risk factor. This paper contributes to our understanding of the interplay between levels of the system by studying the impact of the Good Behavior Game (GBG), a universal preventive intervention delivered in first and second grades and aimed at reducing aggressive, disruptive behavior on later service use for problems with emotions, behavior, or drug or alcohol from elementary school to young adulthood.

1.1 Developmental Epidemiology and Life Course/Social Field Theory

Developmental epidemiology is the prevention science strategy that has guided three generations of preventive intervention trials carried out in partnership with the Baltimore City Public School System (Kellam and Rebok, 1992; Kellam et al., 1999). It draws on three perspectives: community epidemiology, life course development, and theory testing through randomized field trials. Community epidemiology is concerned with the non-random distribution of health problems or related factors in a fairly small population in the context of its environment, such as a neighborhood or a school catchment area. Service use can be conceptualized as a response to these problems. With selection bias controlled, the second perspective of studying variation in developmental antecedents and pathways within the defined population, coupled with the third perspective of directing preventive interventions at hypothesized causal risk factors, allows the study not only of main effects but also of variation in the impact of the preventive intervention on outcomes, including service use. Developmental epidemiology is critical to services research because understanding what interventions work for whom and under which conditions is essential information for providing services in an integrated fashion.

Life course/social field theory (Kellam et al., 1975) has guided the hypotheses of three generations of Baltimore-based field trials and provides a useful framework as we consider use of services throughout childhood and adolescence. A basic tenet of the theory is that individuals move through different social fields across the lifespan. For example, the main social fields for first graders are the classroom and the family. Within each social field, *natural raters*, such as teachers in the classroom, parents in the family, peers in the peer group, supervisors in the workplace, and spouses in the marital social field, define social task demands specific to that social field. The interactive process of demand—as defined by the natural rater—and response—by the individual—is termed *social adaptation*. The judgment of an individual's performance by a natural rater is *social adaptational status* (SAS; Kellam et al., 1975). In contrast to SAS, psychological well-being (PWB) reflects an individual's internal state, as indicated through symptoms and diagnoses. There are two key hypotheses in life course/social field theory. First, in keeping with Cicchetti and Schneider-Rosen (1984), we hypothesize that success in mastering social task demands specific to one stage of development and social field will lead to an increase in later successes both across development and across social fields. Additionally, we hypothesize a potential reciprocal relationship between SAS and PWB.

1.2 Early Aggressive, Disruptive Behavior as a Predictor for Service Use

Aggressive, disruptive behavior, as early as first grade, is a well-established predictor of deleterious outcomes such as later aggressive, disruptive behavior, conduct disorder, antisocial personality disorder, and illicit drug use (Dishion et al., 1996; Ensminger and Slusarcick, 1992; Ensminger et al., 1983; Farrington and Gunn, 1985; Farrington et al., 1988; Ialongo et al., 2001; Kellam and Ensminger, 1980; Kellam and Rebok, 1992; Kellam et al., 1975; Kellam et al., 1983; Kellam et al., 1991; Kellam et al., 1994b; McCord, 1998; Patterson et al., 1992; Robins, 1978). Children often begin to exhibit aggressive, disruptive behavior at an early age in their interactions with family, including parents and siblings, and with peers. An early risk factor for the development of conduct problems, including aggressive, disruptive behavior, throughout childhood is ineffective and coercive parenting practices in response to challenging child behavior, which escalates and reifies the oppositional behavior (Patterson et al., 1992; Reid, 1993; Reid and Eddy, 1997). When these children enter school, they often engage in the same aversive behaviors with the teacher as they did with the parents. Teachers who lack training in classroom behavior management may find themselves engaging in ineffective and coercive interactions with these students, further intensifying the negative behaviors rather than extinguishing them.

The interactive process captured in the ratings of SAS is particularly relevant to child and adolescent services research because by and large, children and adolescents receive services through the intercession of adults who perceive them to need those services. In fact, during the school years, the determination of need for services, followed by referral and use, quite often occurs as a result of poor social adaptation of children to the teacher's behavioral task demands (Gerber and Semmel, 1984; Horcutt, 1996; Lloyd et al., 1991; Mattison et al., 1986; Poduska, 2000; Shinn et al., 1987). Hocutt (1996) refers to the “teachability” of students when discussing factors that teachers consider when thinking about children's need for services. She defines teachability as the “extent to which the student is alert, sustains attention in the classroom, and begins and completes [tasks] on time,” a concise description of classroom SAS.

Further, although both SAS and PWB have been found to be associated with the use of mental health services (Burns et al., 1995; Costello and Janisezwski,1990; Feehan et al.,1990; Garralda and Bailey, 1988; John et al., 1995; Laitinen-Krispijn et al., 1999; Rickwood and Braithwaite, 1994; Staghezza-Jaramillo et al., 1995; Verhulst and der Ende, 1997), there is evidence that children with aggressive, disruptive behavior and diagnoses are more likely to receive services

than children with depressive symptoms or a diagnosis of depression alone (Cohen et al., 1991; Wu et al., 1999). In a recent study, elementary school teachers cited disruptive classroom behavior as the largest mental health problem in their schools and a lack of information/training as the greatest barrier to addressing these problems, with a lack of time as a close second (Walter et al., 2006).

Results published to date on the GBG as it was tested alone in the first generation of trials in Baltimore have shown a positive impact of the intervention on aggressive, disruptive behavior at the end of first grade (Brown, 1993; Dolan et al., 1993), through middle school (Kellam et al., 1994a; Kellam et al., 1998), and to the transition into young adulthood (Kellam et al., in press, this issue; Petras et al., in press, this issue), particularly for the most aggressive, disruptive males. In addition, results from the second generation of trials conducted in Baltimore in which the GBG was combined with an instructional intervention showed at sixth-grade follow-up that children in this combined classroom-based intervention were rated by teachers as having fewer conduct problems and were less likely to have a diagnosis of conduct disorder, to have been suspended from school, or to have been in need of, or to have received, mental health services (Ialongo et al., 2001).

The overarching hypothesis for this paper is that to the extent that service use for problems with behavior, feelings, or drugs and alcohol is associated with aggressive, disruptive behavior, the GBG, a universal intervention aimed at aggressive, disruptive behavior, will reduce the use of services over the course of childhood to young adulthood. In keeping with prior results, we further hypothesize that the impact will be greatest for those youth who exhibited the highest levels of aggressive, disruptive behavior on entrance to first grade. We will examine GBG impact on service use across several specific service sectors in addition to school because children and adolescents receive services for problems with behavior, feelings, and drugs and alcohol through a variety of additional service sectors, including the mental health, medical health, welfare, and justice systems (Burns et al., 1995; Leaf et al., 1986).

2. Methods

2.1 The GBG Intervention

The GBG is directed at the entire class and targets aggressive, disruptive behavior (Barrish et al., 1969). The GBG has been cited as an effective program by the American Federation of Teachers (AFT) in its *Building on the Best, Learning from What Works* series (AFT, 2000) and by the Surgeon General in his report on youth violence (Department of Health and Human Services, 2001). It has been replicated by at least a dozen independent research groups, in the United States and abroad, since its development in 1969 by Barrish, Saunders, and Wolfe. The most extensive and rigorous randomized field trials have been carried out in Baltimore; see Tingstrom et al. (2006) for a review.

The GBG is a group-contingent classroom behavior management strategy that promotes classmates' concern for each child's behavior by rewarding teams that exhibit low levels of inappropriate behavior. This strategy creates an environment in which the behavior of each child on the team becomes a matter of concern to all children on that team because the team reward depends on all team members' behavior. The goal is to encourage students to manage their own behavior through group reinforcement and mutual self-interest.

To begin the game, the teacher posts clear rules for student behavior and teaches the students what constitutes misbehavior. All students in each classroom are assigned to one of three teams that are heterogeneous in behavior and learning, with equal membership by gender. At random times, the teacher announces “We are playing the Good Behavior Game.” During the game, the teacher places a mark on the chalkboard next to the name of the team whenever a team

member breaks a rule by displaying a defined misbehavior, such as “talking out of turn” or “getting out-of-seat without permission.” A team wins if the number of checkmarks does not exceed four at the end of the game period; all teams can win the game. Students on the winning team(s) receive a special reward or engage in a rewarding activity. At the beginning of the year, the GBG is played three times a week for ten minutes each time. Over the course of the year, the amount of time GBG is played increases. By the end of the year, the game is played at different times throughout the day, during different activities, and in a variety of venues. In addition, the rewards change over the course of the year from being tangible (such as stickers or erasers) to being more intrinsically related to the classroom setting, such as extra quiet time to read during the school day.

In the 1985–86 school year when the first cohort of students was in first grade, first-grade teachers received 40 hours of training and support during the year. In the 1986–87 school year when the second cohort of first graders was in first grade, the first-grade teachers received minimal support from the research team or the school system; the focus of training and support was on the second-grade teachers, who received 40 hours of training and support throughout the year as the first cohort of students moved into second grade.

2.2 Design

In this paper, we use data from an epidemiologically defined study of students in the Baltimore City Public School System (BCPSS) who as first and second graders participated in a randomized field trial in which two universal preventive interventions were tested with two subsequent cohorts of first-grade students entering first grade in 1985 and 1986. Two classroom-based interventions were tested separately; each targeted different risk factors. The proximal targets of the Good Behavior Game intervention (GBG; Barrish et al., 1969), which are the focus of this paper, were aggressive, disruptive behavior and shy behavior in first and second grades, confirmed antecedents of later conduct problems, antisocial personality, and substance abuse. The proximal target of the second intervention, Mastery Learning (ML; Block and Burns, 1976; Guskey, 1997), was poor school achievement, an antecedent of later depressive symptoms and disorders.

In the first stage of the design, five urban areas varied in terms of ethnicity, type of housing, family structure, income, unemployment, and violent crime were identified with the involvement of the Baltimore City Public School System (BCPSS) and the Baltimore City Planning Commission (see Vaden-Kiernan et al., 1995, for a detailed description of the areas). Within each urban area, three or four schools were matched with regard to census tract, school level, and first- and second-grade data. These schools were then randomly assigned to serve as 1) schools in which GBG would be tested independently of ML (6 schools); 2) schools in which ML would be tested independently of GBG (7 schools); or 3) schools that would serve as external control schools in which neither of these interventions would take place (6 schools).

In the second stage, children were assigned to classrooms within each school by the school administrator, who used an alphabetized list to sequentially assign children to classrooms. Classes were checked to see whether they were balanced on children's academic and behavior performance as indicated by data from the children's kindergarten experience. In a few instances, children were reassigned to obtain a balance. The third stage of the design involved random assignment of first-grade classrooms and teachers to an intervention condition within each school after baseline assessments, but before the intervention began. Within each intervention school this created intervention classrooms and control classrooms which we call internal GBG or ML controls. The interventions were delivered over first and second grades for the two subsequent cohorts of students. The children were kept together as a class as they moved from first grade into second grade, but they did change teachers; in other words, the first-grade teacher was not the second-grade teacher.

In this paper, we conduct initial analyses using all three control conditions: internal GBG control classrooms, classrooms in the external matched control schools, and the internal ML control classrooms. We then conduct analyses using only the internal GBG control classrooms. We consider the internal GBG control classrooms the most informative test of GBG impact against the GBG classrooms because this comparison controls for school and community variation.

2.3 The Population from First Grade through Young Adulthood

The study population for each cohort is defined as the children who were present at the time of baseline assessments eight to ten weeks into the school year when classroom composition stabilized. Baseline data were collected prior to the start of the intervention. The total population for Cohort 1 was 1,196; for Cohort 2, it was 1,115. There were no significant differences between the intervention groups at baseline for teacher ratings of aggressive, disruptive behavior, reading achievement, or federal free or reduced-price lunch program status, taking into account school as a random factor. There was also a high degree of comparability at baseline between GBG and internal GBG control means for baseline teacher ratings of aggressive, disruptive behavior (Cohort 1 [C1]: $p = 0.30$; Cohort 2 [C2]: $p = 0.89$) and concentration problems (C1: $p = 0.73$; C2: $p = 0.70$); child's self-ratings of anxiety (C1: $p = 0.17$; C2: $p = 0.87$); scores on standardized tests of reading achievement (C1: $p = 0.84$; C2: $p = 0.68$); and participation in the federal lunch program (C1: $p = 0.30$; C2: $p = 0.56$). There was a baseline difference between GBG classrooms and GBG internal control classrooms on depressive symptoms by child self-report for Cohort 1 (C1: $p = 0.01$; C2: $p = 0.49$) and shy behavior by teacher report for Cohort 2 (C1: $p = 0.43$; C2: $p = 0.02$). However, because the correlation between aggressive, disruptive behavior (the risk factor of interest in this paper) and depressive symptoms is 0.02 for Cohort 1 and 0.01 for Cohort 2 and between aggressive, disruptive behavior and shy behavior is 0.05 for Cohort 1 and 0.13 for Cohort 2, we would not expect these differences to influence our analyses and findings.

Six hundred eighty nine (75%) of the 922 students in Cohort 1 and 656 (76%) of the 867 students in Cohort 2 who were in intervention conditions relevant to this paper (GBG, GBG internal control, ML internal control, and external controls) completed an interview at young adulthood, ages 19–21. There were no significant differences between the individuals who completed the interview and those who did not for baseline ratings of teacher-rated aggressive, disruptive behavior (C1: $p = 0.34$; C2: $p = 0.32$) or shy behavior (C1: $p = 0.37$; C2: $p = 0.22$); child's self-ratings of anxiety (C1: $p = 0.26$; C2: $p = 0.60$) and depression (C1: $p = 0.20$; C2: $p = 0.13$); scores on standardized tests of reading achievement (C1: $p = 0.28$; C2: $p = 0.14$); or federal free or reduced-price lunch program status (C1: $p = 0.16$; C2: $p = 0.74$), without taking into account school/classroom as a random factor. There was a significant difference for teacher ratings of concentration problems in Cohort 2 ($p = 0.02$) but not in Cohort 1 ($p = 0.48$). Gender distribution was significantly different in both Cohort 1 ($p < 0.001$) and Cohort 2 ($p < 001$). A greater proportion of females in both cohorts completed a young adult interview (C1: females 382/460 [83%], males 307/462 [66%]; C2: females 346/427 [81%], males 310/440 [70%]).

There is little missing data on the service use question at young adulthood across the different service sectors (missing ≤ 3). For the logistic regression modeling where the missing data in the measure of baseline risk (aggressive, disruptive behavior) also needs to be taken into account, we have 625 to 627 cases available for analyses for Cohort 1 (males: 278 or 279; females: 347 or 348) and 603 to 605 cases available for Cohort 2 (males: 287 to 289; females: 316) depending on the service sector.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

2.4 Data Collection and Measures

2.4.1. Data Collection—Students enrolled in BCPSS and their teachers participated in baseline assessments in the fall of first grade. The young adult follow-up data was collected through a two-hour telephone interview by interviewers who were blind to the intervention condition of the respondents. An attempt was made to contact all students in the population as they entered young adulthood at 19 to 21 years of age. The information in the interview was organized by the main social fields of family of orientation, school, peers, work, intimate relations, and family of procreation. Information on developmental history as well as current status was collected. The WHO version of the Composite International Diagnostic Interview (CIDI) was used to obtain DSM diagnoses (Kessler et al., 1994). All measures and procedures during the school years were approved by the Institutional Review Board (IRB) of Johns Hopkins University and BCPSS; approval for the young adult data collection was granted from the IRBs of Johns Hopkins University and the American Institutes for Research. In addition, the young adult data collection was conducted under the aegis of an external Community and Institution Board that provided guidance, community support, and oversight.

2.4.2 Social Adaptational Status (SAS) in First Grade: Teacher Observation of Classroom Adaptation-Revised (TOCA-R)—The TOCA-R (Werthamer-Larsson et al., 1991) measures each child's adequacy of performance on the core social tasks of the classroom. Baseline teacher interviews were conducted in the fall of first grade eight to ten weeks into the school year once the teacher knew the children. The TOCA-R is a structured interview administered by a trained interviewer who follows a precise script. Teachers first respond to items pertaining to the child's adaptation in the classroom; they are then asked to make global ratings of each child's behavior and progress as a student. The three central domains on which teachers rate each child's performance over the previous three weeks are aggressive, disruptive behavior, shy behavior, and concentration problems. A 6-point scale is used in which the anchor of 1 = never and the anchor of 6 = almost always. Werthamer-Larsson et al. (1991) reported test-retest correlations over a four-month interval with different interviewers of 0.60 or higher for each of the scales and coefficient alphas of 0.96 for aggressive, disruptive behavior, 0.92 for shy behavior, and 0.85 for concentration problems. A correlation of 0.67 was found between ratings on the aggressive, disruptive behavior scale and peer nominations of “gets in trouble.” The aggressive, disruptive behavior scale (the central SAS construct of interest to this paper) is composed of the following items: 1) breaks rules, 2) harms others and property, 3) breaks things, 4) takes others' property, 5) fights, 6) lies, 7) teases classmates, 8) yells at others, 9) is stubborn, and 10) has trouble accepting authority. The coefficient alphas ranged from 0.92 to 0.94 over grades 1 through 7.

2.4.3 Service Use for Problems with Behaviors, Emotions, or Drugs or Alcohol by Young Adulthood—The Services Assessment of Children and Adolescents (SACA; Horwitz et al., 2001) was employed as part of the telephone interview at the young adult follow-up. The SACA, a structured interview, was designed to obtain information on child and adolescent mental health service utilization (Hoagwood et al., 2000; Horwitz et al., 2001; Stiffman et al., 2000). The SACA was used to obtain information about past and present use of mental health and educational services, including the setting (e.g., outpatient, inpatient, school-based, primary care, juvenile justice system). Young adults were asked “Have you received services from…for problems with behaviors, feelings, drugs or alcohol?” about receiving services from twenty-six types of providers. Time of service use was established for each service an individual reported ever receiving through a set of questions to determine the age at first use, use within the past year, and current use. Stiffman et al. (2000) cite the inter-rater reliability between parent and child reports of service use using the SACA as fair to excellent (*kappa* = 0.43 to 0.86), with agreement increasing with the child's age. In turn, the test-retest reliability of parent-reported twelve-month service use is strong (*kappa* = 0.75 to

0.86; Horwitz et al., 2001). Hoagwood et al. (2000) report a *kappa* of .76 between parent reports of a global “any use” service variable and service records. The *kappa* for agreement between parent report and records for inpatient treatment was 1.0.

Of interest for this paper is the cumulative use of services by young adults for problems with behavior, feelings, drugs, or alcohol. Twenty-two of twenty-six types of service providers were coded into five categories of 1) medical or mental health providers, 2) education system, 3) drug treatment, 4) juvenile or adult justice system, and 5) social service system. A dichotomous variable was created for each category with “yes” indicating that services from this type of provider had been used at some point in the individual's life and “no” indicating that services had never been received from this type of provider. *Medical or mental health providers* included receiving services in a hospital; residential treatment center; another place like a summer treatment program or boarding school; a community mental health center or other outpatient center; by a professional psychologist, psychiatrist, social worker, marriage, or family counselor; partial hospitalization or day treatment; emergency room; or pediatrician and family doctors. *School-based services* included being placed in a special school or special classroom for problems with behavior, feelings, or drug or alcohol; receiving special help in the regular classroom; and receiving other counseling or therapy in school. *Drug treatment services* included receiving treatment from drug and alcohol treatment units, drug or alcohol treatment, and self-help groups such as Alcoholics Anonymous. The *juvenile or adult justice system* included detention centers, prisons, and jails and probation officers, juvenile corrections officers, and court counselors. *Social services* included group homes, foster homes, emergency shelters, and respite care providers. Four types of service providers were not used in these analyses: priest, minister, and rabbi; healer, shaman, and spiritualist; acupuncturist and chiropractor; and crisis hotline. The less than 2% of the young adults who had used services only from these types of providers are not represented in the analyses. In addition to the five categories above, a category for “ever received service from any provider type” was created.

2.5 Statistical Methods

To test the impact of GBG on service use by young adulthood, we used three methods of intent-to-treat analyses because there are no published and available statistical methods that can handle all the complexities we wanted to include simultaneously in our models. For example, we tried to fit a three-level mixed effects model by adding urban area as a fixed factor, with individual, classroom, and school as the three-level random effects; however, existing commercial software could not give us reliable results. In the first method, two-level models, with individual- and classroom-level effects, provide the most direct test of the theory of how GBG impacts youth with different risks of aggressive, disruptive behavior. We rely on this model to examine both linear and nonlinear variation in impact based on individual-level risk. We have included classroom-level random effects and, where necessary, classroom-level predictors in these analyses because this group-based randomized trial involved random assignment at the classroom level (Murray, 1998). These two-level linear and additive logistic regression models with random intercepts provide adjustments for potential dependence of observations at the classroom level and for baseline by treatment interactions.

We used large sample tests on the two-level modeling because few if any exact tests are available for dichotomous outcomes being fit with linear logistic and nonlinear random effects modeling. These tests are generally appropriate when the number of subjects is large, the number of classes is moderate, and the magnitude of the intraclass correlations is small ($\rho < 0.05$), as we have found empirically to be the case in these analyses. Models that involve logistic regression analyses are fit using Splus version 7 (Insightful, Seattle, WA). These models include main and interaction effects involving the baseline levels of aggressive, disruptive behavior, intervention condition, and gender if either linear interactions or nonlinear

interactions are indicated (Brown, 1993). Those with nonlinear interactions are fit in R version 2.1.1 (R Development Core Team, http://www.r-project.org) with generalized additive models (GAM) using the logistic link function (Dominici et al., 2002; Hastie and Tibshirani, 1990). Generally, GBG impact on service use was more pronounced for males as revealed by the analyses. When interactions involving gender were significant at the beginning stages of modeling, we conducted subsequent analyses separated by gender. To minimize redundancy of information, when GBG impact was significant for males, but not for females, we present only the final model for males with detail in the tables. When GBG impact was far from significant for both genders, we present the gender-combined result in the tables.

The second and third analytical methods—one based on Mantel-Haenszel statistics and the other on a paired t-test for log odds ratios—are used to condition on school to account for the blocking of this factor in the design. These latter two analyses for binary outcomes provide a legitimate overall test of GBG impact with small sample corrections based on the number of schools, but they fail to take into account variation in individual-level risk. Details of these three methods are described in depth in Brown et al. (in press, this issue) and Kellam et al. (in press, this issue).

3. Results

The results are organized by service sector, beginning with 1) use of services for problems with behavior, emotions, or drugs or alcohol from any provider, followed by use of services from the specific providers: 2) mental or medical health provider, 3) school-based services, 4) drug treatment provider, 5) juvenile or adult justice system, and 6) social services. For each service sector, we present the results for Cohort 1 followed by the results for Cohort 2. Because there are differences in the models by service provider, we present the model building and the final model for each service category, realizing that this presentation is quite repetitive. The descriptive epidemiology of cumulative service use can be found in the supplementary materials on the journal's website.[1]

3.1 GBG Impact on Any Service Use

3.1.1 Cohort 1—Overall unadjusted rates of any service use were lower for youth in the GBG condition (21%) compared with internal controls (31%, $p = 0.06$) and all controls (28%; $p = 0.09$). For males, the difference was significant, with 25% for GBG versus 42% for internal controls ($p = 0.03$) and 38% for all controls, ($p = 0.03$). In contrast, there were no significant differences in rates between GBG females and control females, with 19% for GBG versus 22% for internal controls ($p = 0.66$) and 19% for all controls ($p = 0.95$).

The final model for any service use is a generalized linear mixed effects model that includes baseline aggressive, disruptive behavior, gender, intervention conditions, and interaction between gender and baseline, as well as a three-way interaction involving gender, baseline, and intervention conditions. Because the interaction between gender and baseline aggressive, disruptive behavior was significant, we conducted separate analyses by gender. For females, the level of aggressive, disruptive behavior was positively related to service use for those exposed to GBG, but there was almost no relation for controls. For males, the final model contains baseline aggressive, disruptive behavior level ($p < 0.001$) and intervention conditions. GBG main effect against internal controls is not significant ($p = 0.15$) but is in the expected direction. The final model for males contains classroom variation as the second source of random effects; see Table 1. GBG did not show significant impact for females ($p = 0.84$).

[1]See online supplementary material. http://www.sciencedirect.com/science/journal/03768716

We also report the results of testing for intervention impact on any service use that explicitly take into account the blocking factor of school. Combining females and males, the Mantel-Haenszel continuity corrected test that conditioned on school did not find a significant association between GBG and a lower probability of any service use (χ^2 = 2.211 on 1 *df*, *p* = 0.14). The corresponding small sample test of the value of the log odds ratio accounting for school as a blocking factor was not significant either (*t* = −2.244 on 5 *df*, *p* = 0.07). For males, the Mantel-Haenszel test gave a stronger result but still did not reach the 0.05 significance level (χ^2 = 2.738 on 1 *df*, *p* = 0.10). The corresponding paired t-test of intervention effect using the log odds ratios controlling for school also gave a non-significant result (*t* = −1.912 on 5 *df*, *p* = 0.11). There was no significant difference revealed by these two tests for females (χ^2 = 0.09 on 1 *df*, *p* = 0.87 for the Mantel-Haenszel test and *t* = −0.843 on 5 *df*, *p* = 0.44 for the paired t-test).

3.1.2 Cohort 2—Overall unadjusted rates of any service use were lower for youth in the GBG condition (22%) compared with internal controls (26%, *p* = 0.45) and all controls (23%, *p* = 0.73). For males, the difference was significant, with 21% for GBG versus 38% for internal controls (*p* = 0.03) but was not significant compared with all controls (28%, *p* = 0.25). In contrast, there were no significant differences in rates between GBG females and control females, with 22% for GBG versus 16% for internal controls (*p* = 0.36) and 19% for all controls (*p* = 0.52).

The final Cohort 2 model for any service use is a generalized linear mixed effects model that includes baseline aggressive, disruptive behavior, gender, intervention conditions, and interactions between gender and intervention conditions. Because the interaction between gender and intervention conditions was significant, we conducted separate analyses by gender. For females, neither the baseline (*p* = 0.26) nor GBG impact (*p* = 0.45) was significant. For males, the final model contains the baseline aggressive, disruptive behavior level (*p* = 0.004) and intervention conditions. GBG showed a significant impact (*p* = 0.03). The final model for males contains classroom variation as the second source of random effects; see Table 2.

We also report the results of testing for intervention impact on any service use that explicitly take into account the blocking factor of school. Combining females and males, the Mantel-Haenszel continuity corrected test that conditioned on school did not find a significant association between GBG and a lower probability of any service use (χ^2 = .435 on 1 *df*, *p* = 0.51). The corresponding small sample test of the value of the log odds ratio accounting for school as a blocking factor was not significant either (*t* = −.764 on 5 *df*, *p* = 0.48). For males, both the Mantel-Haenszel test and the paired t-test had significant results. The chi-square statistic is 4.424 on 1 *df* with *p* value equal to 0.04. The corresponding paired t-test had a result of −2.661 on 5 *df* with *p* value equal to 0.04. These two tests revealed no significant difference for females (χ^2 = 0.752 on 1 *df*, *p* = 0.39 for the Mantel-Haenszel test and *t* = 0.490 on 5 *df*, *p* = 0.64 for the paired t-test).

3.2 GBG Impact on Service Use from Mental or Medical Health Provider

3.2.1 Cohort 1—Overall unadjusted rates of service use from mental or medical health providers were 13% in the GBG condition compared with 15% in the internal control condition (*p* = 0.64) and 16% for all controls (*p* = 0.35). For males, rates were 13% for GBG versus 14% for internal controls (*p* = 0.86) and 18% for all controls (*p* = 0.32). For females, the rates were 13% for GBG versus 16% for internal controls (*p* = 0.63) and 15% for all controls (*p* = 0.75).

The final model for receiving services in the mental or medical health sector was a generalized linear mixed effects model that included baseline aggressive, disruptive behavior, gender, intervention conditions, and gender by baseline interaction term. The *p*-value of GBG impact was 0.90. Baseline aggressive, disruptive behavior was a strongly significant predictor (*p* =

0.001) of service use from the mental or medical health sector. The final combined-gender model contains classroom as random effects, ($p < 0.001$); see Table 3. Because the gender by baseline interaction was significant, we conducted separate analyses for males and females but saw no GBG impact ($p = 0.98$ for males and $p = 0.92$ for females, classroom effect included).

We also report the results of testing for intervention impact on service use from mental or medical health providers that explicitly take into account the blocking factor of school. Combining females and males, the Mantel-Haenszel continuity corrected test that conditioned on school did not find a significant association between GBG and a lower probability of service use in this sector ($\chi^2 = 0.03$ on 1 *df*, $p = 0.87$). The corresponding small sample test of the value of the log odds ratio accounting for school as a blocking factor, was also not significant, ($t = -0.26$ on 5 *df*, $p = 0.80$). For males, the Mantel-Haenszel test result was $\chi^2 = 0.03$ on 1 *df*, $p = 0.87$ and the corresponding paired t-test result of intervention effect using the log odds ratios controlling for school was $t = -1.16$ on 5 *df*, $p = 0.30$. For females, the Mantel-Haenszel test result was $\chi^2 = 0.006$ on 1 *df*, $p = 0.93$ and the t-test result was $t = -0.58$ on 5 *df*, $p = 0.59$.

3.2.2 Cohort 2—Overall unadjusted rates of service use from mental or medical health providers were 11% in the GBG condition compared with 18% in the internal control condition ($p = 0.10$) and 15% for all controls ($p = 0.21$). For males, the rate of service use in the GBG condition (9.4%) was significantly lowered than for the internal controls (26%, $p = 0.01$) although not for GBG compared with all controls (16% for all controls, $p = 0.16$). For females, the rates were 13% for GBG versus 12% for internal controls ($p = 0.85$) and 15% for all controls, ($p = 0.71$).

The final model for receiving services in the mental or medical health sector was a generalized linear mixed effects model that included baseline aggressive, disruptive behavior, gender, intervention conditions, and interactions between gender and intervention conditions. Because the interaction between gender and intervention conditions was significant, we conducted separate analyses by gender. For females, neither the baseline ($p = 0.31$) nor GBG impact ($p = 0.84$) was significant. For males, the final model contains baseline aggressive, disruptive behavior level ($p = 0.23$) and intervention conditions. The GBG showed a significant impact ($p = 0.02$). The final model for males contains classroom variation as the second source of random effects, which is significant ($p < 0.001$); see Table 4.

We also report the results of testing for intervention impact on service use from mental or medical health provider that explicitly take into account the blocking factor of school. Combining females and males, the Mantel-Haenszel continuity corrected test that conditioned on school did not find a significant association between GBG and a lower probability of service use in this sector ($\chi^2 = 2.13$ on 1 *df*, $p = 0.14$). The corresponding small sample test of the value of the log odds ratio accounting for school as a blocking factor was also not significant, ($t = -1.66$ on 5 *df*, $p = 0.16$). This effect was pronounced for males, for whom the Mantel-Haenszel test result was $\chi^2 = 6.11$ on 1 *df*, $p = 0.01$ and the corresponding paired t-test result of intervention effect using the log odds ratios controlling for each school was $t = -4.55$ on 5 *df*, $p = 0.01$. For females, the Mantel-Haenszel test result was $\chi^2 = 0.03$ on 1 *df*, $p = 0.86$ and the t-test result was $t = 0.14$ on 5 *df*, $p = 0.89$.

3.3 GBG Impact on Service Use from School-Based Services

3.3.1 Cohort 1—Overall unadjusted rates of school-based services were lower for youth in the GBG condition (6.6%) compared with internal controls (13%, $p = 0.06$) and all controls (9.6%; $p = 0.22$). For males, the rate of service use in the GBG condition (9.1%) was significantly lower than for the internal controls (21%, $p = 0.04$) although not for GBG compared with all controls (14% for all controls, $p = 0.23$). For females, the rates were 4.8% for GBG versus 5.9% for internal controls ($p = 0.63$) and 5.5% for all controls ($p = 0.78$.)

The final model for school-based services was a generalized additive mixed effects model that included main effects of baseline aggressive, disruptive behavior, gender and intervention conditions. It also included gender by baseline interaction and intervention conditions by baseline interactions. The gender and baseline aggressive, disruptive behavior interaction finding reflects a strong predictive relationship of baseline aggressive, disruptive behavior for males. The *p*-value for combined GBG impact was 0.11. Because the gender by baseline interaction was significant, we did a separate analysis on males, which revealed a significant GBG impact ($p = 0.001$). Males whose teachers rated them as exhibiting higher levels of aggressive, disruptive behavior in first grade—the high-risk males—benefited more from GBG than males rated to have lower levels of aggressive, disruptive behavior; see Table 5 for the final model for males. There was no significant GBG impact on females ($p = 0.96$).

School appears to have an independent effect on the probability of receiving school-based services. For the combined sample of males and females, the Mantel-Haenszel continuity corrected test that conditioned on school did not reveal a significant association between GBG and a lower probability of school-based services ($\chi^2 = 2.43$ on 1 *df*, $p = 0.11$); however, the corresponding small sample t-test result of the value of the log odds ratio accounting for school as a blocking factor was $t = -2.28$ on 5 *df*, $p = 0.05$. This effect was most pronounced for males, for whom the Mantel-Haenszel test result was $\chi^2 = 2.70$ on 1 *df*, $p = 0.10$ and the corresponding paired t-test result of intervention effect using the log odds ratios controlling for each school was $t = -4.82$ on 5 *df*, $p = 0.01$. For females, the Mantel-Haenszel test result was $\chi^2 = 0.12$ on 1 *df*, $p = 0.73$ and the t-test result was $t = -2.14$ on 5 *df*, $p = 0.09$.

3.3.2 Cohort 2—Overall unadjusted rates of school-based services were lower for youth in the GBG condition (8.2%) compared with internal controls (11%, $p = 0.41$) and all controls (8.8%; $p = 0.81$). For males, the rate of service use in the GBG condition (9.4%) was lower although not significantly less than for the internal controls (20%, $p = 0.08$) and for all controls (12%, $p = 0.46$). For females, the rates were 7.1% for GBG versus 4.5% for internal controls ($p = 0.50$) and 5.7% for all controls ($p = 0.66$).

The final model for school-based services was a generalized linear mixed effects model that included main effects of baseline aggressive, disruptive behavior, gender, and intervention conditions. It also included all two-way and three-way interactions among gender, baseline, and intervention conditions. Because the gender by baseline interaction was significant, we did a separate analysis on males. The GBG main effect did not appear significant, but the combined GBG impact including the interaction between GBG and baseline was significant ($p = 0.001$). Males whose teachers rated them as exhibiting higher levels of aggressive, disruptive behavior in first grade—the high-risk males—benefited more from GBG than males rated to have lower levels of aggressive, disruptive behavior; see Table 6 for the final model for males.

We also report the results of testing for intervention impact on the use of school-based services that explicitly take into account the blocking factor of school. For the combined sample of males and females, the Mantel-Haenszel continuity corrected test that conditioned on school did not reveal a significant association between GBG and a lower probability of school-based services ($\chi^2 = 0.29$ on 1 *df*, $p = 0.59$). The corresponding small sample t-test result of the value of the log odds ratio accounting for school as a blocking factor was $t = -2.01$ on 5 *df*, $p = 0.10$. For males, the Mantel-Haenszel test result was $\chi^2 = 2.54$ on 1 *df*, $p = 0.11$, and the corresponding paired t-test result of intervention effect using the log odds ratios controlling for each school was $t = -1.96$ on 5 *df*, $p = 0.11$. For females, the Mantel-Haenszel test result was $\chi^2 = 0.19$ on 1 *df*, $p = 0.66$ and t-test result was $t = -0.43$ on 5 *df*, $p = 0.70$.

3.4 GBG Impact on Service Use from Drug Treatment Provider

3.4.1 Cohort 1—Overall unadjusted rates of service use through the drug treatment sector were not significantly lower for youth in the GBG condition (2.7%) compared with internal controls (5.6%, $p = 0.20$) and all controls (5.2%; $p = 0.18$). For males, the rates were 3.9% for GBG versus 11% for internal controls ($p = 0.12$) and 8.8% for all controls, ($p = 0.16$). For females, the rates were 1.9% for GBG versus 1.4% for internal controls ($p = 0.82$) and 2.2% for all controls ($p = 0.86$).

The final model for services used in the drug treatment sector was a generalized linear mixed effects model that included baseline aggressive, disruptive behavior, gender, and intervention conditions. The *p*-value of GBG impact was 0.30. The baseline aggressive, disruptive behavior was a strongly significant predictor ($p < 0.001$). The final combined-gender model contained classroom as random effects, which was significant ($p < 0.001$); see Table 7.

We also report the results of testing for intervention impact on the use of services through the drug treatment sector that explicitly take into account the blocking factor of school. When we combined females and males, the Mantel-Haenszel continuity corrected test that conditioned on school did not find a significant association between GBG and a lower probability of use of services in the drug treatment sector ($\chi^2 = 0.73$ on 1 *df*, $p = 0.39$). The corresponding small sample test result of the value of the log odds ratio accounting for school as a blocking factor was $t = -1.63$ on 5 *df*, $p = 0.16$. For males, the Mantel-Haenszel test result was $\chi^2 = 0.98$ on 1 *df*, $p = 0.32$. The corresponding paired t-test result of intervention effect using the log odds ratios controlling for each school was $t = -2.24$ on 5 *df*, $p = 0.07$. For females, the Mantel-Haenszel test result was $\chi^2 = 0.0003$ on 1 *df*, $p = 0.99$ and the t-test result was $t = -2.89$ on 5 *df*, $p = 0.03$.

3.4.2 Cohort 2—Overall unadjusted rates of service use through the drug treatment sector were not significantly lower for youth in the GBG condition (3.5%) compared with internal controls (6%, $p = 0.33$) and all controls (5.1%, $p = 0.39$). For males, the rates were 4.7% for GBG versus 10% for internal controls ($p = 0.23$) and 5.8% for all controls ($p = 0.71$). For females, the rates were 2.4% for GBG versus 3% for internal controls ($p = 0.81$) and 4.6% for all controls ($p = 0.36$).

The final combined-gender model for services used in the drug treatment sector is a generalized additive mixed effects model that includes main effects of baseline aggressive, disruptive behavior, gender and intervention conditions. The p-value of GBG impact is 0.28. The baseline aggressive, disruptive behavior was not a significant predictor (p=0.30) of the distal outcome. The model contains classroom as random effects, which is not significant (p=0.50), see Table 8.

We also report the results of testing for intervention impact on the use of services through the drug treatment sector that explicitly take into account the blocking factor of school. When we combined females and males, the Mantel-Haenszel continuity corrected test that conditioned on school did not find a significant association between GBG and a lower probability of use of services in the drug treatment sector ($\chi^2 = 0.55$ on 1 *df*, $p = 0.46$). The corresponding small sample test result of the value of the log odds ratio accounting for school as a blocking factor was $t = -1.80$ on 5 *df*, $p = 0.13$. For males, the Mantel-Haenszel test result was $\chi^2 = 1.04$ on 1 *df*, $p = 0.31$. The corresponding paired t-test result of intervention effect using the log odds ratios controlling for each school was $t = -3.01$ on 5 *df*, $p = 0.03$. For females, the Mantel-Haenszel test result was $\chi^2 = 0.06$ on 1 *df*, $p = 0.80$ and the t-test result was $t = -0.84$ on 5 *df*, $p = 0.44$

3.5 GBG Impact on Service Use from Juvenile or Adult Justice System

3.5.1 Cohort 1—Overall rates of service use through the juvenile or adult justice system were 9.3% for youth in the GBG condition compared with 11% in the internal control condition ($p = 0.60$) and 10% for all controls ($p = 0.75$). For males, the rates were 12% for GBG versus 20% for internal controls ($p = 0.21$) and 20% for all controls, ($p = 0.11$). For females, the rates were 7.6% for GBG versus 4.3% for internal controls ($p = 0.39$) and 2.3% for all controls ($p = 0.01$).

The final model for receipt of services through the justice system was a generalized linear mixed effects model that included baseline aggressive, disruptive behavior, gender, intervention conditions and the interactions between gender and intervention conditions. The *p*-value of GBG impact was 0.20. The baseline aggressive, disruptive behavior was a strongly significant predictor ($p = 0.001$). The final combined-gender model contained classroom as random effects, which were significant ($p < 0.001$); see Table 9. Because the interaction between gender and intervention conditions was significant, we did separate analyses on males and females. The GBG impact was not significant ($p = 0.60$) for males. For females, the impact was in a direction contrary to what had been expected but was not significant ($p = 0.20$).

We also report the results of testing for intervention impact on service use through the justice sector, explicitly taking into account the blocking factor of school. When we combined females and males, the Mantel-Haenszel continuity corrected test that conditioned on school did not find a significant association between GBG and a lower probability of using services through the justices sector ($\chi^2 = 0.04$ on 1 *df*, $p = 0.85$). The corresponding small sample test result of the value of the log odds ratio accounting for school as a blocking factor was $t = -0.90$ on 5 *df*, $p = 0.41$. For males, the Mantel-Haenszel test result was $\chi^2 = 0.34$ on 1 *df*, $p = 0.56$. The corresponding paired t-test of intervention effect using the log odds ratios controlling for each school was $t = -1.94$ on 5 *df*, $p = 0.11$. For females, the Mantel-Haenszel test result was $\chi^2 = 0.24$ on 1 *df*, $p = 0.62$ and the t-test results were $t = -0.28$ on 5 *df*, $p = 0.79$.

3.5.2 Cohort 2—Overall rates of service use through the justice system were 7.6% for youth in the GBG condition compared with 9.4% in the internal control condition ($p = 0.60$) and 7.6% for all controls ($p = 0.99$). For males, the rates were 11% for GBG versus 16% for internal controls ($p = 0.36$) and 11% for all controls ($p = 0.98$). For females, the rates were 4.7% for GBG versus 4.5% for internal controls ($p = 0.95$) and 5.0% for all controls ($p = 0.92$).

The final model for receipt of services through the justice system was a generalized linear mixed effects model that included baseline aggressive, disruptive behavior, gender, and intervention conditions. The *p*-value of GBG impact was 0.38. The final combined-gender model contained classroom as random effects, which were significant ($p < 0.001$); see Table 10.

We also report the results of testing for intervention impact on service use through the justice sector, explicitly taking into account the blocking factor of school. When we combined females and males, the Mantel-Haenszel continuity corrected test that conditioned on school did not find a significant association between GBG and a lower probability of using services through the justices sector ($\chi^2 = 0.10$ on 1 *df*, $p = 0.75$). The corresponding small sample test result of the value of the log odds ratio accounting for school as a blocking factor was $t = -0.62$ on 5 *df*, $p = 0.56$. For males, the Mantel-Haenszel test result was $\chi^2 = 0.82$ on 1 *df*, $p = 0.37$. The corresponding paired t-test of intervention effect using the log odds ratios controlling for each school was $t = -1.68$ on 5 *df*, $p = 0.15$. For females, the Mantel-Haenszel test result was $\chi^2 = 0.04$ on 1 *df*, $p = 0.85$ and the t-test results were $t = -0.27$ on 5 *df*, $p = 0.79$.

3.6 GBG Impact on Service Use by Young Adulthood from Social Services

3.6.1 Cohort 1—Overall rates of service use from social services providers for combined gender were marginally significantly lower for youth in the GBG condition (2%) compared with 6% in the internal control condition ($p = 0.06$) and 5% for all controls ($p = 0.05$). For males, the rates were 1% for GBG versus 7% for internal controls ($p = 0.16$) and 6% for all controls ($p = 0.11$). For females, the rates were 2% for GBG versus 4% for internal controls ($p = 0.38$) and 4% for all controls ($p = 0.25$).

The final model for receipt of services through social service providers was a generalized linear mixed model that included baseline aggressive, disruptive behavior, gender, intervention conditions, baseline by intervention conditions interaction terms. Likelihood ratio tests showed that GBG significantly reduced the risk of any social services use when compared with the internal control ($p = 0.02$). There was significant interaction between intervention and baseline aggressive, disruptive behavior ($p = 0.03$). The final combined-gender model contained classroom as random effects, which were not significant ($p = 0.50$); see Table 11.

We also report the results of testing for intervention impact on service use through the social service system, explicitly taking into account the blocking factor of school. For combined females and males, the Mantel-Haenszel continuity corrected test that conditioned on school did not find a significant association between GBG and a lower probability of service use through the social service system ($\chi^2 = 0.67$ on 1 *df*, $p = 0.41$). The corresponding small sample test of the value of the log odds ratio accounting for school as a blocking factor was not significant either ($t = -1.15$ on 5 *df*, $p = 0.30$). The Mantel-Haenszel test result for males was statistically unstable because of a low cell count problem. Meanwhile, the paired t-test of intervention effect using the log odds ratios controlling for each school gave nonsignificant results for males ($t = -1.23$ on 5 *df*, $p = 0.27$). For females, the Mantel-Haenszel test result was not significant ($\chi^2 = 0.006$ on 1 *df*, $p = 0.94$). The corresponding small sample test was not significant either ($t = -1.67$ on 5 *df*, $p = 0.16$).

3.6.2 Cohort 2—Overall rates of service use from social services providers for combined gender were not different for youth in the GBG condition (4.1%) compared with 3.4% in the internal control condition ($p = 0.77$) and 2.5% for all controls ($p = 0.27$). For males, the rates were 3.6% for GBG versus 2% for internal controls ($p = 0.62$) and 0.9% for all controls ($p = 0.10$). For females, the rates were 4.7% for GBG versus 4.5% for internal controls ($p = 0.94$) and 3.8% for all controls ($p = 0.72$). The low count of social services use made these *p*-values not very trustworthy.

The final model for social services was a generalized additive mixed effects model that included main effects of baseline aggressive, disruptive behavior, gender, and intervention conditions. GBG did not show the anticipated impact on lowering the rate of social services use ($p = 0.65$). The baseline aggressive, disruptive behavior was a significant predictor of later service use ($p = 0.04$). The final combined-gender model contained classroom as random effects, which was not significant ($p = 0.50$); see Table 12.

We also report the results of testing for intervention impact on service use through the social service system, explicitly taking into account the blocking factor of school. For combined females and males, the Mantel-Haenszel continuity corrected test that conditioned on school did not find a significant association between GBG and probability of social services use ($\chi^2 = 0.03$ on 1 *df*, $p = 0.87$). The corresponding small sample test of the value of the log odds ratio accounting for school as a blocking factor was not significant either ($t = 0.60$ on 5 *df*, $p = 0.58$). When we separated the analysis by gender, the Mantel-Haenszel tests results were not significant for both ($\chi^2 = 0.16$ on 1 *df*, $p = 0.69$ for males and $\chi^2 = 0.008$ on 1 *df*, $p = 0.93$ for

females). The paired t-test results separated by gender were statistically unstable because of a low cell count problem.

4. Discussion

This study provides evidence of a positive impact of a universal preventive intervention on later service use by males, though not by females, for problems with emotions, behavior, or drugs or alcohol. Impact was seen across both cohorts with regard to the use of school-based services by highly aggressive, disruptive males. The primacy of the education sector in the delivery of mental health services is well established in the services literature; it is the most common sector for service use and often the only sector in which services are received, particularly for children and youth whose problems do not meet a diagnostic criterion (Burns et al., 1995; Farmer et al., 2003; Leaf et al., 1996). This study extends these findings by reporting that males whose teachers rated them high on aggressive, disruptive behavior in the fall of first grade benefited most from GBG with regard to service use through childhood and into young adulthood. In Cohort 1, the proportion of these males in internal control classrooms ever receiving school-based services for problems with behavior, feelings, or drugs and alcohol by young adulthood was 33% compared with 17% of these males in GBG classrooms. In Cohort 2, the proportions were 33% and 18%, respectively.

That the GBG impact was for children at highest risk of poor outcomes, males with high levels of aggressive, disruptive behavior in the fall of first grade, is in keeping with previous reported work from our trials as well as other trials of universal preventive interventions in which intervention impact was found for the students who entered the intervention period at the highest risk (Brown and Liao, 1999; Curran and Muthèn, 1999; Muthèn and Curran, 1997; Muthèn et al., 2002; Stoolmiller et al., 2000). GBG aims at reducing aggressive, disruptive behavior in the first-grade classroom through socializing first graders to the role of student. Although we could not perform formal analysis of mediation, the positive impact of GBG on a subsequent reduction in the use of school-based services appears to be operating through early aggressive, disruptive behavior for males, the target of GBG.

In addition to the consistent pattern of the impact of GBG on use of school-based services by males who entered first grade with higher levels of aggressive, disruptive behavior across both cohorts, there were several less consistent results. Males who had been in GBG classrooms in Cohort 2 reported significantly less service use from mental or medical health professionals than their counterparts in internal GBG control classrooms, 9.4% versus 26%, respectively, whereas in Cohort 1 the GBG and internal GBG control the rates were nearly identical. There was also a positive impact of GBG on ever using social services by young adulthood for the combined population of Cohort 1 females and males (2%) compared with their counterparts in internal GBG control classrooms (6%). This effect was not seen for Cohort 2.

Of note is the consistent positive relationship between early aggressive, disruptive behavior and service use for males. For males in Cohort 1, this relationship was seen across a range of sectors: mental or medical provider, school-based services, drug treatment, juvenile or adult justice system, and social services. The relationship of early aggressive, disruptive behavior and later service use was seen for school-based services and social services for Cohort 2. We saw that GBG disrupted this association for school-based services for males with high levels of early aggressive, disruptive behavior.

For the most part, the absence of any consistent GBG impact for females makes theoretical sense given that early aggressive, disruptive is not consistently associated with later service use for females. However, for females in Cohort 1, there was an unexpected effect such that in GBG classrooms, aggressive, disruptive behavior in the fall of first grade was positively

associated with later service use in some sectors by young adulthood, whereas there was almost no relationship between early aggressive, disruptive behavior and later service use for females in GBG internal control classrooms. This finding is in contrast to the findings reported above for males and is in the opposite direction from our hypotheses. Although this impact of GBG was not specific to any service sector and was not seen for Cohort 2 females, it suggests that our understanding of pathways to service use for females are not as clear as for males.

There are several limitations to consider in this study. The universal preventive intervention, GBG, was experimentally manipulated, but service use from childhood to adulthood was not. Therefore, this study did not allow us to test the overall effectiveness of an integrated service system. However this naturalistic study of service use after the delivery of a preventive intervention does provide evidence that a universal intervention is a worthwhile first stage in crafting an integrated service system. The range of services captured in our definition of service use for each sector is quite broad in terms of cost, intensity, type, and quality. However, the results are strong and establish that a fairly simple universal classroom-based intervention aimed at aggressive, disruptive behavior reduced cumulative service use from school-based mental health services over the remainder of the school years. This paper relied on self-reports of mental health service use across the entire life span to young adulthood. However, it is likely that these young adults were under-reporting, not over-reporting, service use, particularly for services received earlier in life (Horwitz et al., 2001; Leaf et al., 1996).

4.1 Policy Implications

Over the past decade, the prevention of aggressive, disruptive behavior, violence, and drug use have become specific foci of concern to educators in response to schools' concerns for keeping students safe. Increasingly, federal funding for school initiatives is tied to the use of evidence-based programs (i.e., Public Law 107-110). Society at large and communities in particular have to make choices about the allocation of resources and funds. This study, along with the other papers in this issue (Brown et al., in press, this issue; Kellam et al., in press, this issue; Petras et al., in press, this issue; Wilcox et al., in press. this issue) and prior research (Aos et al., 2004), offers data to suggest that it is money well spent to focus on a continuum of interventions and services beginning with the GBG, a relatively inexpensive universal intervention that strengthens the classroom learning environment by giving teachers a tool that many do not have. This study revealed, using an epidemiological sample of first graders followed over development, that for males, early aggressive, disruptive behavior is related to service use across a wide spectrum of service providers and that GBG, a classroom-based universal intervention delivered to all children and aimed at the specific early risk factor of aggressive, disruptive behavior in the classroom, reduced mental health service use received through school-based services by young adulthood. The finding that GBG has an impact on school-based mental health service use has policy implications because service use is a clear, measurable indicator that can be used in economic analysis. By and large, programmatic decisions such as adopting GBG in a school system are most commonly made in the system in which they are implemented. Although the long-term economic impacts of GBG, such as ASPD, drug and alcohol dependence, and suicide ideation, are important to school system personnel from a prevention and theoretical standpoint, these outcomes may be less likely to influence the spending of school system funds than outcomes directly related to school system spending, such as the use of school-based mental health services.

As the first level in an integrated system of services across development, universal interventions such as GBG are important in several ways.

- As we saw in this study, GBG, a universal intervention aimed at a specific risk factor, reduces the proportion of youth who receive services at a later date, specifically males at high risk in first grade.

- Students' responses to GBG and other universal interventions can provide critical information about risk and the types and targets of backup interventions and services needed for students who do not respond.
- Universal interventions often focus on strengthening the environment, which makes them particularly appealing as a first-level intervention. GBG is delivered in the classroom context as an integral part of the school day by the teacher, whose role is precisely defined as providing a positive classroom environment with the students. GBG is in contrast to interventions that focus on individual students or a small group of students and can undermine the interactive group process of teachers and students sharing responsibility for the classroom environment.
- GBG as a universal intervention is aimed at classroom environment and avoids the stigma that children experience by being removed from the classroom.
- GBG is a strategy, not a curriculum. Hence, it does not compete for instructional time. In fact, with less aggressive, disruptive behavior occurring in the classroom, there is more time for instruction in GBG classrooms.
- GBG is relatively inexpensive compared with the benefits (Aos et al., 2004). Although precision in implementation is required, the initial training and continued mentoring and monitoring can be aligned with the professional development programs of a school system.

4.2 Next Stage of Research

The next stage of research will bring together researchers and practitioners from several disciplines, including prevention science, public health, services research, economics, and education research and practice. Above, we note that understanding students' response to GBG over time provides important information about the type, mode, and target for backup services for non-responders across development. Developing multi-level systems of assessment to provide this information is critical to ensuring that all children receive the most appropriate level and kind of care (Ialongo et al., 1999; Kellam and Rebok, 1992; Poduska and Kendziora, 2001). The field of public health has much to offer the field of education, as state and local departments of education create databases that allow following students over time; these databases are reminiscent of the registries common in the public health sector in the middle of the last century and can provide information both across development and across service sectors.

Although GBG has been found to be relatively inexpensive (Aos et al., 2004), economic analyses on the immediate costs to an organization for scaling up and maintaining GBG over time, as well as longer term cost-benefit analyses, are needed. Administrators and policymakers determining resource allocations need this type of information to make well-informed decisions. Epidemiological concepts such as attributable risk, preventive fraction, and positive and negative predictive ability need to be joined with economic analyses to provide practitioners and policymakers with the information they need to make informed decisions. As interventions are deemed to be “effective and evidence-based,” research on moving research into practice is required. Although as an intervention GBG is relatively simple to implement, maintaining a high level of precision in practice is required to ensure the results we report. In the first generation of trials, no structures were in place in the system to support the first-grade teachers continuing the practices into the second year, and there was little support from the research team. In addition, the training provided to the second-grade teachers was minimal compared with the training and support offered to the first-grade teachers. As reported in this issue (Brown et al.; Kellam et al.; Petras et al.; Wilcox et al.), intervention impact for GBG has been strong across a range of outcomes from childhood into young adulthood for the first cohort

of students who participated in the first-generation trial. Although in many cases the same pattern of intervention impact has been found with the second cohort of first graders in the first-generation trial (Kellam et al., in press, this issue) the impact has not been as strong. We hypothesize that the initial train-the-trainer model was somewhat naïve and that without structures in place for ongoing support across the multiple levels of authority and organization, practices are not likely to be maintained. This conclusion is in keeping with the research on sustainability and implementation of programs in schools (Berman and McLaughlin, 1975, 1978; Datnow and Castellano, 2000; Elliott and Mihalic, 2004; Greenwood et al., 1975; McLaughlin, 1990; Olds et al., 2003). We have begun to explore the types of support structures and the level of monitoring and mentoring required for reaching and maintaining precision of practice over time.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

American Federation of Teachers. Building on the best, learning from what works: five promising discipline and violence prevention programs. 2000 [September 28, 2004]. http://www.aft.org/pubs-reports/downloads/teachers/wwdiscipline.pdf

Aos, S.; Lieb, R.; Mayfield, J.; Miller, M.; Pennucci, A. Benefits and Costs of Prevention and Early Intervention Programs for Youth. Washington State Institute for Public Policy; Olympia, WA: 2004.

Barrish H, Saunders M, Wolf M. Good behavior game: effects of individual contingencies for group consequences on disruptive behavior in a classroom. J Appl Behav Anal 1969;2:119–124. [PubMed: 16795208]

Berman, P.; McLaughlin, MW. Federal Programs Supporting Educational Change, Vol IV: The Findings in Review. RAND Corporation; Santa Monica, CA: 1975.

Berman, P.; McLaughlin, MW. Federal Programs Supporting Educational Change, Vol VIII: Implementing and Sustaining Innovations. RAND Corporation; Santa Monica, CA: 1978.

Block, J.; Burns, R. Mastery Learning. In: Shulman, L., editor. Review of Research in Education. Vol. 4. F.E Peacock; Itasca, IL: 1976. p. 3-49.

Brown CH. Analyzing preventive trials with generalized additive models. Am J Community Psychol 1993;21:635–664. [PubMed: 8192125]

Brown CH, Liao J. Principles for designing randomized preventive trials in mental health: an emerging developmental epidemiology paradigm. Am J Community Psychol 1999;27:673–710. [PubMed: 10676544]

Brown CH, Wang W, Kellam SG, Muthén BO, Petras H, Toyinbo P, Poduska J, Ialongo N, Wyman PA, Chamberlain P, Sloboda Z, MacKinnon D, Windham A. Prevention Science Methodology Group. Methods for testing theory and evaluating impact in randomized filed trials: intent-to-treat analysis for integrating the perspective of person, place, and time. Drug Alcohol Depend. in press, this issue

Burns BJ, Costello EJ, Angold A, Tweed D, Stangl D, Farmer EM, Erkanli A. Children's mental health service use across service sectors. Health Aff 1995;14:147–159.

Cicchetti, D.; Schneider-Rosen, K. Toward a transactional model of childhood depression. In: Cicchetti, D.; Schneider-Rosen, K., editors. Childhood Depression: A Developmental Perspective. Jossey-Bass; San Francisco: 1984. p. 5-28.

Cohen P, Kasen S, Brooks JS, Struening EL. Diagnostic predictors of treatment patterns in a cohort of adolescents. J Am Acad Child Adolesc Psychiatry 1991;30:989–993. [PubMed: 1757449]

Costello EJ, Janiszewski S. Who gets treated? Factors associated with referral in children with psychiatric disorders. Acta Psychiatr Scand 1990;81:540–567.

Curran PJ, Muthèn BO. The application of latent curve analysis to testing developmental theories in intervention research. Am J Community Psychol 1999;27:567–595. [PubMed: 10573835]

Datnow A, Castellano M. Teachers' responses to Success for All: how beliefs, experiences, and adaptations shape implementation. Am Educ Res J 2000;37:775–799.

Department of Health and Human Services. Youth violence: A report of the Surgeon General. Rockville, MD: 2001.

Dishion TJ, Spracklen KM, Andrews DW, Patterson GR. Deviancy training in male adolescent friendships. Behav Ther 1996;27:373–390.

Dolan LJ, Kellam SG, Brown CH, Werthamer-Larsson L, Rebok GW, Mayer LS, Laudolff J, Turkkan J, Ford C, Wheeler L. The short-term impact of two classroom-based preventive interventions on aggressive and shy behaviors and poor achievement. J Appl Dev Psychol 1993;14:317–345.

Dominici F, McDermott A, Zeger SL, Samet JM. On the use of generalized additive models in time-series studies of air pollution and health. Am J Epidemiol 2002;156:193–203. [PubMed: 12142253]

Elliott DS, Mihalic S. Issues in disseminating and replicating effective prevention programs. Prev Sci 2004;5:47–53. [PubMed: 15058912]

Ensminger, ME.; Kellam, SG.; Rubin, BR. School and family origins of delinquency: comparisons by sex. In: Van Dusen, KT.; Mednick, SA., editors. Prospective Studies of Crime and Delinquency. Kluwer-Nijhoff Publishing; Boston: 1983. p. 73-97.

Ensminger ME, Slusarcick AL. Paths to high school graduation or dropout: a longitudinal study of first grade cohort. Sociol Educ 1992;65:95–113.

Farmer EMZ, Burns BJ, Phillips SD, Angold A, Costello EJ. Pathways into and through mental health services for children and adolescents. Psychiatr Serv 2003;54:60–66. [PubMed: 12509668]

Farrington DP, Gallagher B, Moorley L, St Ledger RJ, West DJ. Are there any successful men from criminogenic backgrounds? Psychiatry 1988;51:116–130. [PubMed: 3406226]

Farrington, DP.; Gunn, J. Aggression and Dangerousness. John Wiley & Sons; New York: 1985.

Feehan M, Staunton W, McGee R, Silva PA. Parental help-seeking for behavioural and emotional problems in childhood and adolescence. Community Health Stud 1990;14:303–309. [PubMed: 2253467]

Garralda ME, Bailey D. Child and family factors associated with referral to child psychiatrists. Br J Psychiatry 1988;153:81–89. [PubMed: 3224255]

Gerber MM, Semmel MI. Teacher as imperfect test: reconceptualizing the referral process. Educ Psychol 1984;19:137–148.

Gordon RS. An operational classification of disease prevention. Public Health Rep 1983;98:107–109. [PubMed: 6856733]

Greenwood, PW.; Mann, D.; McLaughlin, MW. Federal Programs Supporting Educational Change, Vol III: The Process of Change. RAND Corporation; Santa Monica, CA: 1975.

Guskey, TR. Implementing Mastery Learning. Vol. 2nd. Wadsworth Publishing Co.; Belmont, CA: 1997.

Hastie, T.; Tibshirani, R. Generalized Additive Models. Chapman & Hall; London: 1990.

Hoagwood K, Horwitz SM, Stiffman A, Weisz J, Bean D, Rae D, Compton W, Cottler L, Bickman L, Leaf P. Concordance between parent reports of children's mental health services and service records: the Service Assessment for Children and Adolescents (SACA). J Child Fam Stud 2000;9:315–331.

Hocutt AM. Effectiveness of special education: is placement the critical factor? The Future of Children: Special Education for Students with Disabilities 1996;6:77–102.

Horwitz SM, Hoagwood K, Stiffman AR, Summerfeld T, Weisz JR, Costello EJ, Rost K, Bean DL, Cottler L, Leaf PJ, Roper M, Norquist G. Reliability of the services assessment for children and adolescents. Psychiatr Serv 2001;52:1088–1094. [PubMed: 11474056]

Ialongo, N.; Kellam, SG.; Poduska, J. A developmental epidemiologic framework for clinical child and pediatric psychology research. In: Drotar, D., editor. Conducting Research in Pediatric and Clinical Child Psychology: A Handbook for Practical Strategies. Kluwer Academic/Plenum Press; New York: 1999. p. 1-25.

Ialongo N, Poduska J, Werthamer L, Kellam S. The distal impact of two first-grade preventive interventions on conduct problems and disorder in early adolescence. J Emot Behav Disord 2001;9:146–160.

John LH, Offord DR, Boyle MH, Racine YA. Factors predicting use of mental health and social sciences by children 6-16 years old: findings from the Ontario Child Health Study. Am J Orthopsychiatry 1995;65:76–86. [PubMed: 7733219]

Kellam, SG.; Branch, JD.; Agrawal, KC.; Ensminger, ME. Mental Health and Going to School: The Woodlawn Program of Assessment, Early Intervention, and Evaluation. University of Chicago Press; Chicago: 1975.

Kellam SG, Brown CH, Poduska JM, Ialongo N, Wang W, Toyinbo P, Petras H, Ford C, Windham A, Wilcox HC. Effects of a universal classroom behavior management program in first and second grades on young adult behavioral, psychiatric, and social outcomes. Drug Alcohol Depend. in press, this issue

Kellam, SG.; Brown, CH.; Rubin, BR.; Ensminger, ME. Paths leading to teenage psychiatric symptoms and substance abuse: developmental epidemiological studies in Woodlawn. In: Guze, SB.; Earls, FJ.; Barrett, JE., editors. Childhood Psychopathology and Development. Raven Press; New York: 1983. p. 17-51.

Kellam, SG.; Ensminger, ME. Theory and method in child psychiatric epidemiology. In: Earls, F., editor. International Monograph Series in Psychosocial Epidemiology, Vol 1: Studying Children Epidemiologically. Neale Watson Academic Publishers; New York: 1980. p. 145-180.

Kellam SG, Koretz D, Moscicki EK. Core elements of developmental epidemiologically based prevention research. Am J Community Psychol 1999;27:463–482. [PubMed: 10573831]

Kellam SG, Ling X, Merisca R, Brown CH, Ialongo N. The effect of the level of aggression in the first grade classroom on the course and malleability of aggressive behavior into middle school. Dev Psychopathol 1998;10:165–185. [PubMed: 9635220]Kellam SG, Ling X, Merisca R, Brown CH, Ialongo N. The effect of the level of aggression in the first grade classroom on the course and malleability of aggressive behavior into middle school: Results of a developmental epidemiology-based prevention trial. Erratum Dev Psychopathol 2000;12:107.See also

Kellam, SG.; Rebok, GW. Building developmental and etiological theory through epidemiologically based preventive intervention trials. In: McCord, J.; Tremblay, RE., editors. Preventing Antisocial Behavior: Interventions from Birth through Adolescence. The Guilford Press; New York: 1992. p. 162-195.

Kellam SG, Rebok GW, Ialongo N, Mayer LS. The course and malleability of aggressive behavior from early first grade into middle school: results of a developmental epidemiologically-based preventive trial. J Child Psychol Psychiatry 1994a;35:259–282. [PubMed: 8188798]

Kellam, SG.; Rebok, GW.; Wilson, R.; Mayer, LS. The social field of the classroom: Context for the developmental epidemiological study of aggressive behavior. In: Silbereisen, RK.; Todt, E., editors. Adolescence in Context: The Interplay of Family, School, Peers, and Work in Adjustment. Springer-Verlag; New York: 1994b. p. 390-408.

Kellam SG, Werthamer-Larsson L, Dolan LJ, Brown CH, Mayer LS, Rebok GW, Anthony JC, Laudolff J, Edelsohn G, Wheeler L. Developmental epidemiologically-based preventive trials: baseline modeling of early target behaviors and depressive symptoms. Am J Community Psychol 1991;19:563–584. [PubMed: 1755436]

Kessler R, McGonagle K, Zhao S, Nelson CB, Hughes M, Eshleman S, Wittchen H, Kendler K. Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Arch Gen Psychiatry 1994;51:8–19. [PubMed: 8279933]

Laitinen-Krispijn S, der Ende JV, Wierdsma AI, Verhulst FC. Predicting adolescence mental health service use in a prospcectove record-linkage study. J Am Acad Child Adolesc Psychiatry 1999;38:1073–1080. [PubMed: 10504805]

Leaf PJ, Alegria M, Cohen P, Goodman SH, Horwitz SM, Hoven CW, Narrow WE, Vaden-Kiernan M, Regier DA. Mental health service use in the community and schools: results from the four-community MECA study. J Am Acad Child Adolesc Psychiatry 1996;35:889–897. [PubMed: 8768348]

Lloyd JW, Kauffman JM, Landrum TJ, Roe DL. Why do teachers refer pupils for Special Education? An analysis of referral records. Exceptionality 1991;2:115–129.

Mattison RE, Humphrey FJ 2nd, Kales SN. An objective evaluation of special class placement of elementary schoolboys with behavior problems. J Abnorm Child Psychol 1986;14:251–262. [PubMed: 3722621]

McCord J. Parental behavior in the cycle of aggression. Psychiatry 1988;51:14–23. [PubMed: 3368543]

McLaughlin MW. The rand change agent study revisited: Macro perspectives and micro realities. Educational Researcher 1990;19:11–16.

Mrazek, PG.; Haggerty, RJ. Reducing Risks for Mental Disorders: Frontiers for Preventive Intervention Research. National Academy Press; Washington, DC: 1994.

Murray, DM. Design and Analysis of Group-Randomized Trials. Oxford University Press; New York: 1998.

Muthén BO, Curran PJ. General longitudinal modeling of individual differences in experimental designs: a latent variable framework for analysis and power estimation. Psychol Methods 1997;2:371–402.

Muthén BO, Brown CH, Masyn K, Jo B, Khoo ST, Yang CC, Wang CP, Kellam SG, Carlin JB. General growth mixture modeling for randomized preventive interventions. Biostatistics 2002;3:459–475. [PubMed: 12933592]

Olds DL, Hill PL, O'Brien R, Racine D, Moritz P. Taking preventive intervention to scale: the Nurse-Family Partnership. Cognit Behav Pract 2003;10:278–290.

Patterson, GR.; Reid, JB.; Dishion, TJ. Antisocial Boys. Castalia; Eugene, OR: 1992.

Petras H, Kellam SG, Brown CH, Muthén B, Ialongo NS, Poduska JM. Developmental epidemiological courses leading to antisocial personality disorder and violent and criminal behavior: Effects by young adulthood of a universal preventive intervention in first- and second-grade classrooms. Drug Alcohol Depend. in press, this issue

Poduska JM. Parents' perceptions of their first graders' need for mental health and educational services. J Am Acad Child Adolesc Psychiatry 2000;39:584–591. [PubMed: 10802976]

Poduska, JM.; Kendziora, KT. Integrating public mental health and children's systems of care. Psychiatric Times 18. 2001 [February 20, 2007]. http://www.psychiatrictimes.com/p010950.html

Public Law 107-110. No Child Left Behind Act of 2001. January 8;2002 2001

Reid JB. Prevention of conduct disorder before and after school entry: relating interventions to development findings. Dev Psychopathol 1993;5:243–262.

Reid, JB.; Eddy, JM. The prevention of antisocial behavior: some considerations in the search for effective interventions. In: Stoff, DM.; Breiling, J.; Maser, JD., editors. Handbook of Antisocial Behavior. John Wiley & Sons; New York: 1997. p. 343-356.

Rickwood DJ, Braithwaite VA. Social-psychological factors affecting help-seeking for emotional problems. Soc Sci Med 1994;39:563–572. [PubMed: 7973856]

Robins LN. Sturdy childhood predictors of adult antisocial behavior: replications from longitudinal studies. Psychol Med 1978;8:611–622. [PubMed: 724874]

Shinn MR, Tindal GA, Spira DA. Special education referrals as an index of teacher tolerance: are teachers imperfect tests? Except Child 1987;54:32–40. [PubMed: 3653216]

Staghezza-Jaramillo B, Bird HR, Gould MS, Canino G. Mental health service utilization among Puerto Rican children ages 4 through 16. J Child Fam Stud 1995;4:399–418.

Stiffman AR, Horwitz SM, Hoagwood K, Compton W, Cottler L, Bean DL, Narrow WE, Weisz JR. The Service Assessment for Children and Adolescents (SACA): adult and child reports. J Am Acad Child Adolesc Psychiatry 2000;39:1032–1039. [PubMed: 10939232]

Stoolmiller M, Eddy JM, Reid JB. Detecting and describing preventive intervention effects in a universal school-based randomized trial targeting delinquent and violent behavior. J Consult Clin Psychol 2000;68:296–306. [PubMed: 10780130]

Tingstrom DH, Sterling-Turner HE, Wilczynski SM. The good behavior game: 1969-2002. Behav Modif 2006;30:225–253. [PubMed: 16464846]

Vaden-Kiernan N, Ialongo NS, Pearson J, Kellam S. Household family structure and children's aggressive behavior: a longitudinal study of urban elementary school children. J Abnorm Child Psychol 1995;23:553–568. [PubMed: 8568079]

Verhulst FC, der Ende JV. Factors associated with child mental health services use in the community. J Am Acad Child Adolesc Psychiatry 1997;36:901–909. [PubMed: 9204667]

Walter HJ, Gouze K, Lim KG. Teachers' beliefs about mental health needs in inner city elementary schools. J Am Acad Child Adolesc Psychiatry 2006;45:61–68. [PubMed: 16327582]

Werthamer-Larsson L, Kellam SG, Wheeler L. Effect of first grade classroom environment on child shy behavior, aggressive behavior, and concentration problems. Am J Community Psychol 1991;19:585–602. [PubMed: 1755437]

Wilcox HC, Kellam SG, Brown CH, Poduska J, Ialongo NS, Wang W, Anthony J. The impact of two universal randomized first grade interventions on young adult suicide-related behaviors. Drug Alcohol Depend. in press, this issue

Wu P, Hoven CW, Bird HR, Moore RE, Cohen P, Alegria M, Dulcan MK, Goodman SH, Horwitz SM, Lichtman JH, Narrow WE, Rae DS, Regier DA, Roper MT. Depressive and disruptive disorders and mental health service utilization in children and adolescents. J Am Acad Child Adolesc Psychiatry 1999;38:108.

Table 1

GLMM Model for Any Service Use for Males (N = 279 youth, Cohort 1)

Type of Effect	Coefficient (SE)	*d.f.*	*t-value*	*p-value*
Main Effects				
Intercept	-1.474 (0.488)	247	-3.022	0.003
Baseline Aggression	1.469 (0.303)	247	4.855	0.000
GBG vs Internal GBG Control (Tx1)	-0.861 (0.578)	27	-1.491	0.148
External Control vs Internal GBG Control (Tx2)	-0.006 (0.539)	27	-0.011	0.991
Internal ML Control vs. Internal GBG Control (Tx3)	0.097 (0.583)	27	0.165	0.870
	SD		*p-value*[a]	
Random Effects				
Classroom	0.717		0.015	

[a] for testing zero variance

Table 2

GLMM Model for Any Service Use for Males (N = 289 youth, Cohort 2)

Type of Effect	Coefficient (SE)	*d.f.*	*t-value*	*p-value*
Main Effects				
Intercept	-0.898 (0.363)	255	-2.473	0.014
Baseline Aggression	0.871 (0.306)	255	2.848	0.005
GBG vs Internal GBG Control (Tx1)	-0.928 (0.412)	29	-2.250	0.032
External Control vs Internal GBG Control (Tx2)	-0.832 (0.391)	29	-2.131	0.042
Internal ML Control vs. Internal GBG Control (Tx3)	-0.678 (0.426)	29	-1.593	0.122
	SD		*p-value*[a]	
Random Effects				
Classroom	0.001		0.497	

[a]for testing zero variance

Table 3

GLMM Model for Service Use from Mental or Medical Health Provider for Males and Females Combined (N = 627 youth, Cohort 1)

Type of Effect	**Coefficient (SE)**	***d.f.***	***t-value***	***p-value***
Main Effects				
Intercept	-3.424 (0.732)	593	-4.678	0.000
Baseline Aggression	2.457 (0.767)	593	3.205	0.001
Gender	0.848 (0.371)	593	2.282	0.023
GBG vs Internal GBG Control (Tx1)	-0.058 (0.457)	27	-0.127	0.900
External Control vs Internal GBG Control (Tx2)	0.346 (0.433)	27	0.800	0.431
Internal ML Control vs. Internal GBG Control (Tx3)	0.173 (0.471)	27	0.366	0.717
	Coefficient (SE)	*d.f.*	*t-value*	*p-value*
Interaction Effects				
Gender By Baseline Aggression	-1.437 (0.519)	593	-2.766	0.006
	SD		*p-value*[a]	
Random Effects				
Classroom	0.531		0.000	

[a]for testing zero variance

Table 4

GLMM Model for Service Use from Mental or Medical Health Provider for Males (N = 289 youth, Cohort 2)

Type of Effect	Coefficient (SE)	*d.f.*	*t-value*	*p-value*
Main Effects				
Intercept	-1.156 (0.438)	255	-2.639	0.009
Baseline Aggression	0.448 (0.371)	255	1.206	0.229
GBG vs Internal GBG Control (Tx1)	-1.308 (0.544)	29	-2.404	0.023
External Control vs Internal GBG Control (Tx2)	-1.03 (0.494)	29	-2.087	0.046
Internal ML Control vs. Internal GBG Control (Tx3)	-1.051 (0.553)	29	-1.902	0.067
	SD		*p-value*[a]	
Random Effects				
Classroom	0.464		0.000	

[a]for testing zero variance

Table 5

GAMM for School-based Service Use for Males (N = 278 youth, Cohort 1)

Type of Effect	**Coefficient (SE)**	***t-value***	***p-value***
Main Effects (Linear)			
Intercept	0.472 (1.221)	0.387	0.699
GBG vs Internal GBG Control (Tx1)	-6.110 (3.476)	-1.758	0.080
External Control vs Internal GBG Control (Tx2)	-2.694 (1.642)	-1.641	0.102
Internal ML Control vs. Internal GBG Control (Tx3)	-2.246 (1.485)	-1.512	0.132
	d.f.	χ^2*-value*	*p- value*
Main Effects (Nonlinear)			
Baseline Aggression			
Total	3	15.233	0.002
Linear	1	9.366	0.002
Smooth	2	5.867	0.053
	d.f.	χ^2*-value*	*p-value*
Interaction Effects (Nonlinear)			
Baseline Aggression * GBG vs. internal control (Tx1)			
Total Baseline * Tx1	3	6.650	0.084
Linear (Baseline * Tx1	1	0.237	0.626
Smooth (Baseline*Tx1)	2	6.413	0.040
Baseline Aggression * external control vs. internal GBG control (Tx2)			
Total Baseline * Tx2	3	6.609	0.085
Linear (Baseline * Tx2)	1	0.644	0.422
Smooth (Baseline*Tx2)	2	5.965	0.051
Baseline Aggression * internal ML vs. internal GBG control (Tx3)			
Total Baseline * Tx3	3	5.519	0.137
Linear (Baseline * Tx3)	1	0.115	0.734
Smooth (Baseline*Tx3)	2	5.404	0.067
	SD	*p-value*[a]	
Random Effects			
Classroom	0.656	0.500	

[a]for testing zero variance

Table 6

GLMM Model for School-based Service Use for Males (N = 289 youth, Cohort 2)

Type of Effect	**Coefficient (SE)**	***d.f.***	***t-value***	***p-value***
Main Effects				
Intercept	-3.247 (1.045)	252	-3.108	0.002
Baseline Aggression	2.750 (1.231)	252	2.234	0.026
GBG vs Internal GBG Control (Tx1)	-0.602 (1.435)	29	-0.419	0.678
External Control vs Internal GBG Control (Tx2)	0.377 (1.250)	29	0.302	0.765
Internal ML Control vs. Internal GBG Control (Tx3)	0.459 (1.243)	29	0.370	0.714
	Coefficient (SE)	*d.f.*	*t-value*	*p-value*
Interaction Effects				
GBG By Baseline Aggression	-0.691 (1.555)	252	-0.445	0.657
External Control By Baseline Aggression	-1.944 (1.440)	252	-1.350	0.178
Internal ML Control By Baseline Aggression	-1.578 (1.469)	252	-1.074	0.284
	SD		*p-value*[a]	
Random Effects				
Classroom	0.001		0.493	

[a]for testing zero variance

Table 7

GLMM Model for Drug Treatment Service Use for Males and Females Combined (N = 625 youth, Cohort 1)

Type of Effect	Coefficient (SE)	*d.f.*	*t-value*	*p-value*
Main Effects				
Intercept	-3.158 (0.890)	592	-3.547	0.000
Baseline Aggression	1.755 (0.386)	592	4.546	0.000
Gender	-0.722 (0.350)	592	-2.062	0.040
GBG vs Internal GBG Control (Tx1)	-0.920 (0.874)	27	-1.052	0.302
External Control vs Internal GBG Control (Tx2)	-0.588 (0.826)	27	-0.712	0.482
Internal ML Control vs. Internal GBG Control (Tx3)	0.379 (0.855)	27	0.443	0.661
	SD		p-value[a]	
Random Effects				
Classroom	1.258		0.000	

[a]for testing zero variance

Table 8

GAMM Model for Drug Treatment Service Use for Males and Females Combined (N = 605 youth, Cohort 2)

Type of Effect	**Coefficient (SE)**	***t-value***	***p-value***
Main Effects (Linear)			
Intercept	-2.233 (0.696)	-3.210	0.001
Gender	-0.268 (0.366)	-0.734	0.463
GBG vs Internal GBG Control (Tx1)	-0.640 (0.594)	-1.078	0.281
External Control vs Internal GBG Control (Tx2)	-0.228 (0.519)	-0.438	0.661
Internal ML Control vs. Internal GBG Control (Tx3)	-0.447 (0.604)	-0.739	0.460
Main Effects (Nonlinear)	*d.f.*	*χ^2-value*	*p- value*
Baseline Aggression			
Total Baseline	3	2.140	0.544
Linear (Baseline)	1	1.044	0.295
Smooth (Baseline)	2	1.096	0.593
	SD	*p-value*[a]	
Random Effects			
Classroom	0.416	0.500	

[a]for testing zero variance

Table 9

GLMM Model for Justice System Service Use for Males and Females Combined (N = 625 youth, Cohort 1)

Type of Effect	Coefficient (SE)	*d.f.*	*t-value*	*p-value*
Main Effects				
Intercept	-1.868 (1.114)	589	-1.676	0.094
Baseline Aggression	1.846 (0.311)	589	5.933	0.000
Gender	-1.451 (0.752)	589	-1.930	0.054
GBG vs Internal GBG Control (Tx1)	-1.789 (1.352)	27	-1.324	0.197
External Control vs Internal GBG Control (Tx2)	1.465 (1.361)	27	1.077	0.291
Internal ML Control vs. Internal GBG Control (Tx3)	1.993 (1.559)	27	1.279	0.212
	Coefficient (SE)	*d.f.*	*t-value*	*p-value*
Interaction Effects				
Gender By Tx1	1.374 (0.898)	589	4.5296	0.127
Gender By Tx2	-0.808 (1.012)	589	-0.799	0.425
Gender By Tx3	-1.297 (1.211)	589	-1.071	0.285
	SD		*p-value*[a]	
Random Effects				
Classroom	0.871		0.000	

[a]for testing zero variance

Table 10

GLMM Model for Justice System Service Use for Males and Females Combined (N = 605 youth, Cohort 2)

Type of Effect	Coefficient (SE)	*d.f.*	*t-value*	*p-value*
Main Effects				
Intercept	-1.073 (0.622)	570	-1.724	0.085
Baseline Aggression	0.274 (0.344)	570	0.798	0.425
Gender	-0.831 (0.320)	570	-2.595	0.010
GBG vs Internal GBG Control (Tx1)	-0.410 (0.462)	29	-0.888	0.382
External Control vs Internal GBG Control (Tx2)	-0.399 (0.434)	29	-0.919	0.366
Internal ML Control vs. Internal GBG Control (Tx3)	-0.482 (0.496)	29	-0.971	0.339
	SD		*p-value*[a]	
Random Effects				
Classroom	0.284		0.000	

[a]for testing zero variance

Table 11

GLMM Model for Social Service Use for Males and Females Combined (N = 625 youth, Cohort 1)

Type of Effect	**Coefficient (SE)**	***d.f.***	***t-value***	***p-value***
Main Effects				
Intercept	-2.209 (1.257)	589	-1.757	0.079
Baseline Aggression	-1.817 (1.682)	589	-1.080	0.281
Gender	-0.096 (0.535)	589	-0.179	0.858
GBG vs Internal GBG Control (Tx1)	-4.705 (2.225)	27	-2.114	0.044
External Control vs Internal GBG Control (Tx2)	-0.724 (1.013)	27	-0.714	0.481
Internal ML Control vs. Internal GBG Control (Tx3)	-1.450 (1.109)	27	-1.307	0.202
	Coefficient (SE)	*d.f.*	*t-value*	*p-value*
Interaction Effects				
GBG By Baseline	4.934 (2.261)	589	2.182	0.030
External Control By Baseline Aggression	2.230 (1.817)	589	1.265	0.206
Internal ML Control By Baseline Aggression	2.825 (1.909)	589	1.480	0.139
	SD		*p-value*[a]	
Random Effects				
Classroom	0.006		0.500	

[a]for testing zero variance

Table 12

GAMM Model for Social Service Use for Males and Females Combined (N = 603 youth, Cohort 2)

Type of Effect	**Coefficient (SE)**	***t-value***	***p-value***
Main Effects (Linear)			
Intercept	-5.424 (1.108)	-4.897	0.000
Gender	1.236 (0.538)	2.297	0.022
GBG vs Internal GBG Control (Tx1)	0.292 (0.637)	0.459	0.647
External Control vs Internal GBG Control (Tx2)	-0.345 (0.652)	-0.53	0.597
Internal ML Control vs. Internal GBG Control (Tx3)	-0.768 (0.872)	-0.881	0.379
	d.f.	*χ^2-value*	*p-value*
Main Effects (Nonlinear)			
Baseline Aggression			
Total Baseline	3	5.744	0.124
Linear (Baseline)	1	4.149	0.042
Smooth (Baseline)	2	1.594	0.451
	SD	*p-value*[a]	
Random Effects			
Classroom	0.416	0.500	

[a]for testing zero variance