The Long-Term Effects of After-School Programming on Educational Adjustment and Juvenile Crime: A Study of the LA's BEST After-School Program

Final Deliverable – June 2007

Pete Goldschmidt and Denise Huang, Project Directors
National Center for Research on Evaluation, Standards, and Student Testing (CRESST)
University of California, Los Angeles

U.S. Department of Justice
National Institute of Justice
Award # 2004-SI-FX-0032

National Center for Research on Evaluation, Standards, and Student Testing (CRESST)
Center for the Study of Evaluation (CSE)
Graduate School of Education & Information Studies
University of California, Los Angeles
GSE&IS Building, Box 951522
Los Angeles, CA 90095-1522
(310) 206-1532

The work reported herein was supported under the National Institute of Justice, PR/Award Number 2004-SI-FX-0032, as administered by National Institute of Justice, U.S. Department of Justice.

The findings and opinions expressed in this report do not reflect the positions or policies of the National Institute of Justice, U.S. Department of Justice.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

THE LONG-TERM EFFECTS OF AFTER-SCHOOL PROGRAMMING ON EDUCATIONAL ADJUSTMENT AND JUVENILE CRIME: A STUDY OF THE LA'S BEST AFTER-SCHOOL PROGRAM

Pete Goldschmidt, Denise Huang & Marjorie Chinen

CRESST/University of California, Los Angeles

Abstract

Widespread interest in the impact of after-school programs on youth development has increased dramatically over the past several years. Although research has investigated the short-term impact of programs on academic and social student development, there is limited research on the long-term effectiveness of after-school programs in lowering rates of juvenile crime. This study bridges that research gap and presents results from an evaluation of the effectiveness of LA's BEST - the largest urban-based, after-school program in Los Angeles County - on long-term academic achievement growth and juvenile crime. This research tracked the academic and juvenile crime histories for a sample of 6,000 students, 2,000 students participating in LA's BEST and 4,000 matched control students not participating in LA' s BEST. We used multilevel propensity scores to match control to treatment students and applied multilevel longitudinal models and multilevel survival analyses methods to analyze the data. Results indicate that students' engagement in the program is a strong mediating factor of program effectiveness. The key element of positive program impact is student engagement, as indicated by a medium to high average monthly attendance, and by significant adult contact of at least one additional adult (volunteer) per day. Student participants, who attended sites with a higher average of adult volunteerism, demonstrate modest achievement gains compared to students who did not participate in LA's BEST. Likewise, students who consistently attended LA's BEST demonstrate a substantively significant reduction in the juvenile crime hazard compared to participants with inconsistent attendance, and compared to students in the control group. Benefit-cost analyses indicate that results are sensitive to assumptions regarding the value of avoided costs from prevented crimes.

Introduction

Each year hundreds of millions of dollars are spent on funding after-school programs in the United States. For the 2003 fiscal year, Congress appropriated approximately one billion dollars to be used for this purpose (U.S. Department of Education, 2002). While this reflects the importance that the public places on after-school programs, there has been little long-term assessment of the effectiveness of after-school programs in lowering incidences of juvenile crime. After-school programs potentially have many positive effects on juveniles, and given that the annual cost of juvenile crime is estimated to be approximately $56.7 billion dollars (Caldwell, Vitacco & Rybroek, 2006), the impact of after-school programs on juvenile crime warrants continued analysis. In addition, few studies have included a benefit-cost analysis of after-school programs on the effects of juvenile crime. This study contributes to the research gap in understanding the connection between LA's BEST (Los Angeles' Better Educated Students for Tomorrow), the largest after-school program in Los Angeles County, long-term academic achievement, and juvenile crime.

This study has two major goals[1]. The first goal is to examine the long-term relationship between participation in LA's BEST and academic achievement. The second goal is to investigate the impact of LA's BEST on reducing the juvenile crime hazard, and conducting a benefit-cost analysis based on the effectiveness results. Accordingly, the three main research questions for this study are as follows:

- Is there a difference in the long-term educational outcomes of LA's BEST participants in comparison with non-participants?
- Is there a difference in the students' rate of committing juvenile crimes among LA's BEST participants and non-participants?
- What is the cost-effectiveness of LA's BEST in terms of students' long-term juvenile crime hazard?

[1] It should be noted here that the study originally planned to examine the relationship between LA's BEST attendance and the likelihood a student would drop-out of high school. However, the district has only recently begun keeping track of drop-out status in its individual data base. In addition, because procedures for calculating drop-out status have changed in the last two years, we did not have reliable nor consistent data for the purposes of the current study.

Since millions of dollars of public and private funding are invested in after-school programs each year, this study on the long-term effects of LA's BEST on juvenile delinquency and educational adjustment will be particularly salient for policymakers, law enforcement officials, and educators. Even though studies on the impact of after-school programs are growing in number, the majority of studies focus solely on the short-term social and academic outcomes. There is a clear lack of research on long-term outcomes, particularly with regard to the impact of after-school participation on juvenile criminal offenses. This study intends to fill this research gap. Further, this study examines the moderating and mediating factors that potentially impact program effectiveness.

Review of the Literature

Positive Effects of After-School Programs

The research literature indicates that there are a multitude of risk factors associated with juvenile delinquency, and that these risks are present in the everyday lives of many urban children and adolescents. Mayer (2001) argues that aversive or punitive environments in the home, community, and school contribute to antisocial behaviors such as aggression, vandalism, rule infraction, defiance of adult authority, and other violations of social norms. Hawkins and colleagues (2000) assert that the following individual factors all contribute to youth violence: physical health and involvement in or beliefs favorable towards antisocial behavior; family factors such as home stability, parent involvement, and parental values; school factors such as academic failure, low bonding to school, and truancy; peer-related factors such as sibling and peer aspirations and gang membership; and community and neighborhood factors such as poverty, community disorganization, and exposure to drugs, criminal adults, violence, and racial prejudice. Similarly, Carr and Vandiver (2001) contend that the risk factors associated with youth offenders include engagement in problem behavior such as dropping out of school, poor self-concept and low self-esteem, interpersonal inadequacy, poor educational expectations, troublesome attitudes, poor parenting and family stability, negative peer relationships, large number of siblings at home, drug use, and poor academics and school attendance.

In order to counter juvenile delinquency, Carr and Vandiver (2001) affirm that children need to have access to protective buffers that will decrease the likelihood of them engaging in problematic antisocial and anti-school behaviors, and increase the likelihood of them developing into competent and successful adolescents. Masten, Hubbard, Gest, Tellegen, Garmezy and Ramirez (1999) state that children and adolescents develop competence through psychosocial resources, and maladaptive adolescents tend to have faced adverse conditions with fewer resources over time. Encouragingly, these researchers also declare that adolescents with access to resources, even those in adverse environments, could develop good competency skills.[2] (Cairns & Cairns, 1994). Their findings echo the results of resiliency research, which indicates that healthy and successful individuals share common properties of high self-esteem, positive attitudes toward life, future aspirations, and impulse control (Garmezy, 1991; Krovetz, 1999; Reynolds, 1998; Richman & Fraser, 2001; Wang, Haertel & Walberg, 1998; Werner & Smith, 1992). These youth are able to achieve resiliency in part because they have had opportunities to develop affirming personal relationships, learn about the importance of school, and gain a sense of well-being. According to a study of aggressive boys from elementary school to their pre-teen years, O'Donnell, Hawkins, and Abbott (1995) aver that social skills, academic achievement, and protection from antisocial peers and adults are significant in preventing delinquency. Such opportunities may arise from quality after-school programs with a focus on building resiliency against academic, social, and other risks faced by students (Cassell, Chow, Demoulin, & Reiger, 2000).

Newman, Fox, Flynn, and Christeson (2000) stress that after-school programs can reduce juvenile crime and violence, reduce substance abuse, reduce teen sex and pregnancies, and boost academic success and school completion. After-school programs are beneficial to student resiliency and the prevention of juvenile delinquency in three critical ways. First, after-school programs provide children with supervision during a time when they might normally fall prey to deviant or antisocial behaviors. Research shows that the rates for both violent juvenile crimes and victimization of juveniles peak between 3 and 6 p.m. on school days (Newman et al., 2000; Richardson, Radziszweska, Dent, & Flay, 1993; U.S. Department of Education & U.S. Department of Justice, 2000). In addition, school-based interventions can increase students' feelings of attachment to school and provide them with skills needed to avoid delinquent behaviors (Greenwood, Model, Rydell, & Chiesa, 1998). According to DeKalb (1999), after-school programs can also reduce student truancy, which is a key predictor of juvenile delinquency. Secondly, after-school programs provide experiences that may

[2] Skills which are positively associated with educational attainment and life choices

benefit students' social skills and classroom conduct. Children who participate in these programs tend to exhibit better behavior in school and higher academic achievement, better social skills and self-control, and improved self-confidence through the development of positive relationships with adults and peers (Scott-Little, Hamann & Jurs, 2002). Students can also benefit from the extra-curricular activities that many after-school programs offer. According to the Carolina Longitudinal Study (Cairns & Cairns, 1994), extracurricular activity participation is associated with low rates of early school drop-out (Mahoney & Cairns, 1997) and low rates of criminal arrest in young adulthood (Mahoney, 2000). Cassell et al. (2000) posit that heavy extracurricular involvement helps to dissuade students from becoming involved with delinquency. Finally, after-school programs may help improve academic achievement (Fashola, 1998). Students who participate in these programs often are more positive about school and their own schoolwork, and are more likely to have ambitions to graduate from high school and attend college (U.S. Department of Education & U.S. Department of Justice, 2000).

There are a number of after-school program elements that can foster youngsters' resiliency. In a longitudinal risk-and-resiliency study of 178 African American elementary students in Milwaukee, Wisconsin, Vandell, and Posner (1999) hypothesize that fifth graders' school adjustment is negatively associated with exposure to negative elements in their home area. However, they believe that high-quality neighborhood programs such as after-school programs could provide a safe haven and contribute positively to children's development. In another study of 71 resilient, at-risk, elementary school students, Westfall and Pisapia (1994) conjecture that these students spend their spare time involved in a variety of extracurricular activities and hobbies that provide them with meaningful relationships and social-psychological support. In examining the self-concept and motivational patterns of 36 urban Hispanic youth, Gordon (1996) finds that resilient youth, as opposed to non-resilient youth, has more confidence in their cognitive abilities, such as understanding materials in class and believing that they could do well in homework. These students are also more academically engaged—a characteristic that has been shown to help reduce truancy and eventual delinquency. In a study of 1,170 black urban youth drawn from the Chicago Longitudinal Study, Reynolds (1998) asserts that the factors most highly correlated with resilience are academic achievement, classroom adjustment, perceived competence, and parental participation. All of these factors can be fostered in extracurricular programming of after-school programs.

While the immediate effects of after-school programs have been well studied, relatively few have examined the enduring effects of after-school programs on students' academic and social well-being. One of the few long-term effect studies was conducted by Smith, Hill, and Bandera (1997), who surveyed high school students who had been in an after-school program as fifth-graders. While this study finds no academic differences between the students and the control group, some social differences are noticed. High school students, who were former after-school program participants, generally like school better and plan to stay in school longer—helping to prevent truancy or school drop-out.

The Importance of a Cost-Benefit Analysis

Several recent studies have moved beyond traditional educational program evaluation and utilized benefit-cost analyses to observe whether programs are efficient and have a positive net present value or a benefit-cost ratio greater than 1[3] (Hanushek & Lockeed, 1988). For this reason, recent efforts have examined the link between the roots of student delinquency and the benefits gained by program participants, especially focusing upon interventions that deal with academic improvement and social skills. Several studies demonstrate net present values to child and youth interventions. Programs such as the Federal program for Women, Infants, and Children (WIC), the Abcederen project, Perry Preschool, Success for All, and Chicago Child Parent Centers returned between $3 and $9 for each $1 spent on program costs (Reynolds, Temple, Robertson & Mann, 2002; Barnett; 1995; Schweinhart, Barnes & Weikart, 1993; Karoly et al, 1998). As noted, the majority of these analyses are based on either short term effects and/or small samples. The most notable benefit-cost evaluation examining Perry Preschool, for example, is based on about 230 participants. Another recent analysis that yields benefit-costs ratios of approximately 7 (Caldwell et al., 2006) is based on a sample of 200.

The reported benefits of after-school programs tend to rely on estimates of avoided costs associated with crime as these costs makeup the majority of the benefits.

[3] It is possible to compare the relative effectiveness among disparate programs and consider whether individual programs are the best use of resources, if program effects are assigned to a common metric, such as dollars. This notion expands much of the educational program evaluation literature that often underestimates the potential benefits associated with an intervention by focusing on achievement outcomes.

The estimated benefits from crime avoidance are greatly influenced by victims' costs (Caldwell et al., 2006), of which 65% are intangible costs (Cohen, 1998).

Further, summary reports of the profits result from after-school programs combine all the benefits derived from several programs to estimate total benefits potentially accruing to a single program (Brown, Frates, Rudge & Tradewell, 2002). However, there is no evidence that any single program can reproduce all the benefits from all potential sources.

In summary, research indicates that after-school programs are a potentially powerful resource that can help reduce juvenile delinquency rates. Quality after-school programs teach students the academic and social skills they need to avoid the anti-school behaviors and attitudes that contribute to juvenile delinquency. For instance, after-school programs have been found to increase academic engagement, which in turn helps to prevent truancy and future delinquency. After-school programs also provide potentially structured and safe environments and offer a variety of extra-curricular activities during a time when juvenile crime rates peak. Participation in extra-curricular activities can provide opportunities for advancing students' interpersonal competence, inspiring challenging life goals, promoting educational success, thus reducing juvenile crime. However, despite the important role that after-school programs potentially have in the prevention of juvenile delinquency, this topic has been largely overlooked in research. This study fills the research gap by examining the long-term impact of after-school programming, specifically LA's BEST[4], on juvenile delinquency and academic achievement; it tackles issues in examining the long-term educational and social adjustments of after-school participants. This research also contributes in our understanding of why after-school programs offer positive long-term effects by estimating benefit-cost ratios of programs on juvenile crime reduction. This study argues that after school programs like LA's BEST not only reduce negative effects for disadvantaged youth, but provide more of a benefit to cultivating resilient individuals who do better in school.

First, a brief description of the LA's BEST program is provided.

[4] a comprehensive after school program that fosters resiliency and school success for at-risk youth

LA's BEST-The Program

LA's BEST was first implemented in the fall of 1988. The program is under the auspices of the Mayor of Los Angeles, the Superintendent of the Los Angeles Unified School District (LAUSD), a board of directors, and an advisory board consisting of leaders from business, labor, government, education, and the community.

LA's BEST seeks to provide a safe haven for at-risk students in neighborhoods where gang violence, drugs, and other types of anti-social behaviors are common. The program is housed at selected LAUSD elementary schools and is designed for students in kindergarten through fifth/sixth grade. The LA's BEST sites are chosen based on certain criteria, such as low academic performance and their location in low-income, high-crime neighborhoods. For optimal program success and to ensure buy-in from the principals and the school staff, the school principals have to officially write a letter of request for the program to be placed in their school site.

LA's BEST is a free program open to all students in the selected sites on a first come first serve basis. Students who sign up for the program are expected to attend five days a week in order to reap the full benefits of the program offerings. Currently, LA's BEST serves a student population of approximately 30,000 with about 80% Hispanic and about 12% Black elementary students. English language learners comprise of at least half of the student population from most sites. Of this population, the majority's primary language is Spanish; while the other percentage of the English learner population is composed of those who's first language is of Asian/Pacific origin.

Parents often mention homework help and proper supervision as the primary incentives for enrolling their children to the program. Students are also recommended by teachers to attend LA's BEST due to behavioral or academic needs. Students enjoy the program due to its supportive staff and positive environment conducive for academic achievement and engagement of extracurricular activities. A brief history and an organization chart of LA's BEST are available in Appendix E.

Program Offerings

Since its inception in 1988 LA's BEST has adapted and updated their goals in response to educational policies, research, and theory. Over the years, the program has moved past its initial emphasis on providing a safe environment and educational enrichment to an emphasis on the development of the whole child. In developmental theory, a whole child curriculum is one that cultivates the development of students' intellectual, social, and emotional well-beings so that they can achieve their full potential (Schaps, 2006; Hodgkinson, 2006). At LA's BEST, it provides a whole child education by centering on their 3 & 1/2 beats to enhance the students' intellectual, social-emotional, and physical developments:

Cognitive beat & Homework beat

Intellectual development such as:

- responsibility and positive work habits- through emphasis on the importance of completing assignments, teaching learning strategies and study skills, and providing a learning climate that enforces positive attitudes towards school
- love of learning- through active participation, explorations, and engaging research-based activities
- self-efficacy- through guided experiences, challenging activities, and relationship building between staff and students
- future aspirations-through high expectations, activities that build self-reliance, value of education, collaborations, and critical thinking

Recreational beat

Physical, Social & emotional development such as:

- sense of safety & security- through providing students with a safe and nurturing environment.
- healthy life style-through curriculum and activities that promotes drug and gang prevention, healthy eating habits, and plenty of exercises.
- social competence- through demonstrating and enhancing students' respect for self and others, and providing students with opportunities to form friendships and develop trust and respect with peers and adults.
- sense of community- through providing students with opportunities to participate in community sponsored events, volunteer in community assignments, and offering field trips to local business and organizations.
- respect for diversity- through role modeling and curriculum that enhances awareness and responsibility to each other within their diverse community.

To summarize, the mission of LA's BEST is to provide engaging settings so that each student learns in an intellectually challenging environment that is physically and emotionally safe for both students and adults; furthermore, each student can be actively engaged in learning activities that is connected to their school and broader community; and most importantly, each student also has the access to extra-curricular activities, academic enhancements, and to qualified, caring adults.

Since the central theme of the LA's BEST mission is to empower both staff and student members, and to build on students' daily life experiences with program offerings; Thus as a program policy, each individual LA's BEST sites may be autonomous in how they structure their specific programs as long as the site coordinators and staff adhere to the foundational principles of the program.[5] As a result, each site has its distinct characteristics and program themes (such as arts, self-esteem, conflict resolution, technology, etc). Subsequently, relationships with the day school, and levels of school[6] and community supports also tend to vary with each site (see Huang, Miyoshi, La Torre, Marshall, Pérez & Peterson, 2006).

The following list provides an overview of the different educational and enrichment activities offered:

Cognitive/Academic

This includes homework time, tutoring, academic incentive programs, math and science activities, reading and writing activities, computer activities, and psychological programs addressing conflict resolution skills.

Recreational

This includes arts and crafts, cooking, games, holiday activities, and sports such as aerobics, karate, and team sports.

[5] The snack and homework periods are the common components of all LA's BEST. The education and enrichment sessions are grounded on the principles of being: 1) cognitive/ academic(activities in school subject matter; 2) recreational (physical fitness); 3) part of the performing arts (i.e. dance, drama, etc.).

[6] In a qualitative study of six LA's BEST sites, Huang and colleagues (2006) found that most principals had a cooperative working relationship with LA's BEST site staff.

Performing and Visual Arts

This includes choir and music, dance, drama/theater, flag/drill team, museum visits, art camps, etc.

Health and Nutrition

This includes study of nutrition, healthy habits, and exercises programs such as tennis, skating, and BEST Fit community health fair.

Community and Cultural

This includes community programs, such as adopt-a-grandparent, and community days; and cultural programs, such as those dedicated to Black history, "Folklorico," and other cultural holiday celebrations

Parental Involvement Activities

These activities include:

- celebrations, for example: Halloween Kidfest, Community Jam, and Awards Days;
- programs for children, for example: parents' volunteering for daily activities and field trip supervision;
- programs for parents, for example: parent workshops and parent education speakers;
- communication/information, for example: open house events, assemblies, and parent-teacher meetings; and
- field trips, for example: a variety of field trips to performing arts events and visits from artists to LA's BEST sites.

The above activities mainly come from three different sources: 1) curricula purchased from education vendors, such as KidzLit[7] and KidzMath[8]; 2) activities

[7] Afterschool KidzLit is an enrichment program that emphasizes literacy skills, written expression, core values, connections, and thinking skills by having children read and talk about books. The program is research based and is aligned with the National Council of Teachers of English (NCTE) standards.
[8] Afterschool KidzMath is an enrichment program that emphasizes the enjoyment and development of math skills. Lessons are structured around the use of math games and math-themed children's books.

developed by the education and staff development departments at LA's BEST operations; and, 3) activities designed by the site staff[9].

Quality Assurance

For continuous improvements LA's BEST employs both internal and external evaluators. Their operations office includes both a Director of evaluation and research analysts. The internal evaluation team conducts regular meetings with field staff to provide a forum for sharing experiences and examples of what works and what does not work with staff and administrators at the operation office. External evaluations often involve feedbacks from staff, school teachers, students and parents; it gauges the short and/or long term effects on specific program components, or overall program effects.

Results from evaluations are discussed at site coordinator meetings, and are used to determine whether individual sites and the program are meeting goals and objectives.

Study Design and Methods

Since the formation of LA's BEST in 1988, CRESST[10] has been conducting evaluations of the program. CRESST has established a longitudinal database on these students as well as a longitudinal database on a comparison group of control students. This longitudinal database[11] includes student demographics and academic information such as student achievement scores on Reading and Mathematics standardized tests.

The basis for the sample is comprised of the LAUSD student dataset that CRESST has collected and stored since the 1992-93 school year. The first step in building a sample consists of generating a sampling frame. We accomplish this task by going back through the historical records and tracking all available information for all students from the 1994-95 school year through the 2002-2003 school year.

---

[9] Site staff members receive support from the activities consultants and their site coordinators in developing and/or implementing activities.

[10] CRESST- The National Center for the Research in Educational Standards and Student Testings in UCLA.

[11] The school information is partially obtained from the National Center for Education Statistics and Los Angeles School Police data. The latter is a longitudinal database between the years 1995 and 2002. Furthermore, we also utilized 1990 census data to characterize the neighborhoods of the treatment and control schools.

We also consider the changes in schools and communities in which LA's BEST and comparison schools are located. The analyses of demographic changes over the past 10 years in LAUSD include comparisons of school and community characteristics. This allows us to account for potential contextual confounding factors, and to consider how these factors have changed over time. The following describes the study design and the data analyses strategies for this study. We describe the data sources, variables, and sample in detail in the following section.

Study Design

This study employs a quasi-experimental design that consists of longitudinal sample of both academic and juvenile crime data. The sample is comprised of 2,331 students from LA's BEST programs, 2,331 students who attend the same schools as those in the LA's BEST programs but do not participate in LA's BEST, and 1,914 who attend schools that had no LA's BEST program. The base years for these students were in 1994-95, 1995-96, and 1996-1997. We took advantage of this panel structure and applied hierarchical growth modeling to academic outcomes as well as hierarchical survival analysis to crime outcomes. These methods allowed us to examine students' academic growth and likelihood of particular events. In both instances, we compared LA's BEST students against non-LA's BEST students. Given that we had student background information, we examined moderating factors such as gender, race/ethnicity, language proficiency, and SES. We also examined potential programmatic mediating factors. Further, we used available cost information to derive benefit-cost ratios.

Data Analysis Methods

The importance and advantages of using multilevel analyses in program evaluations have been discussed in Seltzer (2004) and Raudenbush and Bryk (2002) for cross-sectional designs, and Osgood and Smith (1995) and Singer and Willet (2003) for longitudinal studies. The important aspect that we consider is that students are clustered within schools, and those students do not represent independent observations (Raudenbush & Bryk, 2002). This clustering leads to under-estimation of standard errors (Raudenbush & Bryk, 20002) and errors in interpretation when analysis examines multiple levels of data (Burstein, 1980). To counter this aspect, we utilized both the growth and survival models within the general framework of hierarchical (random coefficient) models. This allowed us to explicitly handle multiple levels of data efficiently (Raudenbush & Bryk, 2002).

We also utilized the longitudinal nature of the data and followed academic and social student development over time. The benefit of this longitudinal structure is twofold. First, it allows us to move beyond traditional pre/post analysis, which is limited by data requirements and explanatory possibilities (Rogosa, Brandt, & Zimowski, 1982; Raudenbush & Bryk, 2002). We employed growth-modeling techniques that examined individual trajectories (Rogosa et al., 1982) and have more flexible data requirements.[12] Second, we separated initial status from growth, thus avoiding spurious negative correlations between where students start and their ensuing growth (Bloomquist, 1977).

To sample comparable control schools and students, we estimated the propensity score. We then used these scores to select units from a large reservoir of potential controls applying a systematic matching procedure. The propensity score is the conditional probability of being assigned to the treatment condition given a set of observed covariates. It is commonly estimated using a logistic link function.

In order to examine the effects of LA's BEST on achievement and achievement growth, we employed a hierarchical linear model (HLM) design that has the advantages of directly modeling growth trajectories and being more flexible than traditional analyses. Since observations are nested within individuals, time intervals need not be constant nor the same across individuals as in traditional repeated measures analyses (Raudenbush & Bryk, 2002), and the number of observations per person may vary. HLM allows flexible specification of the covariance structure at every level of the analysis (Snijders & Bosker, 1999).

The HLM analysis is based on a three-level model. At level 1, student development is represented by a growth trajectory dependent upon a set of parameters. The outcomes are nested within students. At level 2, these individual growth parameters become outcomes that depend upon student-level characteristics; and at level 3, student characteristic effects become outcomes dependent on school and neighborhood-level characteristics.

It is important to utilize a three-level model in this context because we were particularly interested in examining the effect of LA'S BEST on student achievement

[12] such as not requiring balanced data (Raudenbush & Bryk, 2002) and managing missing data due to attrition (Hox, 2002)

status and growth. By using a three-level model, we were able to divide the variation in achievement into between-student, between-school, and error components. This is important to do because data containing multiple levels of aggregation can lead to errors in interpretation of results when these multiple levels are ignored (Aitkin & Longford, 1986; Burstein, 1980). For example, socioeconomic status, measured at the student level, represents a measure of a student's home resources; while the aggregated mean student SES at a school measures the average resources available in a community (Burstein, 1980). Ignoring the nested nature of the data and simply analyzing outcomes aggregated to the teacher or school level upwardly biases results of student-level predictors because within school student-level variation is lost upon aggregation (Freedman, Pisani, & Purves, 1978). Not only are the student effects biased, but it also becomes unclear whether the estimated effect is due to a group effect or whether the aggregated variable proxies for an unrepresented student effect (Burstein, 1980).

The three-level model is constructed in the following manner.

In general at level 1:

$$Y_{tij} = \pi_{0ij} + \pi_{1ij}\alpha_{tij} + \pi_{2}C_{tij} + e_{tij}, \quad (1)$$

where Y_{tij} is the outcome at time t for student i with school j, α is the time parameter measured in years, and C is the time varying covariate, such ELL status during any year. Our specification includes additional level 1 parameters because preliminary plots of the data revealed significant non-linearity. We also include additional fixed test-effects that account for changes in assessments over time. We detailed the parameterization in the results section that follows. Since growth trajectories are assumed to vary across students, at level 2 for the status[13] at time = 0:

$$\pi_{0ij} = \beta_{00j} + \beta_{01j}X_{1ij} + \ldots + \beta_{0Pj}X_{Pij} + r_{0ij}, \quad (2)$$

where there are p = 1 to P student-level predictors. For the growth trajectories:

$$\pi_{1ij} = \beta_{10j} + \beta_{11j}X_{1ij} + \ldots + \beta_{1Pj}X_{Pij} + r_{1ij}, \quad (3)$$

and for the time varying covariate:

$$\pi_{2ij} = \beta_{20j} + \beta_{21j}X_{2ij} + \ldots + \beta_{2Pj}X_{Pij} + r_{2ij}, \quad (4)$$

[13] Status is achievement in 1998.

The time varying covariate is also free to vary among students. The effects of student characteristics are assumed to vary across schools at level 3. For example, for initial status:

$$\beta_{00j} = \gamma_{000} + \gamma_{001}Z_{1j} + \ldots + \gamma_{00Q}Z_{Qj} + u_{00j}, \quad (5)$$

where there are q = 1 to Q school-level predictors. For the status student-level effects:

$$\beta_{0pj} = \gamma_{0p0} + \gamma_{0p1}Z_{1j} + \ldots + \gamma_{0pQ}Z_{Qj} + u_{0pj}, \quad (6)$$

For the mean student growth trajectory:

$$\beta_{10j} = \gamma_{100} + \gamma_{1pq}Z_{1j} + \ldots + \gamma_{1pQ}Z_{Qj} + u_{10j}, \quad (7)$$

and for each student effect:

$$\beta_{1pj} = \gamma_{1p0} + \gamma_{1pq}Z_{1j} + \ldots + \gamma_{1pQ}Z_{Qj} + u_{1pj}, \quad (8)$$

Finally, the time varying covariate is allowed to vary randomly at level 3 as well.

$$\beta_{20j} = \gamma_{200} + \gamma_{2pq}Z_{2j} + \ldots + \gamma_{2pQ}Z_{Qj} + u_{20j}, \quad (9)$$

In other words, we estimate an average status and an average growth trajectory. These estimates are allowed to vary among students, and this variation among students was modeled by various student-level predictors (e.g., gender). The mean status and the mean growth trajectory, as well as the student-level predictors, are allowed to vary among school. Hence, each of these is modeled as a function of school-level predictors.

In general, we employ the following steps in building a prudent model that explained student growth and the variation in growth among students. Equations 1 through 8 are combined to build a three-level growth model, which describes Reading and Mathematics achievement trajectories for each student and how these growth trajectories varied between students and schools. The first step is to use an unconditional model (a model with only a growth parameter, but no other predictors) to examine various growth trajectories and provide baseline statistics to evaluate various level 2 and level 3 models. This also provides an estimate of the mean intercept and an estimate of the mean growth trajectory. Additionally, the unconditional model determines whether these estimates are significant and whether they vary significantly between students and schools. Further, this provides an estimate of the true correlation between the initial status and the growth rate. Normal pre/post designs generally provide spurious negative correlations because the error variance of the pretest is negatively correlated with growth (Bloomquist, 1977).

The next stage in the analysis is to expand the unconditional model one level at a time (from level 1 to level 3) using the available student and school information. Specifically, the model tests for the impact of LA's BEST programs by using an LA's BEST indicator or indicators[14]. This variable is included in the student-level model as each student is either participating in LA's BEST after-school programs or not. Hence, the simplest treatment contrast is to code the LA's BEST indicator variable as 1 if the student attended LA's BEST, or as 0 if the student was part of the matched control. The model for the growth trajectory (equation 3) is expanded to include a parameter for the effect of LA's BEST.[15]

$$\pi_{1ij} = \beta_{10j} + \beta_{11j}X_{1ij} + \ldots + \beta_{1Pj}X_{Pij} + \lambda_{11j}\text{LA's BEST}_{1ij} + r_{1ij}\,, \qquad (3b)$$

Using the λs in both the intercept and the slope models yields estimates that identify whether there are differences between treatment and non-treatment groups in status (at the beginning of middle school); contrasts the treatment versus the non-treatment group in growth trajectories; and estimates an average treatment effect over the span of the data (Osgood & Smith, 1995).

The rationale behind the methodology we use to examine crime data is the same as for the analysis of achievement data. Confounding issues arise because the outcome of interest – crime - is an event thereby requiring changes to the models applied, as well as interpretations of the parameters associated with time. The distinction necessitates a shift from growth trajectories to hazard functions. In order to accomplish this while explicitly accounting for the clustering of students within schools, we utilize a Multilevel Discrete-Time Hazard (MDTH) analysis that has been implemented by several previous analyses in the literature (Barber, Murphy, Axinn, & Maples, 2000; Reardon, Brennan, & Buka, 2002; Callens & Croux, 2005) and has been described in Goldstein (1999). However, practical application of MDTH models in educational evaluation settings has been limited.

Much of the recent emphasis in education has been student assessment results. However, in terms of impact on continued education and postsecondary education outcomes, other events contribute significantly. For example, dropping out of school

---

[14] We examined various specifications that best capture potential treatment effects. These will be presented along with the results discussed below.

[15] Equations 2 and 4 are expanded in the same way.

significantly impacts a student's earning potential (Rumberger & Lamb, 2003; Goldschmidt & Wang, 1999). Another example is youth crime, which if prevented can benefit society tremendously (Cohen, 1998). Given that the interest lies in whether and when the event occurs, survival analysis should be utilized (Singer & Willett, 2003). MDTH analysis is appropriate for estimating survival functions, comparing those functions among groups (e.g., treatment and control) and examining the relationship of explanatory variables to survival time (Kleinbaum, 1996). A key consideration for the current study is that complete data is not available for all subjects, and by the end of the study, subjects have either attenuated or the event occurred (Willett, Singer, & Martin, 1998; Singer & Willett, 2003). In this case, we focus on the first arrest of juveniles as part of the evaluation of LA's BEST's effectiveness.

Longitudinal and Discrete-Time Hazard models can account for time-varying covariates (Singer & Willett, 2003) as can MDTH at both the individual and unit or school levels (Barber et al., 2000). A time varying covariate at the unit level represents the potential that changes in context over time and mediates the relationship between an individual level explanatory variable and the outcome of interest. Previous research has examined situations in which unit level context changes (Barber et al., 2000).

The key to using the MDTH model is properly setting up the datasets prior to analysis. This is described in detail by Singer and Willett (2003) and Barber et al. (2000). We followed the design outlined by this previous work. The basic MDTH model then takes the form of:

LEVEL 1 MODEL (bold: group-mean centering; bold italic: grand-m

$$\text{Prob}(\text{CRIME3}=1|\beta) = \varphi$$

$$\text{Log}[\varphi/(1 - \varphi)] = \eta$$

$$\eta = \beta_0 + \beta_1(\text{YEAR}) + \beta_2(\text{YEARSQR})$$

LEVEL 2 MODEL (bold italic: grand-mean centering)

$$\beta_0 = \gamma_{00} + u_0$$

$$\beta_1 = \gamma_{10} + u_1$$

$$\beta_2 = \gamma_{20} \qquad (11)$$

Here we use the natural log likelihood function to estimate parameters of interest (Willett & Singer, 2003). In this case the basic specification include two terms to track time: year and year squared. This allows us to model a non-linear hazard function. We explore the fitting of this model to the actual hazard in the results section.

As with achievement modeling, we specify the effect of both intercept and time as being random or varying across schools. Hence, the level 2 model allows us to examine whether there is significant variation among schools in the hazard function.

The final parameterized model includes both student and school level covariates and is specified as:

LEVEL 1 MODEL (bold: group-mean centering; bold italic: grand-m

Prob(CRIME3=1|β) = φ

Log[φ/(1 - φ)] = η

$\eta = \beta_0 + \beta_1$(YEAR) + β_2(YEARSQR) + β_3(TREATMEN) + β_4(LABATMED)
β_5(LABATHI) + β_6(FEMALE) + β_7(HISPANIC) + β_8(BLACK) + β_9(ASI/
β_{10}(EVERDSP) + β_{11}(PEDUHI) + β_{12}(DURAT2) + β_{13}(LEP_SUM)

LEVEL 2 MODEL (bold italic: grand-mean centering)

$\beta_0 = \gamma_{00} + \gamma_{01}$(***BLACK_PG***) + γ_{02}(***PEDUHI_P***) + γ_{03}(LABEST_F) + γ_{04}(l
+ γ_{05}(***POVERTYP***)

$\beta_1 = \gamma_{10} + \gamma_{11}$(***BLACK_PG***) + γ_{12}(***PEDUHI_P***) + γ_{13}(LATERLB) +
γ_{14}(***POVERTYP***) + u_1

$\beta_2 = \gamma_{20}$

$\beta_3 = \gamma_{30}$

$\beta_4 = \gamma_{40}$

$\beta_5 = \gamma_{50}$

$\beta_6 = \gamma_{60}$

$\beta_7 = \gamma_{70}$

$\beta_8 = \gamma_{80}$

$\beta_9 = \gamma_{90}$

$\beta_{10} = \gamma_{100}$

$\beta_{11} = \gamma_{110}$

$\beta \quad = \gamma$ (12)

We discuss the details of this specification in the results section. Specifically, we present the particulars of coding and the rationale behind our choices.

Building the Dataset

In this section we briefly describe the data sources utilized in building the dataset. Building the dataset requires not only collecting data to analyze but first sampling students to include in the analyses. Given the quasi-experimental design, we have to carefully select a control group to serve as a counterfactual group. We detail the sampling procedure in the following section. We first present the list of variables coded at the student and school levels and used for the sampling and statistical analyses.

The data set we present below is unique for several reasons: one, the sample size; two, the contiguous nature of data elements; three, the number of years of data; four, the objective nature of many data elements; and five, the careful selection mechanisms used to sample counterfactuals. Studies examining after-school program effects are generally limited in sample size, focusing on specific sites. Our sample contains approximately 6,000 students and 48 schools. Studies focusing on long—term effects are often forced to use retrospective data based on student perceptions rather than extant data collected in real time of a 10 year period – which includes achievement data from a district that tested students annually in grades 3-11.[16] Because the data elements have been collected of the past 10 years and the criminal offenses are based on actual probation data files, neither educational histories nor criminal behavior is based on survey responses. Finally, the matching scheme, while a second best option to randomization, we present below matches each individual student based on a large set of characteristics (including test scores) and explicitly incorporates the nested nature of students within schools.

[16] Pre-NCLB many states and districts only tested students in one elementary, one middle, and one high school grade.

Table 1 Description of the Variables Tested in the Subsequent Models

	Name	Variable	Type
Students' time-variant covariates			
1	Age	Age	Continuous
	Grade Level	Grade	Ordinal
	Language proficiency: EO	EO	Dummy – reference LEP
	Language proficiency: RFEP	RFEP	Dummy – reference LEP
	In travel program or not	Travel	Dummy
	Student residence equal to school location or not	Dresident	Dummy
	Reading Total (rdptl4)	Readnce	NCE Scale
	Mathematics Total (mtptl4)	Mathnce	NCE Scale
Students' time-invariant covariates			
	Cohort 2	Cohort2	Dummy – reference cohort 3
	Female	Gender	Dummy reference –Male
	Ethnicity: Hispanic	Hispanic	Dummy – reference white
	Ethnicity: African American	Black	Dummy – reference white
	Ethnicity: Asian	Asian	Dummy – reference white
	Ethnicity: Other	Other	Dummy – reference white
	Home Language: Spanish	Hlspanish	Dummy – reference English
	Home Language: Asian	Hlasian	Dummy – reference English
	Home Language: Other	Hlother	Dummy – reference English
	Years in lunch program	yearslunch	Dummy reference-full paid or non- participant
	Parent education: complete college or more	Peduhi	Dummy – reference some college or less education
	Parent education: some college	Peduhi2	Dummy – reference complete high school education or less
	Ever in special education or not	everdsp	Dummy
	Ever gifted student or not	Evergate	Dummy
	Ever retained in elementary	Retain1	Dummy – reference never retained in elementary
	Ever retained in high school	Retain2	Dummy – reference never retained in high school
	In Track A or not	Tracka	Dummy
	In track B or not	Trackb	Dummy
	In track C or not	Trackc	Dummy
	In track D or not	Trackd	Dummy
	No track	Notrack	Dummy
Treatment Indicators			
	In the program or not	Labest	Dummy
	Total days attended	intensity	Continuous
	Total years attended	duration	Ordinal

	Name	Variable	Type
	Average days of attendance per year	Engagement	Continuous
	Indicator of intensity: high	labathigh	Dummy
	Indicator of intensity: Medium	Labatmed	Dummy
	Total days attended 1993	Days93	Continuous
	Total days attended 1994	Days94	Continuous
	Total days attended 1995	Days95	Continuous
	Total days attended 1996	Days96	Continuous
	Total days attended 1997	Days97	Continuous
School Level variables & zipcode indicators			
	Proxy of Implementation: Monthly number of volunteer hours	movoluho	Continuous
	Number workshops attended by LA's BEST staff	nworkshops	Continuous
	Started in program in 1988	Start88	Dummy
	Started in program in 1989	Start89	Dummy
	Started in program in 1990	Start90	Dummy
	Started in program in 1994	Start94	Dummy
	Started in program in 1995	Start95	Dummy
	Started in program in 1996	Start96	Dummy
	Pupil teacher ratio	Puptch97	Continuous
	% white in zipcode	Pcwhite	Continuous
	% African American in zipcode	Pcblack	Continuous
	% Asian in zipcode	Pcasian	Continuous
	% native American in zipcode	Pcnative	Continuous
	% other in zipcode	Pcother	Continuous
	% households with 7 or more people in zipcode	Pchouse7	Continuous
	% households with income less than 10,000 in zipcode	Pchi10	Continuous
	% households with income less than 15,000 in zipcode	Pchi14	Continuous
	% population 25 years old over with less than 9 grade in zipcode	Pcedu9	Continuous
	% population 25 years old over with 9 to 12th grade no diploma in zipcode	Pcedu912	Continuous
	% population 25 years old over HS graduate in zipcode	pceduhs	Continuous
	Crime trend between census years	Tcrime	Crime trend by zipcode

Data Sources

As mentioned previously, LA's BEST[17] serves a population of low-income, low-performing, and predominately Hispanic (about 75%) and Black (about 12%) elementary students in the Los Angeles Unified School District (LAUSD). Since the inception of LA's BEST in 1988, CRESST has been conducting evaluations of the program. CRESST has established a longitudinal database on these students as well as a longitudinal database on a comparison group of control students. This longitudinal database includes student demographics and academic information such as student achievement scores on Reading and Mathematics standardized tests. We combined this database with school level information partially obtained from the National Center for Education Statistics and Los Angeles School Police data. The latter is a longitudinal database between the years 1995 and 2002. Furthermore, we also utilized 1990 census data to characterize the neighborhoods of the treatment and control schools.

The basis for the sample is comprised of the LAUSD student dataset that CRESST has collected and stored since the 1992-93 school year. The first step in building a sample consists of generating a sampling frame. We accomplish this task by going back through the historical records and tracking all available information for all students from the 1994-95 school year through the 2002-2003 school year.

We also consider the changes in schools and communities in which LA's BEST and comparison schools are located. The analysis of demographic changes over the past 10 years in LAUSD includes comparisons of school and community characteristics. This allows us to account for potential contextual confounding factors, and to consider how these factors have changed over time.

Selecting the Treatment Students

It is very important to establish a sample that carefully matches students who attended LA's BEST with those that did not attend LA's BEST so that valid inferences can be generated. While we have the advantage of utilizing longitudinal data in which students can be their own controls, we must also generate comparison samples in order to make further generalizations about programmatic effects within a quasi-experimental

[17] Currently serving students in more than 178 Los Angeles Unified elementary school sites, its aim is to combine academic and recreation activities that focuses on improving the life choices of at-risk students.

setting. Hence, without random assignment we need to carefully consider the likelihood of unobservable factors that explain both LA's BEST membership and subsequent outcomes. Thus, in order to reduce biases from potential confounding factors we take the following steps to analyze and construct the sample.

First, elementary student participation in LA's BEST is measured as the number of days attended during the academic year. The maximum number of days for program attendance in each school site that operates for 9 months is 180 days and 240 days for year-round schools. Examination of student attendance patterns indicates that many students participate sparingly and then drop out of the program. Therefore, it is necessary to set a criterion for minimum attendance to distinguish treated students against those with very low or casual attendance, thereby refining the analysis and reducing the intent to treat phenomenon, to some extent, at the outset[18]. Thus, students who attend the program less than four times per month are considered as untreated students. In contrast, those students that attend at least 36 days (i.e., at least once a week during 9 months) are included in the sample as part of the treated sample of students[19]. The following table shows the number of students with attendance information including the mean, the standard deviation, and the minimum and maximum attendance values.

[18] However, subsequent analyses used the number of days attended as a covariate to adjust for variable exposure experienced by even consistent enrollees.

[19] For the year-around schools, the cut-point of 36 did not satisfy the minimum criterion but with the goal of using a standard criterion, the minimum cut-point was maintained for these schools.

Table 2

Descriptive Statistics of Attendance in the LA's BEST After-School Program

Attendance	N	Mean	SD	Min	Max
1991	462	32.8	32.7	0	154
1992	282	12.9	5.7	0	20
1993	4,364	29.7	32.0	0	205
1994	7,109	62.9	48.1	0	203
1995	8,438	75.5	58.4	0	240
1996	9,028	76.6	58.4	0	240
1997	7,338	67.6	46.8	0	195
1998	---	---	---	---	---
1999	---	---	---	---	---
2000	20,451	83.0	61.9	0	240
2001	25,440	90.1	65.7	0	240
2002	32,478	118.1	63.2	1	240

Note. Attendance for each of the years was corrected from outliers.

In the following table we present the number of students that satisfy three different criteria of attendance for the 11-year span.

Table 3

Number of Students with LA's BEST Attendance Information for the 1993-2002 School Years

N	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
With attend>0	462	282	4,318	7,084	8,409	8,852	7,289	---	---	20,434	24,222	32,478
With attend =>24	129	0	1,622	5,121	6,215	6,728	5,503	---	---	13,751	19,504	28,013
With attend =>36	103	0	1,072	4,380	5,525	6,031	4,892	---	---	12,290	17,876	26,553

Note. Sample restricted to schools that have more than 20 students with attendance information.

It is important to note that the number of students with attendance information is very low for 1991 and 1992; in fact, there are no students that satisfy the minimum attendance criterion in 1992. This is likely due to the fact that collecting attendance information is not systematic in the early years of the program. In subsequent years after 1992, the number of students meeting the attendance criteria increases. For example, the percentage of students with attendance equal to or greater than 36 is 25% in 1993 and 62% in 1994. For the years 1995, 1996, and 1997, these percentages are 66%, 68% and 67% respectively. However, attendance information for 1998 and 1999 is not available. After 2000, a significant improvement in the attendance rates was observed. In the years 2000, 2001, and 2002 the number of students with attendance higher than zero increases to 20,434, 24,222, and 32,478 respectively.

Table 4 shows the number of schools for each of the years that satisfy the three different attendance criteria. Given the scarcity of information for the years of 1991, 1992, 1998 and 1999, the data is not reported.

Table 4

Number of Schools that Offered the LA's BEST Program during 1993-2002 School Years

Schools	1993	1994	1995	1996	1997	2000	2001	2002
N with attend data >0	18	22	22	25	22	52	77	105
N with attend >36	10/9	22	22	25/24	22/21	50	77	104
N with attend >24	14/13	22	22	25/24	22/21	50	77	104

Table 4 presents two numbers for some years. The first number represents the total number of schools that includes students who fulfill the minimum attendance requirements. However, very few students satisfy the criterion in some of these schools. In order to distinguish these schools a second number is reported. For example, during 1993 a total of 10 schools have students that satisfy the 36-day attendance criterion. However, only nine of these schools have a meaningful number of students who actually attend the program.

Since data for 1991, 1992, 1998 and 1999 are either unreliable or unavailable, we restrict the years from which to sample students to the period between 1994 and 1996. However, we also checked if the sampled students attended the program during 1993 and 1997. Table 5 presents the distribution of students with attendance equal to or greater than 36 days by school.

Table 5

Number of Students at Schools with Attendance > 36 For Years 93-97

School	N of students				
	1993	1994	1995	1996	1997
1	39	164	369	377	319
2	52	389	408	361	287
3	8	124	175	183	182
4	124	169	174	251	189
5	162	105	247	149	234
6	44	228	296	249	233
7	259	273	299	340	294
8	240	226	175	296	247
9	82	220	220	233	185
10	3	124	103	126	163
11		71	83	90	105
12		116	267	255	233
13		236	283	241	314
14		324	420	441	418
15		110	180	203	209
16		207	210	190	180
17		175	235	208	201
18		183	256	234	194
19		328	383	378	317
20				17	27
21				234	156
22				253	205
23		254	267	232	
24		134	167	183	
25	59	220	308	307	
Total	1,072	4,380	5,525	6,031	4,892

During this period, 10 of the 25 schools reported students with attendance higher than 36 days for five years. Nine of these 25 schools were started in 1994, another 3 schools were started in 1996.

Several critical pieces of information are used to build the sampling frame. As noted, we excluded the years 1991, 1992, 1998 and 1999 due to poor data quality. Another consideration for sampling is the number of years students participated in the program. Given that students can participate in the program from first through fifth grade, we need to be able to follow students for five years in order to gain an accurate picture of attendance. Due to retention, the number of grades we need to observe vary with each student.

Table 6 presents the possible cohorts that could be included in the study and the number of years that data are available. Four different cohorts are displayed. In addition, Table 5 also presents the grade level of the student.

Table 6

Sampling Scheme by Cohort, Year, and Grade

				Sampling Years				Follow-up Years				
Cohort-Grade	91-92*	92-93*	93-94	94-95	95-96	96-97	97-98	98-99**	99-00**	00-01	01-02	02-03
I				Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9
II			Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9	Grade 10
III		Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9	Grade 10	Grade 11
IV	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9	Grade 10	Grade 11	Grade 12

*There was no demographic information for this year. **There was no attendance information for this year.

Students belonging to cohort I began the after-school program in first grade during the 1994-1995 school year. Given that a student can participate in the program until fifth grade, attendance information is needed until 1998 to verify the number of years of participation. Since attendance data is not available in 1998, we are unable to verify attendance status for this year. Hence, we have to either impute attendance or limit the analysis to contain a maximum of four years of attendance for this cohort. The latter approach likely results in over-estimated treatment effects for this cohort[20].

Students in cohort II began the program in first grade during the 1993-1994 school year. In order to obtain complete exposure information, we check attendance information through 1997. Therefore, we are able to identify the status of the student (i.e., participant vs. non-participant) for all the necessary years for this cohort. For this reason, students from cohort II are selected to participate in this study.

In the case of cohorts III and IV, we need attendance information for 1991 and 1992, which as previously noted, is less reliable and scarce in 1992. However, since these are the beginning years for the program, it may be assumed that the attendance rate is relatively low. Hence, if we assume that the students in any of these cohorts do not participate in the program during either of the two early years, we are not likely to affect estimates significantly[21]. Following this rationale, we include cohort III in the sample.

Additionally, there are other necessary considerations in defining the potential sample of students for the study. A key consideration is to select only students that attend the program in the same schools. This consideration is important in order to avoid cross-classification problems since the quality of implementation likely varies from school to school. In other words, we consider only students who attend LA's BEST at a single school. Further, the students are sampled to be in high school at the beginning of the 2000-2001 year.

Characteristics of the Sampled Cohorts. The following two tables present the sample characteristics of cohorts II and III. The numbers underlined represent the number of students for each year counted in the sample. For example in 1994, a total of 753 students received the treatment in 2nd grade (679 students started in 1994). From these students, 147 remained in the program until 5th grade in 1997. In 1995, 502 new students entered the program in 3rd grade, and again some of them remained in the program during the following years. In 1996, 445 students started the program in 4th grade, and 179 continued until 1997. Finally in 1997, 225 students started the program in 5th grade. The total sample of unique students in those five years was 1,917. Restricting

[20] Analyses using survival analysis techniques were not affected as survival analysis explicitly takes truncated samples into account.
[21] We imputed attendance and conducted sensitivity analysis to determine effects of assumptions.

the sampling years to 1994, 1995, and 1996 resulted in a cohort II of 1,692 students who attended the after-school program in the same school.

In the case of cohort III, the sample of unique students that attended the program in the same school was 1596 students. The total sample of two cohorts of students was 3288.

Table 7

Cohort II: Students in 1993 were in Grade 1 (if attended >36 days)

Cohort ⇒	93: Grade 1	94: Grade 2	95: Grade 3	96: Grade 4	97: Grade 5
Years	**93**	**93&94**	**93&94&95**	**93&94&95&96**	**93&94&95&97**
N	140	74	50	37	21
Years		**94**	**~93&94&95**	**~93&94&95&96**	**~93&94&95&96&97**
N		753 ~93&94: 679	389	233	147
Same school		745/753			
Years			**95**	**~93~94&95&96**	**~93~94&95&96&97**
N			959 ~93~94&95= 502	227	129
Same school			502/502		
Years				**96**	**~93~94~95&96&97**
N				994 ~93~94~95&96= 445	179 1(grade4)
Same school				445/445	
Years					**97**
N					780 ~93~94~95~96&97= 225

Table 8

Cohorts III: Students in 1993 were in Grade 1 (if attended >36 days)

	93: Grade 2	94: Grade 3	95: Grade 4	96: Grade 5
Years	**93**	**93&94**	**93&94&95**	**93&94&95&96**
N	157	73	49	35
Years		**94**	**~93&94&95**	**~93&94&95&96**
N		781 ~93&94: 708	399	270
Same school		744/781		
Years			**95**	**~93~94&95&96**
N			922 453	221 2
Same school			439/453	
Years				**96**
N				996 425
Same school				413/425

Table 9 illustrates attendance in the program as various combinations of years. For example, of the students that started the program in 1994, 235 attended only that year, while 199 attended two years, 129 three years, 161 four years, and finally only 21 students remained in the program for five years. In this table we can observe that once a student started in the program, it was more common to continue the following year rather than stop one year and then continue in a subsequent year.

Table 9

Combination of Years Attending the Program for Cohort II

Students that started in 1994					
	Combination of years				
One Year	94				
N	235				
Two Years	93&94	94&95	94&96	94&97	
N	21	139	21	18	
Three Years	93&94&95	93&94&96	94&95&96	94&95&97	94&96&97
N	13	2	86	15	13
Four Years	93&94&95&96	93&94&96&97	94&95&96&97		
N	16	1	144		
Five Years	93&94&95&96&97				
N	21				
Students that started in 1995			**Students that started in 1996**		
Combination of years			Combination of years		
One Year	95		One Year	96	
N	250		N	265	
Two Years	95&96	95&97	Two Years	96&97	
N	99	24	N	180	
Three Years	95&96&97				
N	129				

The attendance pattern for cohort III is displayed in Table 10.

Table 10

Combination of Years Attending the Program for Cohort III

Students that started in 1994			
	Combination of years		
One Year	94		
N	266		
Two Years	93&94	94&95	94&96
N	22	125	33
Three Years	93&94&95	93&94&96	94&95&96
N	14	2	247
Four Years	93&94&95&96		
N	35		
Students that started in 1995		**Students that started in 1996**	
	Combination of Years	**Combination of**	**Year**
One Year	95	One Year	96
N	228	N	413
Two Years	95&96		
N	211		

Considering that students can attend the LA's BEST program for up to 5 years and for 240 days within each year, both the number of years and days attended are accounted for by measuring the level of individual exposure to the program. The following figures show boxplots of the relationships between duration and intensity and duration and the average attendance per year in the program.

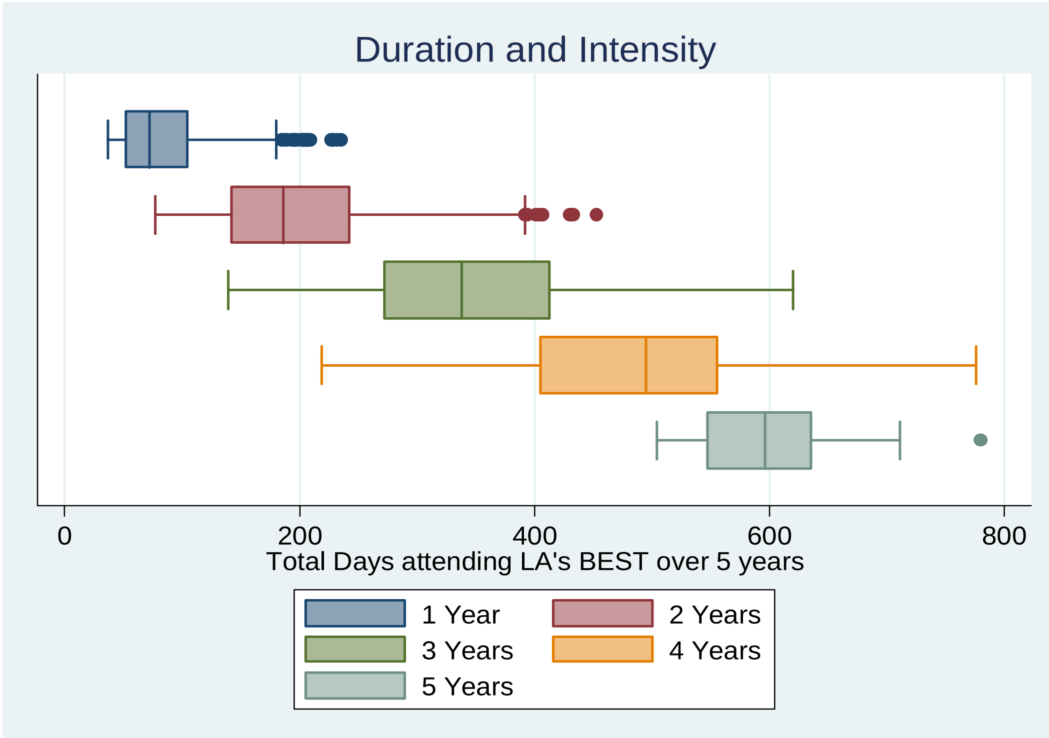


Figure 1. Boxplots of attendance intensity by duration of attendance

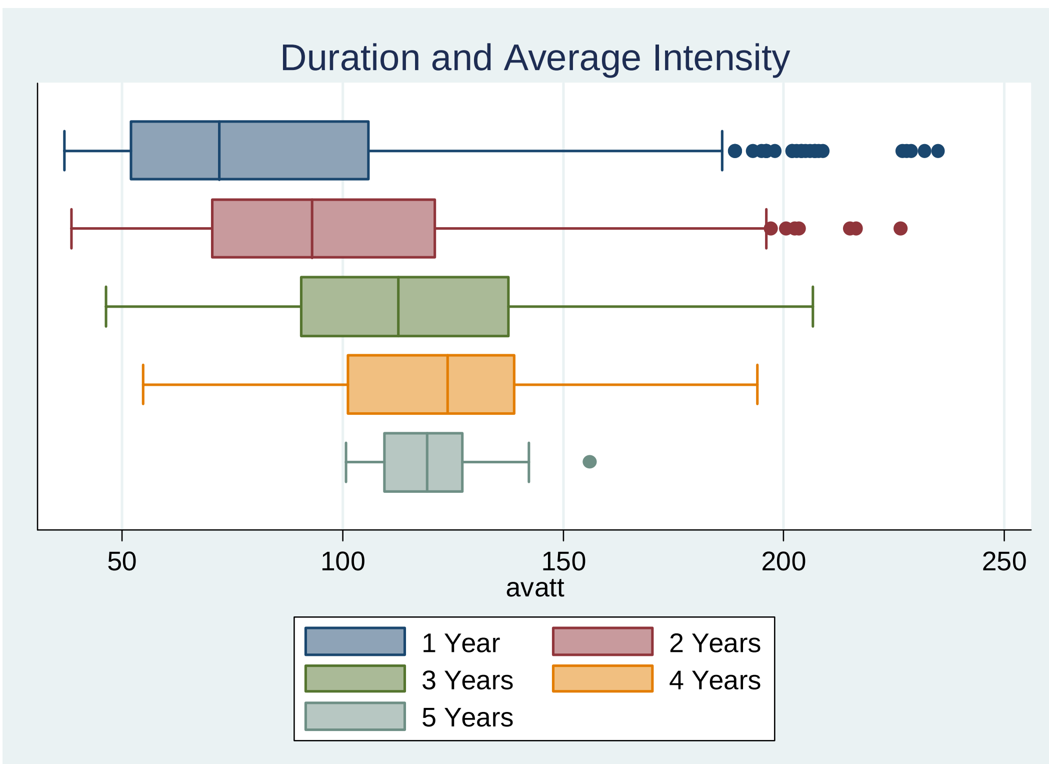


Figure 2. Boxplots of the average intensity of attendance by duration of attendance

Figure 1 demonstrates that overall intensity is highly correlated with exposure. Figure 2 illustrates that average annual intensity is only moderately correlated with exposure. It is important to note that we define exposure as the number of years a student attended LA's BEST; we define intensity as the total number of days a student attended LA's BEST; and, we define engagement as the average number of days per year that a student attended LA's BEST.

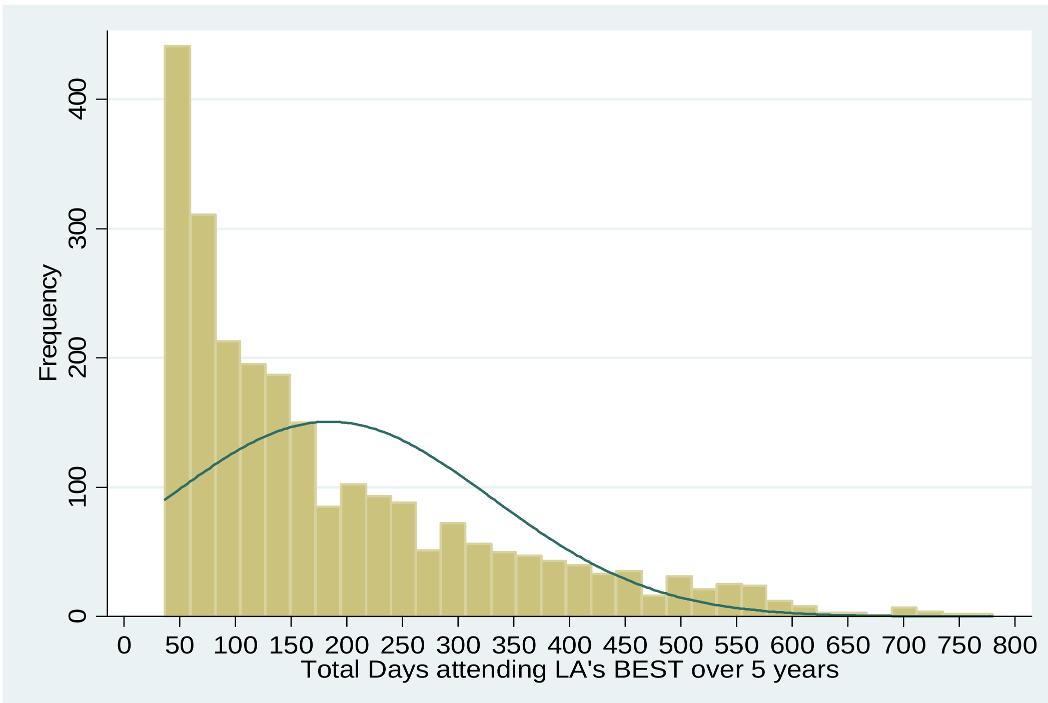


Figure 3. Distribution of days of attendance over a period of 5 years

Note. Range in a month scale vary from 2 to 39 months.

Selecting the Control Students Using Propensity Score

The following criteria is used to select the control students within the treatment schools:

1. The sample of treatment schools is restricted to the set of 24 schools that implemented the after-school program between the years of 1994 and 1996.

2. Once the treatment group is selected, we also select "potential" control students from the same years and grade levels. Then the propensity scores are estimated separately for each of the six samples using a Multilevel Logistic Model.

Variables potentially relate to the decision for program participation is included in the selection model. Within schools, a student's self-selected decision to participate in the program is, to a large extent, a function of student characteristics. We also consider that the school's characteristics may interact and influence not only the way the program is implemented and delivered but also the way students perceive the program; thus, it may be hypothesized that the decision to participate was a function of school level characteristics (Chinen & Goldschmidt, 2006). The clustering of students within schools is explicitly modeled by specifying multilevel logistic models to estimate the propensity scores. We ran six different models, one for each year and cohort and estimated propensity scores for each of these samples.

3. Once the propensity score is estimated, each treated student is matched to a student from his or her own elementary school. Within each school, the treatment and control conditions share a series of student characteristics such as socioeconomic status, race/ethnicity, and language proficiency, as well as school characteristics. Students are affected by the same school policies, amount of resources, facilities, etc. This procedure controls all observed and unobserved pretreatment variables that are constant for all students within a school (Rosenbaum, 1986). Therefore, to a great extent, treatment conditions within the same school are probabilistically similar.

4. The matching procedure applied is 1-1 nearest neighbor algorithm within a 0.6s caliper and with no replacement. Even though the use of a caliper restricts the number of treated units matched, this procedure guarantees that the nearest neighbor algorithm does not match cases far apart in distance. We test two different calipers 0.1s and 0.6s, one more stringent than the other one. Even though using a caliper of 0.1s produces matched samples closer in terms of distance, it reduces the sample of matched units considerably. However, using a caliper of 0.6s, the matched cases are not very close in distance but groups show balance in terms of the observed covariates. Also, the sample of matched cases increases significantly. Given our goal of maintaining a large treatment sample, we use matches based on the 0.6s caliper.

   The selection of students in each of the six samples has to be sequential given that the same group of students was followed over the course of 3 years. In that regard, the matching is without replacement because once a control student is matched in one year, it is removed from the reservoir of controls for the following year. The subsequent steps are followed to sample controls from cohort II across the 3 years:

   First, we start with those students that are in grade 2 in 1994 and estimate the propensity score as well as treat students with control students.

Second, we continue with those students that are in grade 3 in 1995 but remove the pupils already pulled in 1994.[22] We match treated students with the remaining control students.

These steps generated the treatment sample as well as one of the control samples. To select control students in non-treatment schools, the following criteria were implemented:

To check treatment effect consistency and to make sure that the matching and adjustment by observed covariates sufficed in estimating treatment effects, a second control group in non-treatment schools was sampled. Because of the possibility that unobserved reasons may not be captured by the set of observed covariates in the selection model,[23] having a second control group can be very valuable. For both cases we expect treatment effects to be very similar between the two designs. The selection of control students in non-treatment schools implies two steps:

1. The first step involves the selection of control schools as comparable as possible to the treatment schools to serve as a valid counterfactual group. For this purpose, all schools from the same school district are pre-selected as tentative controls. Pre-treatment school level variables and community indicators from the baseline year 1993 are used to estimate the probability of being a treatment school. The principal criteria for a school to qualify, and receive the treatment is poverty. [24]We include those key selection predictors along with some community variables that capture other relevant dimensions of poverty. The selection model is estimated using a logistic regression model. The estimated propensity score is used to match treated and control schools by the nearest neighbor algorithm within a caliper (0.6s). The structure is 1-1 matching.

2. Once the matched pairs of treated and control schools are identified, within the control schools we select the same grade levels as the ones pulled for the treatment group. Subsequently, we estimate the probability of being a treated pupil by using a logistic regression model as a function of student level variables. Finally, within the matched pair of schools, treated students are matched with control pupils from other schools using the same matching algorithm mentioned before.

In this section, we illustrate the process of matching using the propensity scores, and the sequence of results obtained for the grade 2, 1994 sample.

[22] These are the sample of potential controls.

[23] This explains the decision of not participating in the program within treatment schools (unobserved biases)

[24] This is measured by the percentage of students in the school receiving free lunch, and academics

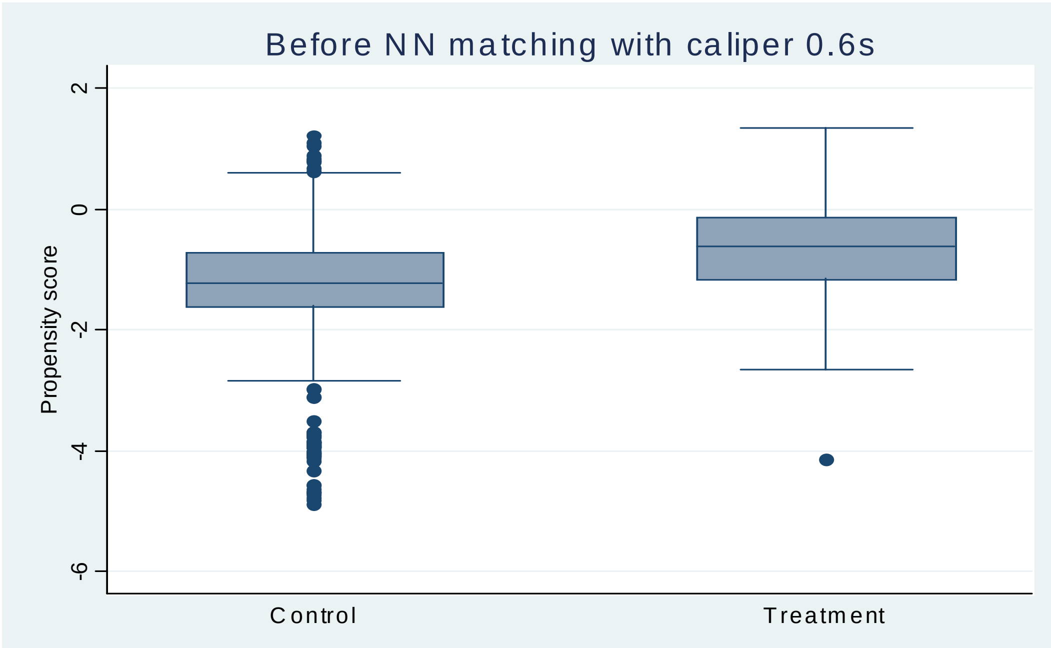


Figure 4. Boxplot of the propensity score by group: Before matching students within treatment schools

Note. NN stands for the Nearest Neighborhood matching process.

This figure illustrates the extent to which the treatment and control groups overlapped in terms of the propensity score. In general, treatment students have propensity scores that overlap to a great extent with those of the control students. Only those pupils with propensity scores located in the upper adjacent zone of the boxplot, or students more likely to receive the treatment, are less likely to find a match pair in the reservoir of control students. Including these treatment students in the sample is likely to compromise the two desired properties of covariates balance and minimum distance. By restricting the matching within a caliper, this group of treated students is expected to be dropped from the sample.

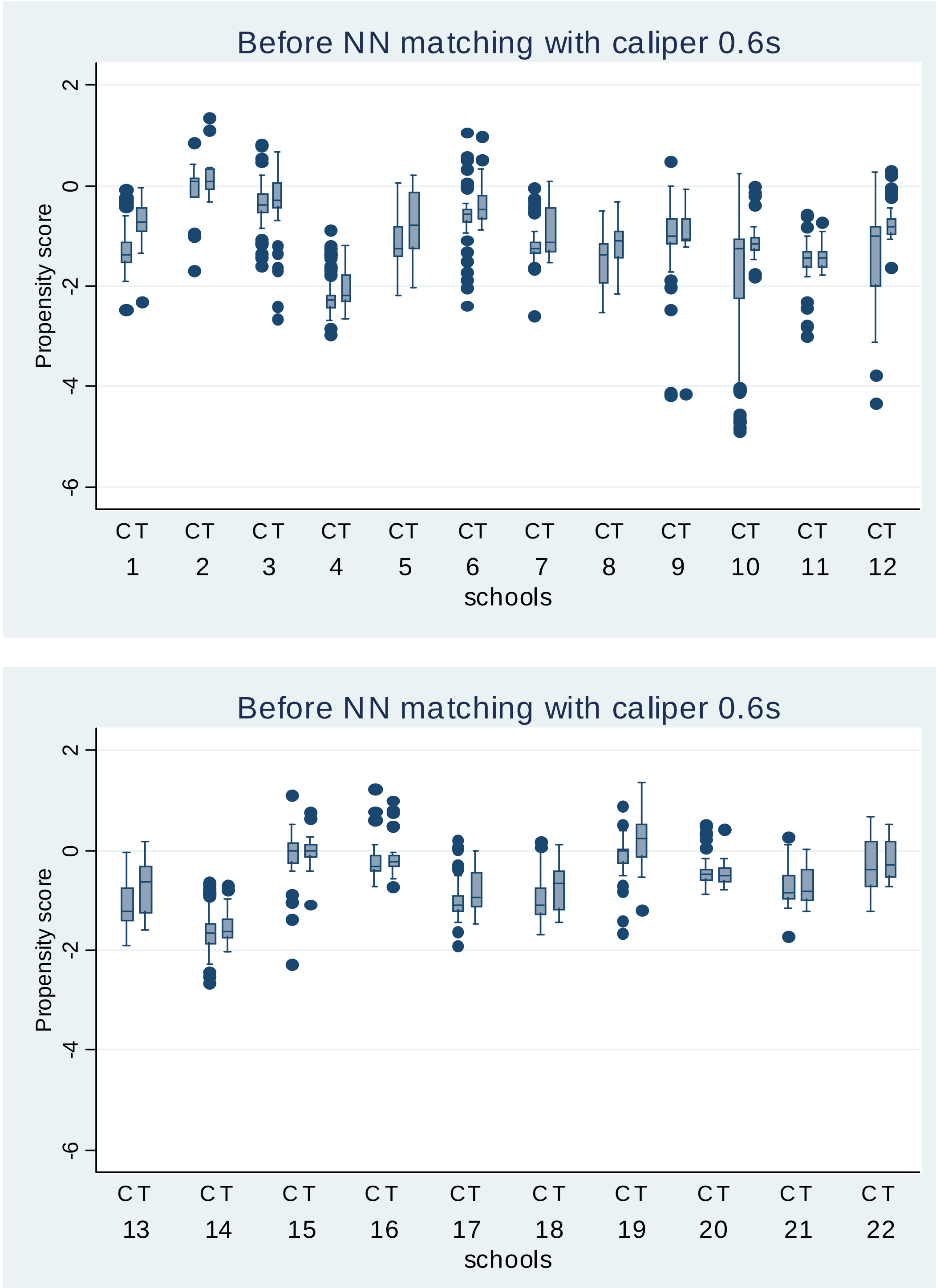


Figure 5. Distribution of the propensity score before matching by school

Although the distribution of the propensity scores for treatment and control groups overlapped greatly, we observe a different situation within schools. Figure 5 shows the boxplot of the propensity score for 22 treatment schools. Within some schools not only is there little overlap between the two groups, but also the reservoir of controls is very small. This provides the insight that within some schools finding a matched control is more difficult especially because of the restriction of matching within a caliper of 0.6s. Also, in schools with little overlap, the average distance is large. Finally, the final sample of treated students is further reduced by restricting the matching within schools.

The following figures present the distribution of the propensity score after matching.

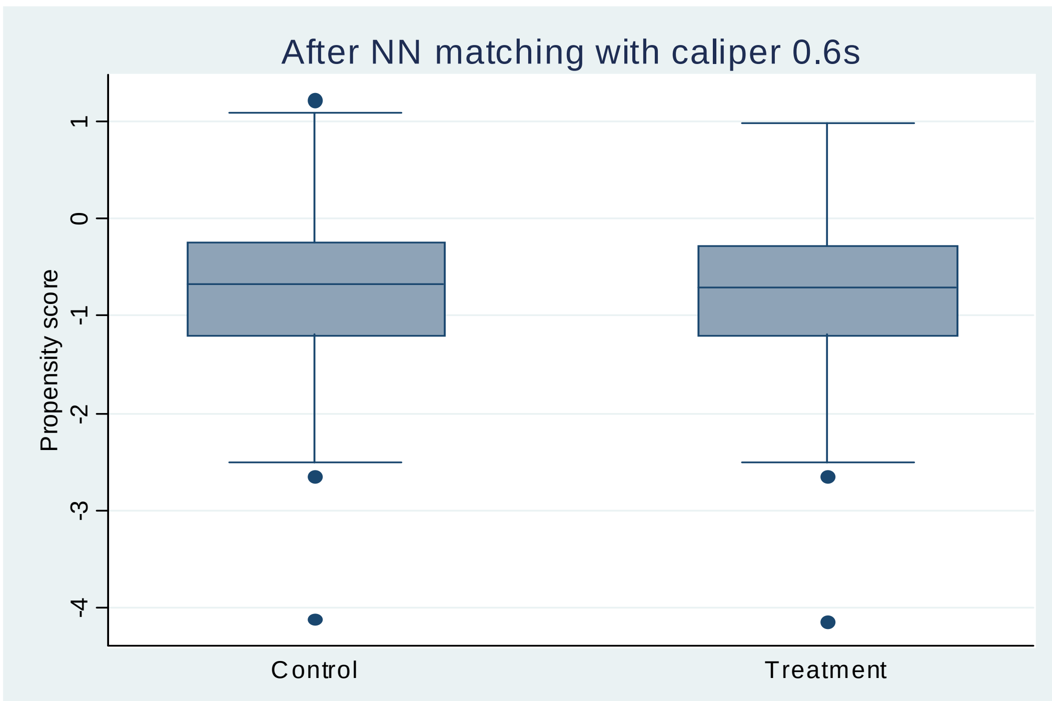
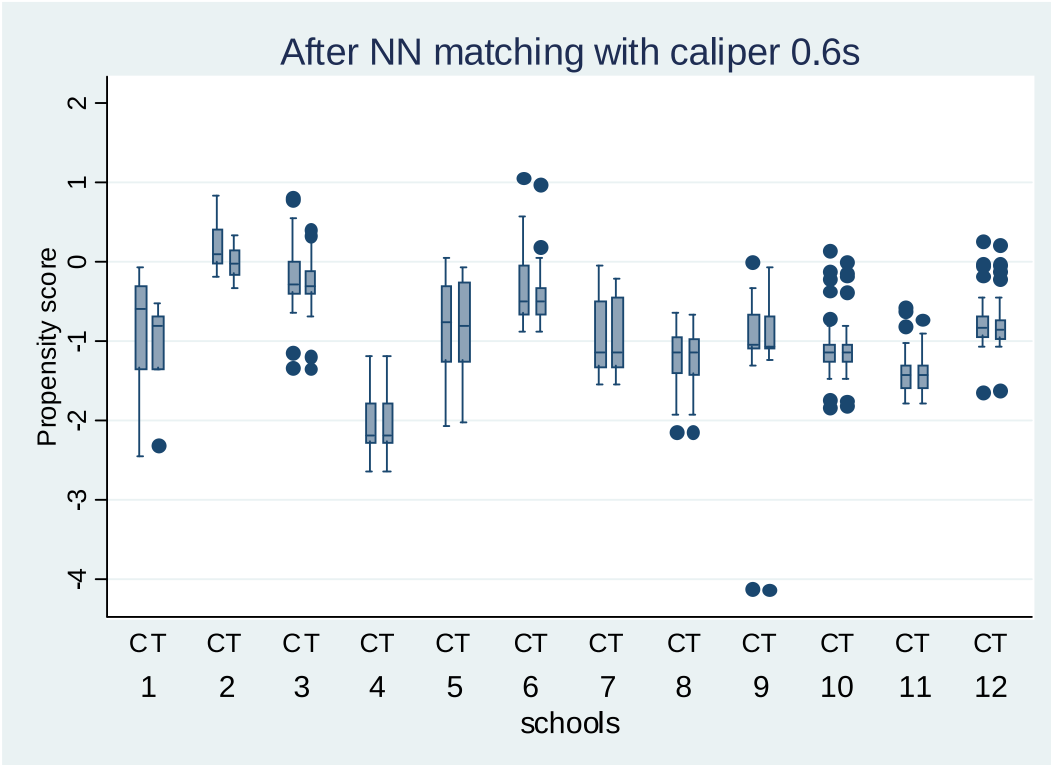


Figure 6. Boxplot of the propensity score by group: After matching students within treatment schools

Comparing the boxplots before and after matching makes it clear that the matching process helps to select the most comparable control students; it eliminates those unique treatment students who do not have comparable matches. This process makes the treatment and control groups more homogeneous.



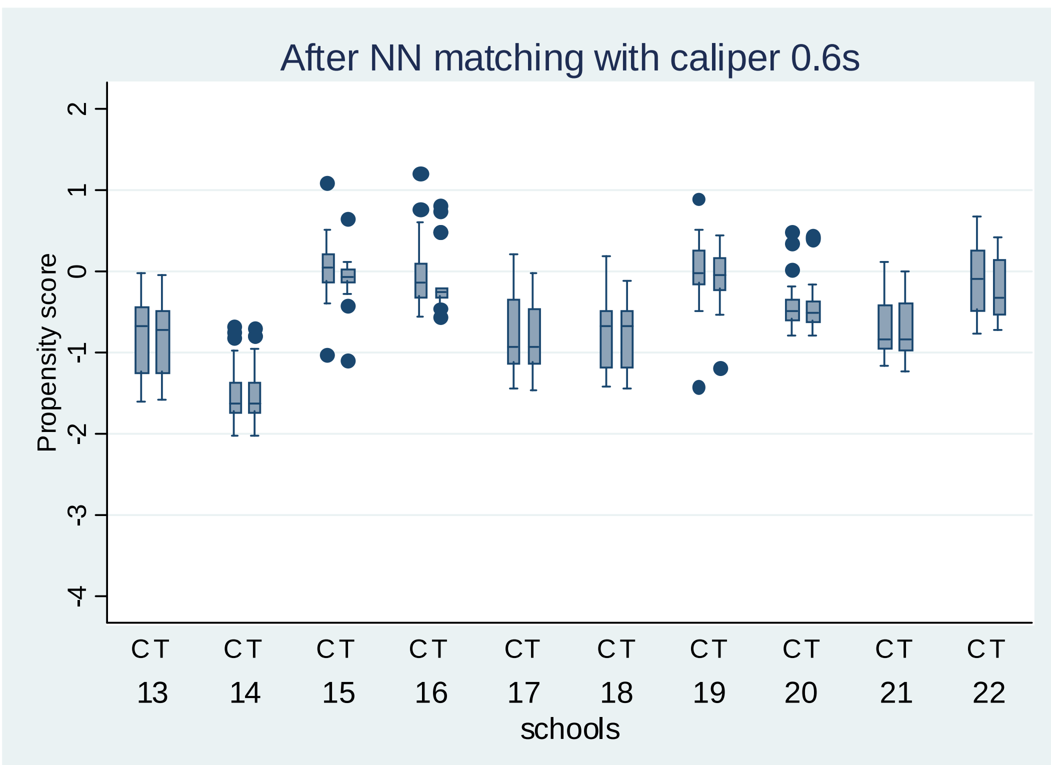


Figure 7. Distribution of the propensity score after matching by school

Within schools, controls and treatment students are clearly more alike after the matching process.

Figures 8 and 9 demonstrate the significant improvement in the overlap of estimated propensity scores by school after matching.

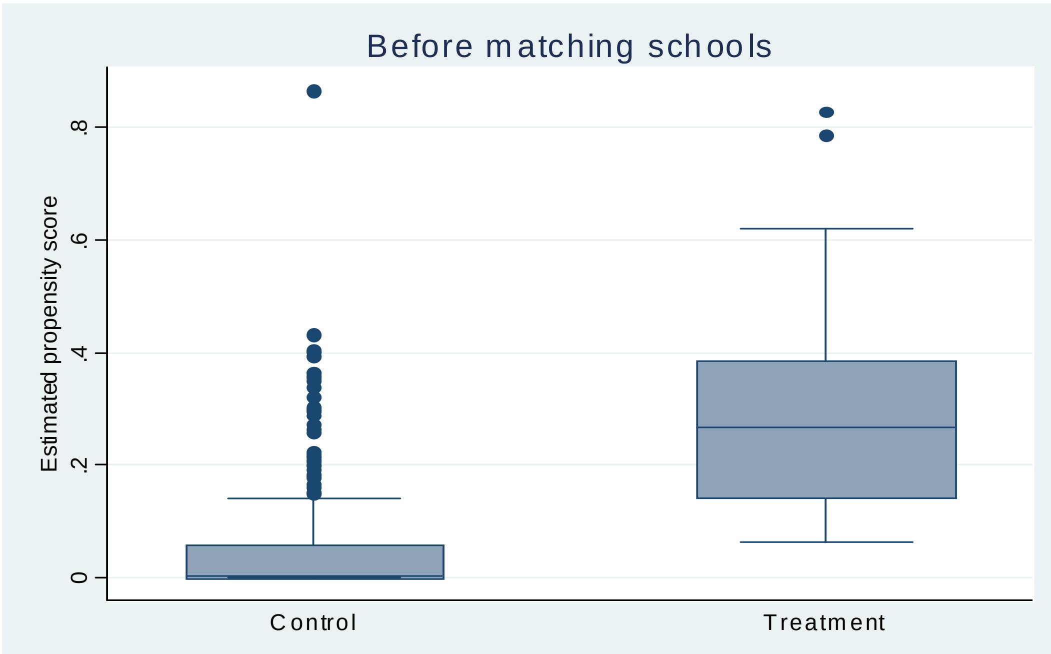


Figure 8. Boxplot of the propensity score by group: Before matching schools

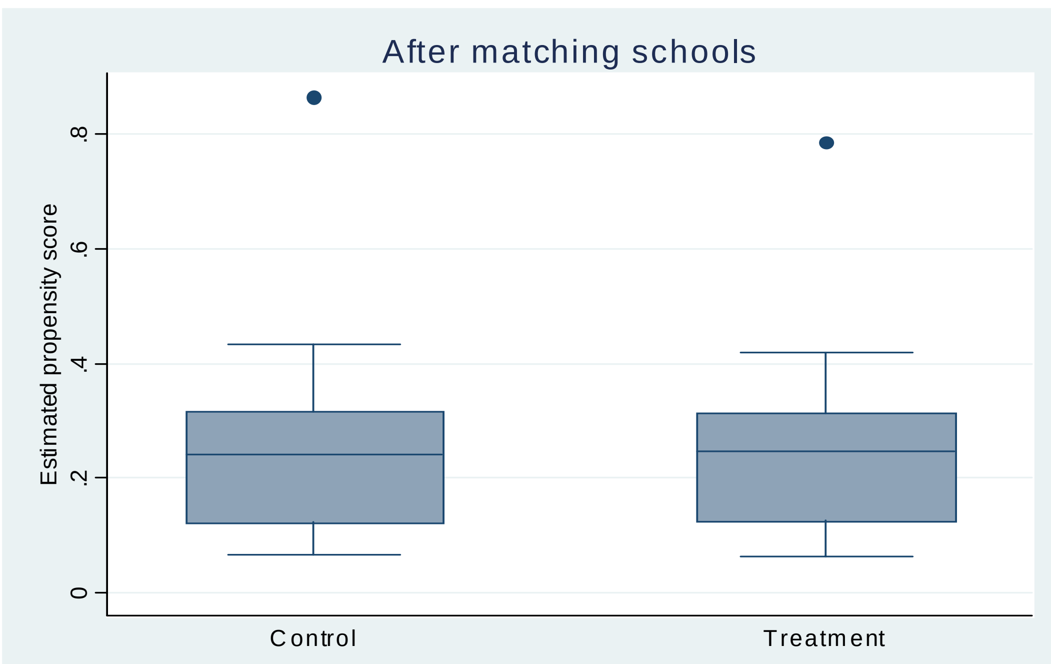


Figure 9. Boxplot of the propensity score by group: After matching schools

Once the control schools are selected, matched pairs of treatment and control schools are specified. We assume that if the selection model at the school level is correctly specified, then the matched schools are very similar in terms of educational and socio-economical factors. Given that, students within these matched pairs of schools are expected to be very similar, and treated pupils could potentially find matched control pairs.

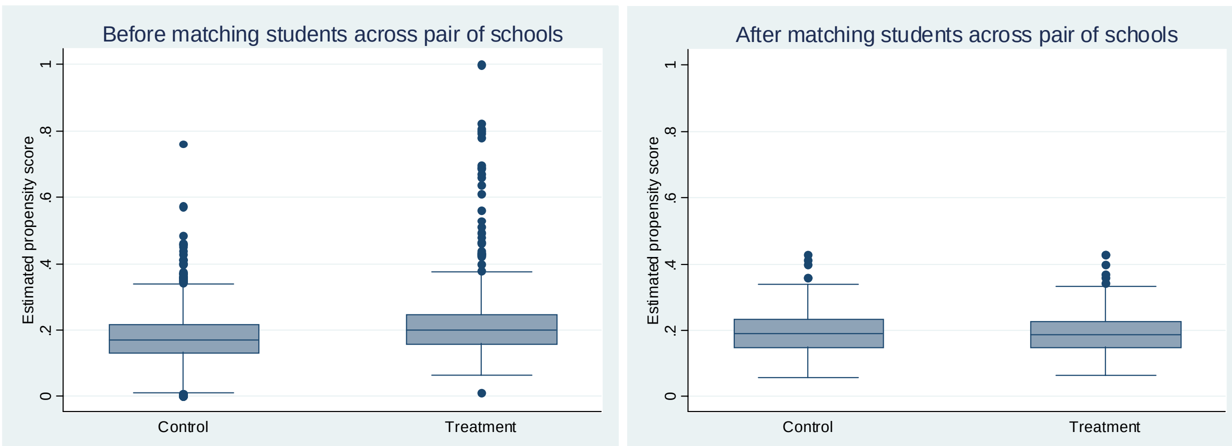


Figure 10. Boxplots of the propensity score by group: Before and After matching students across schools

Demographic Analysis

We examined demographic time to determine how representative schools are of the surrounding neighborhood in which they are located. The purpose of this is two-fold: first, to establish whether contiguous neighborhoods are the best option for matching control and treatment schools; and secondly, to establish a current and historical demographic context that potentially accounts for between-school variation in juvenile behavior. We used 1990 and 2000 census data by zip code to compare these schools' demographic composition to the community. The results are displayed in Tables 11 and 12. Given the strong correlations, we are confident that census data is an appropriate proxy for average family resources available to students in a particular school. Hence, we incorporate census-based family income and wealth information to set the school economic context as a principle, between-school moderating variable.

Table 11

Correlation of Ethnicity Variables – 1990 Census & 1993 School District Data

	1990 Census				
	% White	% Black	% Native American	% Asian	% Hispanic(1)
1993 School District					
% Female	-0.087	0.012	-0.017	-0.073	0.158
% White	**0.835**	-0.434	-0.083	-0.144	-0.710
% Black	-0.403	**0.807**	-0.233	-0.211	-0.202
% Asian	0.157	-0.247	-0.042	**0.588**	-0.264
% Hispanic	-0.465	-0.144	0.237	0.066	**0.849**
% Other	0.081	-0.175	0.134	0.469	-0.170

(1)Hispanic is defined as Other in the U.S. Census data.

Table 12

Correlation of Ethnicity Variables – 2000 Census & 2002 School District Data

	2000 Census				
	% White	% Black	% Native American	% Asian	% Hispanic(1)
2002 School District					
% Female	-0.130	0.084	0.079	-0.102	0.150
% White	**0.798**	-0.357	-0.504	-0.027	-0.713
% Black	-0.315	**0.805**	-0.165	-0.229	-0.226
% Asian	0.180	-0.255	-0.223	**0.593**	-0.315
% Hispanic	-0.496	-0.202	0.579	0.077	**0.818**
% Other	0.541	-0.241	-0.161	0.028	-0.513

(1) Hispanic is defined as Other in the U.S. Census data.

Next, we examined the extent to which demographics have changed over the 10-year period. Figure 11 presents descriptive results for the relevant student characteristics. The results in Figure 11 indicate that the proportion of non-minority (i.e., White) students decrease over time. Also, the percentage of low SES (i.e., Title 1) students increases over time. The results also imply a bifurcation in student academic potential as both GATE (Gifted and Talented) and SWD (Students with Disabilities) student representation have increased.

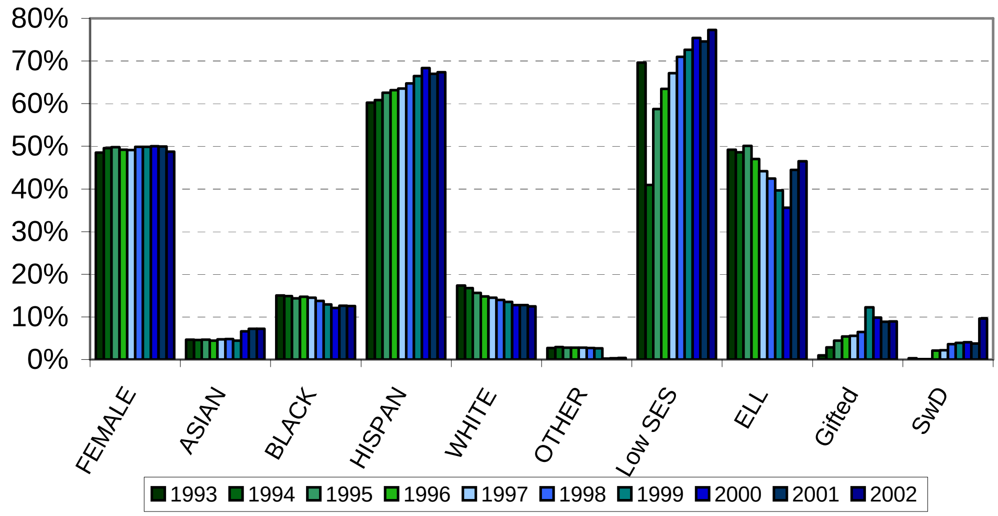


Figure 11. Descriptive summary of student characteristics over time

While the descriptive results highlight the demographic composition of the district changed over time, it does not identify how these changes vary among schools. Figure 12 highlights this for SWD. Figure 12 depicts how both the mean percentage of SWD[25] increased as well as the variation among schools.

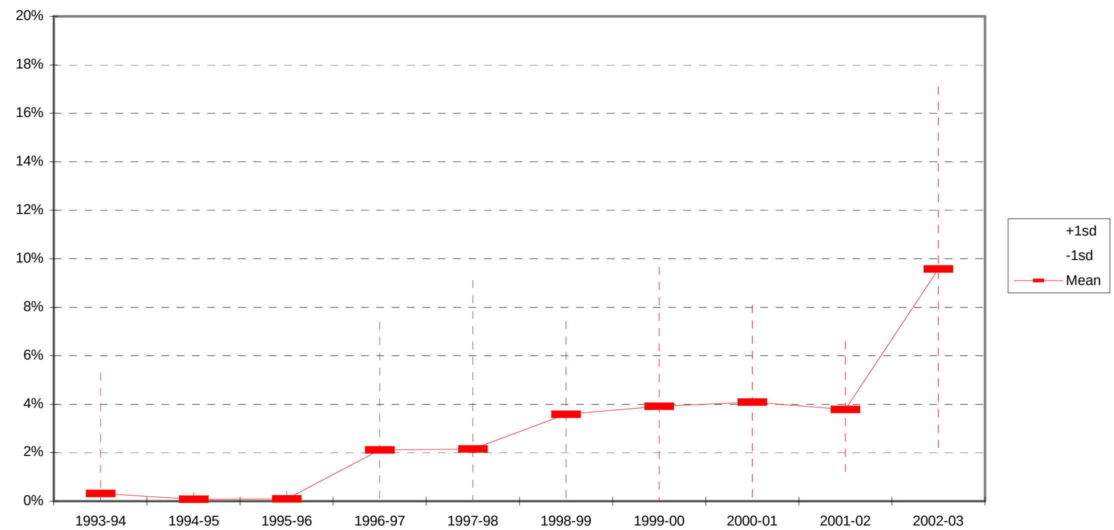


Figure 12. Percentage of SWD and average variation among schools

Contextual, demographic information is included in subsequent models. For example, multilevel survival analyses of students committing crimes might contain time covariates that account for changes in the community that potentially moderate differences between treatment and control student behavior.

We specifically examined zip code or neighborhood demographic characteristics for sampled schools. Table 13 presents the number of schools contained in each city within LAUSD. Table 14 demonstrates that except for one zip code, schools in the sample were unique to the neighborhood. This provides support for using neighborhood data as potential mediating effects on the outcomes under study.

[25] SWD percentages were confounded with increasing district accuracy in identifying SWD. In fact, we used 2002-03 classifications and assigned those to students in prior years.

Table 13

Geographical Information: City of Treatment and Non-Treatment Schools

Mailing city	Frequency	Percent
Non-LA's BEST schools		
1	17	73.91
2	1	4.35
3	1	4.35
4	1	4.35
5	1	4.35
6	1	4.35
7	1	4.35
Total	23	100.00
LA's BEST schools		
8	2	8.33
9	17	70.83
10	1	4.17
11	1	4.17
12	1	4.17
13	1	4.17
14	1	4.17
Total	24	100.00

Table 14

Geographical Information: Zip code of Treatment and Non-Treatment Schools

Non-LA's BEST schools			LA's BEST schools		
Mailing Zip Code	Frequency	Percent	Mailing Zip Code	Frequency	Percent
90001	1	4.35	90002	1	4.17
90002	1	4.35	90003	1	4.17
90003	1	4.35	90006	1	4.17
90005	1	4.35	90008	1	4.17
90011	1	4.35	90011	1	4.17
90019	2	8.70	90012	1	4.17
90022	1	4.35	90017	1	4.17
90026	1	4.35	90019	1	4.17
90031	1	4.35	90026	1	4.17
90033	1	4.35	90028	1	4.17
90043	1	4.35	90031	1	4.17
90044	1	4.35	90032	1	4.17
90047	2	8.70	90033	1	4.17
90061	1	4.35	90037	1	4.17
90065	1	4.35	90044	2	8.33
90291	1	4.35	90059	1	4.17
90731	1	4.35	90291	1	4.17
90744	1	4.35	90744	1	4.17
91331	1	4.35	91303	2	8.33
91405	1	4.35	91324	1	4.17
91605	1	4.35	91342	1	4.17
			91343	1	4.17
Total	23	100.00	Total	24	100.00

Table 15 presents school neighborhood demographic information. This additional descriptive information may provide contextual material beyond that which is provided by the school and is common to many treatment and control schools.

Table 15

Characteristics of the Zip code Where the Sampled Schools are Located - LA'S BEST Schools

Zip Code	Census 1990 total population	Ethnicity (%)					Household Information		
		White	African American	Native American	Asian	Other	2 Person	7 or More Persons	Income in 1989 below poverty
90044	84086	7.92	61.45	.17	1.06	29.38	4948	2312	7326
90026	75214	39.43	2.34	.23	23.59	34.38	5053	1974	4464
90017	21817	35.43	1.54	.66	4.21	58.13	986	696	2101
91343	48702	65.42	5.98	.50	10.04	18.04	4437	891	1666
90002	41154	8.08	57.28	.50	33.86	.26	1971	1605	3956
90028	30705	68.98	6.99	.49	6.08	17.44	3130	458	3300
90012	28487	20.24	18.20	.43	39.75	21.36	1652	327	1402
90291	32882	72.41	9.58	.49	3.20	14.29	4779	373	1691
91303	19517	60.71	3.46	.60	8.52	26.69	1892	361	747
90044	84086	7.92	61.45	.17	1.06	29.38	4948	2312	7326
90744	49236	55.70	4.03	.62	5.65	33.98	2239	1423	2445
90033	57188	52.99	1.92	.31	5.08	39.67	2038	2006	4231
90003	*5369*	*11.66*	*51.66*	*.13*	*.38*	*36.15*	*2301*	*2117*	*4825*
91342	69353	54.31	8.08	.87	4.22	32.50	5302	1575	1609
90011	96057	15.90	27.11	.14	1.19	55.64	2895	4454	7635
90032	46474	38.71	1.70	.65	13.68	45.23	2501	1300	1974
90006	63241	24.15	5.13	.42	20.68	49.60	3185	1544	4836
90019	64737	18.09	42.97	.35	13.93	24.63	5617	1078	3570
90008	32887	4.88	83.41	.45	4.96	6.27	4225	201	2431
90037	56816	12.72	42.12	.02	1.33	43.78	2607	1960	4648
90031	40111	36.42	1.05	.54	28.04	33.92	1673	1484	2543
91324	23270	71.61	2.43	.20	11.42	14.32	2687	315	735
90059	34714	8.89	63.38	.13	.03	27.54	1733	1239	3347

Source: Census 1990

Table 16

Characteristics of the Zip code Where the Sampled Schools are Located - Non - LA'S BEST Schools

Zip Code	Census 1990 total population	Ethnicity (%)					Household Information		
		White	African American	Native American	Asian	Other	2 Person	7 or More Persons	Income in 1989 below poverty
90061	21358	10.15	62.37	.09	.48	26.88	1214	692	1829
90022	65016	32.60	.23	.50	1.18	65.46	2770	2222	3426
90019	64737	18.09	42.97	.35	13.93	24.63	5617	1078	3570
90001	51635	19.65	26.59	.23	.39	53.12	1674	2246	3850
90065	45008	48.37	1.61	.35	16.63	33.02	3157	1096	1820
91605	50546	49.77	3.76	.41	11.85	34.18	3149	1403	1924
90047	48295	4.42	83.02	.30	1.03	11.19	4118	764	2568
91331	87640	35.42	8.14	.43	5.77	50.22	3426	3866	2654
90731	58567	71.05	6.77	.98	5.28	15.89	6186	566	2976
90005	35606	28.54	7.97	.50	24.87	38.10	2421	588	2816
91405	39606	65.02	6.54	.21	9.25	18.96	4077	568	1578
90031	40111	36.42	1.05	.54	28.04	33.92	1673	1484	2543
90002	41154	8.08	57.28	.50	.26	33.86	1971	1605	3956
90744	49236	55.70	4.03	.62	5.65	33.98	2239	1423	2445
90011	96057	15.90	27.11	.14	1.19	55.64	2895	4454	7635
90043	45519	7.51	80.29	.15	1.25	10.77	4367	765	2684
90003	53699	11.66	51.66	.13	.38	36.15	2301	2117	4825
90033	57188	52.99	1.92	.31	5.08	39.67	2038	2006	4231
90291	32882	72.41	9.58	.49	3.20	14.29	4779	373	1691
90026	75214	39.43	2.34	.23	23.59	34.38	5053	1974	4464
90044	84086	7.92	61.45	.17	1.06	29.38	4948	2312	7326

Table 17 presents, by school, the date that LA's BEST is implemented at their site. This information provides the number of years of experience that a particular school has with running the program. Again, this information is potentially a useful proxy for program quality, assuming that program quality improves with experience.

Table 17

Year When the Treatment School Started the LA's BEST Program

School Code	1988	1989	1990	1994	1995	1996
1001	.	X	.	.	.	.
1002	X	.	.	.	.	.
1003	.	.	X	.	.	.
1004	.	.	X	.	.	.
1005	.	.	.	X	.	.
1006	.	X	.	.	.	.
1007	X	.	.	.	.	.
1008	X	.	.	.	.	.
1009	.	.	.	X	.	.
1010	X	.	.	.	.	.
1011	.	X	.	.	.	.
1012	.	X	.	.	.	.
1013	.	.	.	.	X	.
1014	.	.	.	.	.	X
1015	X	.	.	.	.	.
1016	.	X	.	.	.	.
1017	X	.	.	.	.	.
1018	.	.	X	.	.	.
1019	X	.	.	.	.	.
1020	X	.	.	.	.	.
1021	.	.	X	.	.	.
1022	X	.	.	.	.	.
1023	.	.	.	.	.	X
1024	X	.	.	.	.	.

Table 18

School Demographic Characteristics by Groups

Variable	Non-LA's BEST schools			LA's BEST schools		
	Obs	Mean	Std. Dev	Obs	Mean	Std. Dev
Female	24	.50	.06	24	.49	.03
Hispanic	24	.82	.17	24	.81	.18
African American	24	.16	.18	24	.16	.20
Asian	24	.00	.01	24	.01	.02
Other Ethnicity	24	.00	.00	24	.00	.00
Parents' Edu ≥ College	24	.14	.04	24	.15	.04
ELL 1993	24	.94	.04	24	.93	.07
RFEP 1993	24	.00	.00	24	.00	.00
EO 1993	24	.05	.04	24	.06	.07
FRL 1993	24	.01	.05	24	.00	.00
Students' resident different from school location 1993	24	.04	.10	24	.03	.03
Reading CTBS Score 1993	24	33.57	5.25	24	33.96	5.02
Mathematics CTBS Score 1993	24	36.26	6.51	24	38.23	5.07

Figures 13 and 14 highlight another important proxy for program quality – the number of volunteer hours at a site. The histogram displayed in Figure 14 indicates that there is substantial variability among schools in the number of volunteer hours per month available to schools. Figure 13 highlights this by plotting the average number of volunteer hours per month that schools routinely receive.

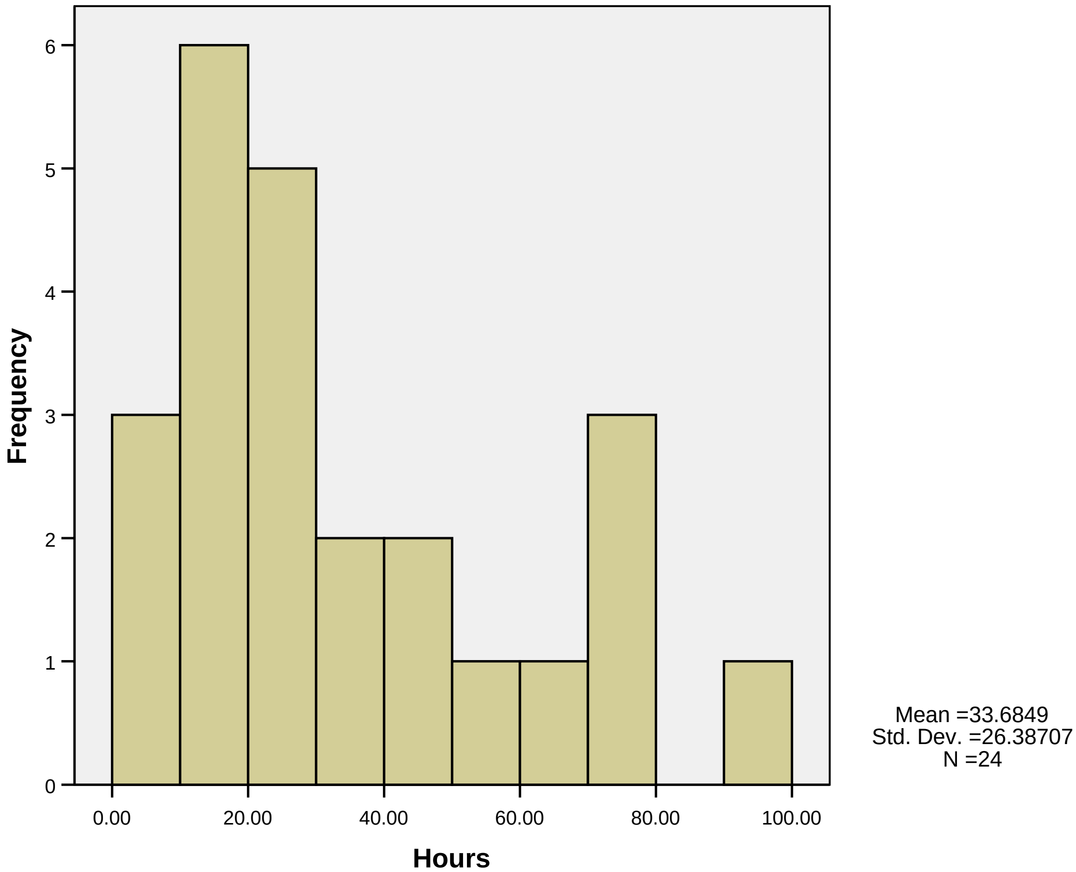


Figure 13. Distribution of the approximate monthly number of volunteer hours across LA's BEST schools

Note. The approximate monthly number of volunteer hours was obtained from cumulative statistics for the dates between May 2001 and February 2006.

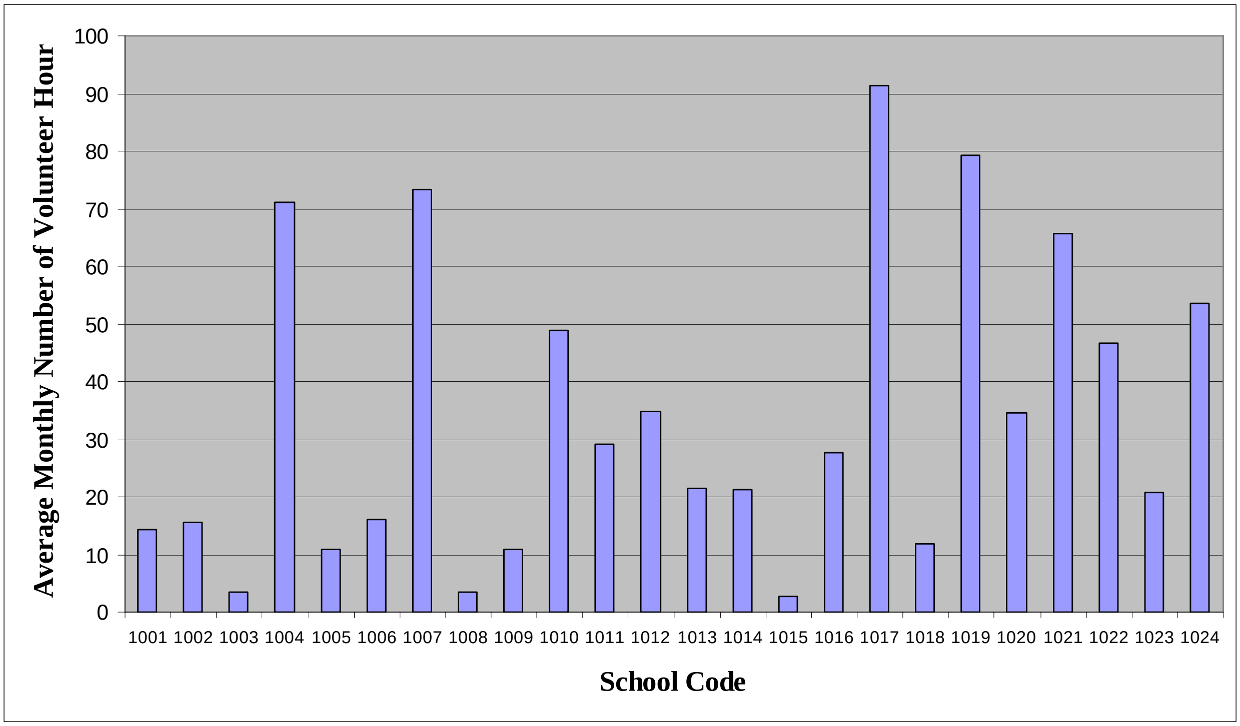


Figure 14. Average monthly number of volunteer hours across LA's BEST schools

Finally, in the following table, we present descriptive statistics at the student level by treatment groups. In this table, control group I refers to the matched students enrolled in the same schools as the students attending LA's BEST; control group II consists of students matched to the treatment students, but who are enrolled in schools without LA's BEST after-school programs.

Table 19

Baseline Characteristics of the Sampled Groups in 1993

Variables	Control 2: In different schools			Control 1: Within LA's BEST schools			LA's BEST Group		
	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.
Female	1902	.50	.50	2210	.50	.50	2458	.49	.50
Hispanic	1902	.85	.35	2210	.81	.39	2458	.81	.39
African American	1902	.13	.33	2210	.16	.37	2458	.16	.37
Asian	1902	.00	.06	2210	.00	.09	2458	.01	.11
Other Ethnicity	1902	.00	.05	2210	.00	.06	2458	.00	.05
Parents' Edu ≥ College	1902	.14	.35	2210	.16	.37	2458	.16	.36
ELL 1993	1445	.94	.22	1523	.94	.23	1787	.93	.25
RFEP 1993	1445	.00	.04	1523	.00	.03	1787	.00	.02
EO 1993	1445	.05	.22	1523	.05	.23	1787	.06	.25
FRL 1993	1902	.01	.10	2210	.00	.09	2458	.00	.07
Students' resident different from school location 1993	1766	.02	.15	2007	.03	.18	2256	.03	.18
Reading CTBS Scores 1993	1379	33.58	21.07	1508	34.42	21.61	1750	34.77	21.16
Mathematics CTBS Scores 1993	1433	35.89	20.71	1561	38.69	21.52	1814	39.08	20.97
GATS 3	1902	.00	.05	2210	.00	.04	2458	.00	.02
SWD 3	1902	.00	.06	2210	.00	.04	2458	.00	.04
FRL 3	1250	.95	.20	1422	.92	.26	1656	.93	.24

Results

The focus of our analysis is on the effect of LA's BEST on student social outcomes; however, we also present results on student academic achievement, as student achievement is related to post-schooling outcomes including labor force participation (Goldschmidt, 1997; Murnane, Willet & Levy, 1995). In the absence of intervention, it is important to reiterate the underlying tenet that when evaluating a program or intervention, the students receiving the intervention would be indistinguishable from control students; they would have outcomes similar to those not receiving the intervention (Campbell & Stanley, 1963). Ideally, we would assign students randomly to a treatment and control group and compare outcomes – the process of randomization ensuring that rival hypotheses do not account for observed differences.

The trade-off of having access to a large urban dataset as well as one of the largest after-school programs in the U.S. is that randomization is not feasible. We rely on both sampling methods and analysis methods to eliminate the effect of potential confounding factors as much as possible. The sampling as described above use propensity scores based on multilevel models to identify two control groups as counterfactual to the randomly selected treatment group (LA's BEST participants).

The analysis methods we rely on are based on multilevel longitudinal modeling – modeling student academic achievement over time, as well as event occurrence over time. Using multilevel survival analysis, we model two events of interest: the time period up to first crime (any crime) and the time up to misdemeanors and felonies, separately.

Given that our analyses emphasize data and outcomes occurring over time, it is important to first consider the nature of the outcome over time (Raudenbush & Bryk, 2002; Osgood & Smith, 2000; Singer & Willett, 2003). This applies to both academic achievement and event occurrence outcomes. This important step is required to adequately model effects of concomitant variables and treatment effects. Below we present results for the two outcomes of interest, beginning with a brief description of the base models that describe the outcome over time.

Student Academic Achievement Results

We examined both Reading and Mathematics achievement over a ten-year period from 1993 to 2003. To more readily isolate treatment effects, we analyzed the achievement results as two samples. Analysis Sample 1 (AS1) includes only students attending LA's BEST schools and compares treatment students to non-treatment students in the same 24 schools. Analysis Sample 2 (AS2) includes all 48 sampled schools and compared treatment students to non-treatment students not attending LA's BEST schools. A common justification for using multiple control groups is that each of these control groups resembles the treatment group in some dimensions but not necessarily in others (Rosenbaum, 1987).

Both AS1 and AS2 non-treatment students (referred as Control 1 and Control 2) are matched on all observable information, theoretically reducing potential unobservable effects, but still allowing for complimentary analyses that offer more robust inferences regarding treatment effects. AS1 treatment and non-treatment (control) students are very similar to each other in that they attended the same schools and generally lived in the same neighborhoods. Arguably, the only difference between treatment and non-treatment (control 1) students in AS1 is that one group attends the LA's BEST after-school program and the other does not. Of course this entails that given the option of attending the program, a group of students otherwise the same on observable characteristics do not choose to receive the treatment – implying a potentially significant, unobservable effect on achievement. AS2 non-treatment students (control 2) are also matched on observable characteristics, but are inherently different as a group because they attend different schools and live in different neighborhoods. To the extent that the Control 2 students are similar to the AS2 students, they might have attended LA's BEST had it been available to them at their school. In this respect, AS2 provides a potentially more objective comparison. The following table shows how the samples are grouped.

	Participants attending LA's BEST schools	Non-participants attending LA's BEST schools (Control 1)	Non-participants attending non- LA's BEST schools (Control 2)
Participant Sample 1 (AS1)	X	X	
Participant Sample 2 (AS2)	X		X

As noted above, the first step is to adequately capture the pattern of achievement over time. Figures 15 and 16 summarize the pattern of Reading and Mathematics achievement over the 10-year span for treatment students as well as for students in Control 1 and control 2. It is important to note that over the 10-year span, the district uses three un-equated assessments; the Comprehensive Test of Basic Skills (CTBS), the Stanford Achievement Test (SAT) v.9, and the California Achievement Test (CAT) v.6. We adjust for shifts in scoring based solely on changes in tests and avoided scaling issues by using Normal Curve Equivalent scores (NCE). NCE scores have limited use for describing absolute achievement growth (Thum, 2002), but they provide unbiased, consistent estimates of relative performance (Goldschmidt, Choi & Martinez-Fernandez, 2003) thus allowing for objective evaluation of program effects. The achievement trends presented in Figures 15 and 16 are consistent with expectations of the effects of changing assessments, as described by Linn (2000). Examining these unconditional patterns is important in order to rule out potential confounding effects of test changes.

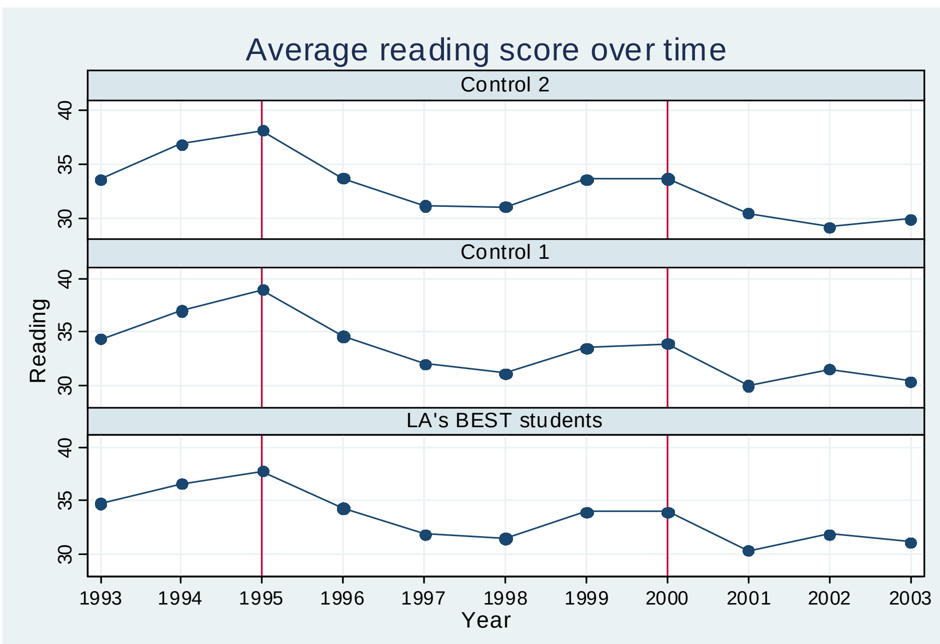


Figure 15. Average Reading score over time

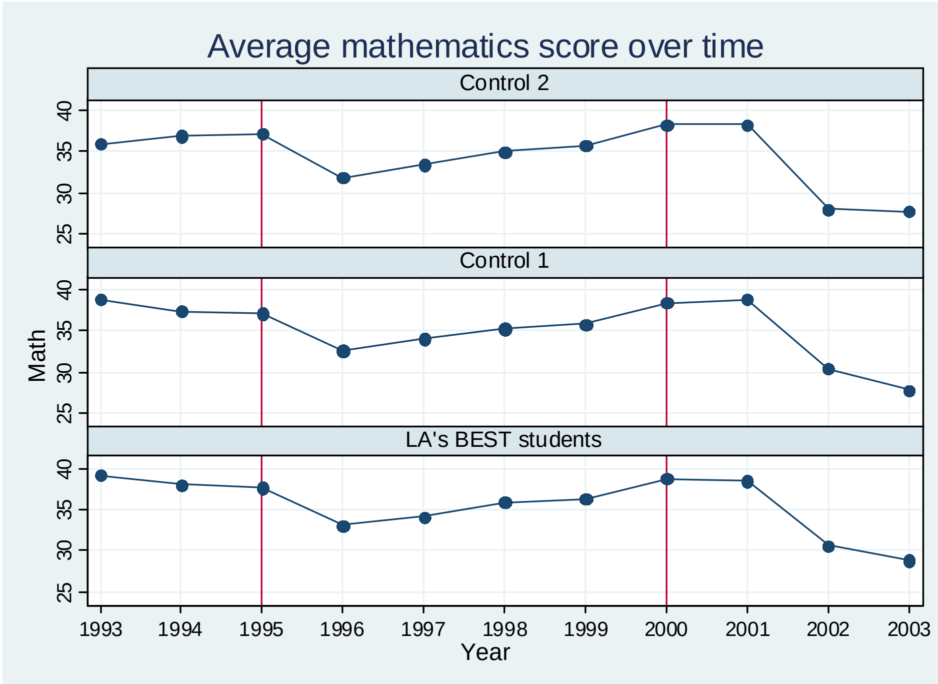


Figure 16. Average Mathematics score over time

Figures 15 and 16 exhibit relatively consistent results across LA's BEST students and the two control groups for Math and Reading achievement scores. In order to capture both the fluctuations in student performance as well as the test change effect, we specify a base model that includes three terms for time and two test change indicators. The results in Tables 20 and 21 indicate that a linear, quadratic and, cubic time (years) are needed to adequately capture achievement changes over the 10-year period.

Results from Tables 20 and 21 show slightly higher mean Reading and Math scores for LA's BEST students (Analysis Sample 1) as compared to the total group (group), of .25 and .07 respectively. Exploring all other growth categories, the results are quite comparable although mixed, with Analysis Sample 2 outperforming Analysis Sample 1 in some areas.

Table 20

Longitudinal Hierarchical Linear Model: Model 1 Reading

	Sample 1		Sample 2	
Variable	Parameter Estimate	S.E.	Parameter Estimate	S.E.
Mean Achievement 1998	31.39	0.55 **	31.14	0.48 **
Linear Growth	-0.65	0.08 **	-0.54	0.08 **
Quadratic Growth	-0.09	0.02 **	-0.09	0.02 **
Cubic Growth	0.02	0.00 **	0.02	0.00 **
Test Effect (CTBS)	5.26	0.36 **	5.54	0.38 **
Test Effect (CAT 6)	1.28	0.58 *	0.11	0.61

Table 21

Longitudinal Hierarchical Linear Model: Model 1 Mathematics

	Sample 1		Sample 2	
Variable	Parameter Estimate	S.E.	Parameter Estimate	S.E.
Mean Achievement 1998	33.47	0.70 **	33.40	0.52 **
Linear Growth	1.00	0.07 **	0.99	0.07 **
Quadratic Growth	-0.09	0.02 **	0.06	0.02 **
Cubic Growth	-0.02	-0.00 **	-0.01	0.00 *
Test Effect (CTBS)	4.35	0.34 **	5.17	0.35 **
Test Effect (CAT 6)	-8.27	0.55 **	-9.58	0.56 **

Figure 17. Model-based Reading and Mathematics trend over time

Figure 17 displays the model-based achievement trends over time and captures the fluctuations in average student performance. The results vary somewhat depending on whether Analysis Sample 1 (LA's BEST schools only) or Analysis Sample 2 (all sampled schools) are included, but in each case the formulation for time provides the best estimates of student performance. The status result reported in Tables 20 and 21 represents student academic achievement in 1998. As with the earlier results, the first few years of intervention have led to higher achievement scores, followed by a period of decline and upward trending data for the final three years.

The achievement models we test examine four components that potentially affect student achievement and achievement growth: time varying student characteristics; time invariant student characteristics; school contextual effects; and treatment effects. Time varying student characteristics can change over time with the status of the particular characteristic having potential academic performance impacts in that year. We include three time varying student characteristics: student travel, English Language Learner (ELL) status and Re-designated English Proficient (REP) status. Travel indicates whether a student is attending his or her home school or is traveling to another school. This is an important student characteristic to consider because it could potentially serve as a proxy for unobservable student and parent motivation, with parents requesting permission to enroll their child in a school other then the designated home school. Although propensity score matching methods take these characteristics into account at the time of sampling (ensuring that both the treatment and control students are equally representative of the various student characteristics) – propensity score matching methods might not control for potential effects of time varying covariates on achievement growth over time.

In order to further account for differences among students in academic achievement and achievement growth we include time invariant student characteristics. These include student demographic characteristics such as gender and race/ethnicity as well as other student background data such as disability status, GATE (Gifted and Talented) status, and cohort. We also include two SES proxies: FRL (Free and Reduced price lunch) eligibility status and parent education level. Each of the student characteristics account for between-student variability in achievement status and achievement growth. This further enhances the precision of treatment effect estimates and thus increases the power of the study (or the probability of detecting the treatment effects under study).

It is important to isolate potential mediating effects of the treatment as well. Often this includes considering program implementation (Seltzer, 2004). The analyses must consider both student inputs and program inputs to examine potential program effects. As part of student inputs, we consider both the variation in exposure and intensity. We define exposure as the number of years of attendance and intensity as the number of days of attendance. Given that the treatment occurs in 24 schools, there is variability in program fidelity or quality. This variation is captured with two school- level implementation proxies: number of volunteer hours per month and number of workshops attended by staff. We also include additional school contextual variables to examine whether the treatment is affected by school level features.[26]

Tables 22 through 29 summarize the results of three Growth HLMs. We present these three models as they summarized the modeling process.[27]

---

[26] These include the percentage of minority students in a school, the percentage of ELL students in a school and the average parental educational levels of the students in a school

[27] We first describe achievement over time, then model between-student effects on achievement before including treatment indicators, and finally model the treatment indicators and between school effects on achievement

The tables are divided into the 1998 achievement scores[28]. It is as follows:

Table 22

Student Achievement Status in 1998: Reading Sample 1

Variable	Model 2 Parameter Estimate	Model 2 S.E.	Model 3 Parameter Estimate	Model 3 S.E.
Mean Achievement 1998	32.98	2.52 **	32.86	2.54 **
Ave volunteer hrs/month			-0.02	0.01
Exposure (years of LA's BEST attend)			-0.28	0.33
Intensity (Daily attendance (ln)			0.19	0.15
Cohort (latter vs. earlier)	1.59	0.36 **	1.59	0.36 **
Female vs. males	1.81	0.36 **	1.83	0.36 **
Hispanic vs. White	-5.25	2.42 *	-5.29	2.42 *
African American vs. White	-10.45	2.49 **	-10.46	2.51 **
Asian vs. White	-3.56	2.92	-3.74	2.92
Other vs. White	-8.32	4.09 *	-8.35	4.09 *
Gifted and Talented vs. not	18.6	0.73 **	18.67	0.73 **
Years in Free or Reduced lunch	0.52	0.1 **	0.52	0.10 **
SWD vs. not SWD	-14.93	0.54 **	-14.94	0.54 **
Parent some College (or more vs. less)	1.04	0.5 *	1.04	0.50 *
Effect of Ever Retained	-5.13	0.38 **	-5.14	0.38 **
Track A vs. all other tracks	0.82	0.42 *	0.77	0.41

[28] corresponds to results at the beginning of middle school, and achievement growth; models the linear portion of achievement growth

Table 23

Student Achievement Growth: Reading Sample 1

	Model 2		Model 3	
Variable	Parameter Estimate	S.E.	Parameter Estimate	S.E.
Average Annual Growth	0.64	0.64	0.58	0.64
Pct White at School			0.008	0.003 **
Ave volunteer hrs/month			0.009	0.002 **
Exposure (years of LA's BEST attend)			0.05	0.07
Intensity (Daily attendance (ln)			0.005	0.03
Cohort (latter vs. earlier)	0.23	0.08 **	0.21	0.08 **
Female vs. males	-0.2	0.08 **	-0.19	0.08 **
Hispanic vs. White	-2.23	0.61 **	-2.18	0.61 **
African American vs. White	-1.29	0.64 *	-1.20	0.64
Asian vs. White	-0.26	0.7	-0.20	0.70
Other vs. White	-3.82	1.01 **	-3.84	1.01 **
Years in Free or Reduced lunch	0.05	0.02 **	0.05	0.02 **
Effect of Ever Retained	-0.58	0.08 **	-0.59	0.08 **
<u>Time Varying Covariates</u>				
Attend Non-resident School	3.13	0.42 **	3.07	0.41 **
English Language Learner (vs. English only)	1.19	0.61	1.17	0.61
Redesignated ELL (vs. English only)	1.01	0.72	0.99	0.72

Table 24

Student Achievement Status in 1998: Mathematics Sample 1

	Model 2		Model 3	
Variable	Parameter Estimate	S.E.	Parameter Estimate	S.E.
Mean Achievement 1998	36.94	2.4 **	36.59	2.40 **
Ave volunteer hrs/month			-0.03	0.01 *
Exposure (years of LA's BEST attend)			-0.25	0.32
Intensity (Daily attendance (ln)			0.23	0.14
Cohort (latter vs. earlier)	0.83	0.34 **	0.83	0.35 **
Female vs. males	-1.04	0.34 **	-1.03	0.34 **
Hispanic vs. White	-4.93	2.29 *	-4.91	2.29 *
African American vs. White	-11.46	2.36 **	-11.43	2.35 **
Asian vs. White	5.98	2.78 *	5.89	2.78 *
Other vs. White	-5.77	3.87	-5.86	3.87
Gifted and Talented Vs not	20.12	0.7 **	20.17	0.70 **
Years in Free or Reduced lunch	0.67	0.09 **	0.66	0.09 **
SWD vs. not SWD	-13.56	0.51 **	-13.56	0.51 **
Parent some College (or more vs. less)	1.21	0.48 **	1.22	0.48 **
Effect of Ever Retained	-5.52	0.37 **	-5.51	0.37 **
Track A vs. all other tracks	0.94	0.4 *	0.91	0.40 *

Table 25

Student Achievement Growth: Mathematics Sample 1

	Model 2		Model 3	
Variable	Parameter Estimate	S.E.	Parameter Estimate	S.E.
Average Annual Growth	1.52	0.59 **	1.53	0.59 *
Pct White at School			0.00	0.00
Ave volunteer hrs/month			0.00	0.00
Exposure (years of LA's BEST attend)			0.02	0.06
Intensity (Daily attendance (ln)			-0.01	0.03
Cohort (latter vs. earlier)	-0.08	0.07	-0.08	0.07
Female vs. males	-0.16	0.07 *	-0.16	0.07 *
Hispanic vs. White	-1.05	0.57	-1.03	0.57
African American vs. White	-1.09	0.59	-1.05	0.59
Asian vs. White	-1.46	0.65 *	-1.44	0.65 *
Other vs. White	-1.77	0.93	-1.75	0.93
Years in Free or Reduced lunch	0.06	0.02 **	0.06	0.02 **
Effect of Ever Retained	-0.43	0.07 **	-0.43	0.07 **
<u>Time Varying Covariates</u>				
Attend non-resident School	2.84	0.38 **	2.84	0.38 **
English Language Learner (vs. English only)	-0.74	0.49	-0.77	0.49
Redesignated ELL (vs. English only)	-0.4	0.47	-0.42	0.47

Table 26

Student Achievement Status in 1998: Reading Sample 2

	Model 2		Model 3	
Variable	Parameter Estimate	S.E.	Parameter Estimate	S.E.
Mean Achievement 1998	27.37	2.4 **	26.47	2.42 **
Ave volunteer hrs/month			-0.01	0.01
Exposure (years of LA's BEST attend)			-0.58	0.34
Intensity (Daily attendance (ln)			0.82	0.24 **
Cohort (latter vs. earlier)	1.49	0.36 **	1.52	0.36 **
Female vs. males	1.5	0.36 **	1.49	0.36 **
Hispanic vs. White	0.72	2028	0.63	2.28
African American vs. White	-4.6	2.38 *	-4.70	2.38 *
Asian vs. White	3.78	2.94	3.43	2.93
Other vs. White	-2.81	3.97	-2.88	3.97
Gifted and Talented Vs not	17.08	0.67 **	17.18	0.67 **
Years in Free or Reduced lunch	0.46	0.1 **	0.45	0.10 **
SWD vs. not SWD	-14.28	0.53 **	-14.24	0.53 **
Parent some College (or more vs. less)	0.78	0.5	0.76	0.50
Effect of Ever Retained	-4.41	0.38 **	-4.37	0.38 **
Track A vs. all other tracks	0.74	0.4	0.69	0.40

Table 27

Student Achievement Growth: Reading Sample 2

Variable	Model 2 Parameter Estimate	Model 2 S.E.	Model 3 Parameter Estimate	Model 3 S.E.
Average Annual Growth	-0.18	0.56	-0.24	0.57
Pct White at School			0.00	0.00
Ave volunteer hrs/month			0.01	0.00 **
Exposure (years of LA's BEST attend)			0.11	0.07
Intensity (Daily attendance (ln)			-0.08	0.05
Cohort (latter vs. earlier)	0.25	0.08 **	0.23	0.08 **
Female vs. males	-0.29	0.08 **	-0.29	0.08 **
Hispanic vs. White	-1.25	0.53 **	-1.25	0.53 *
African American vs. White	-0.84	0.57	-0.86	0.57
Asian vs. White	0.68	0.66	0.68	0.66
Other vs. White	-1.98	0.94 *	-1.97	0.94 *
Years in Free or Reduced lunch	0.07	0.02 **	0.06	0.02 **
Effect of Ever Retained	-0.6	0.08 **	-0.61	0.08 **
<u>Time Varying Covariates</u>				
Attend non-resident School	3.38	0.43 **	3.34	0.43 **
English Language Learner (vs. English only)	0.27	0.61	0.25	0.61
Redesignated ELL (vs. English only)	1.02	0.55	0.97	0.55

Table 28

Student Achievement Status in 1998: Mathematics Sample 2

	Model 2		Model 3	
Variable	Parameter Estimate	S.E.	Parameter Estimate	S.E.
Mean Achievement 1998	32.53	2.39 **	31.81	2.40 **
Ave volunteer hrs/month			-0.05	0.02 **
Exposure (years of LA's BEST attend)			-0.58	0.34 *
Intensity (Daily attendance (ln)			0.92	0.24 **
Cohort (latter vs. earlier)	0.87	0.35 *	0.89	0.36 **
Female vs. males	-1.18	0.35 **	-1.20	0.36 **
Hispanic vs. White	-0.53	2.25	-0.68	2.25
African American vs. White	-6.82	2.36 **	-7.02	2.36 **
Asian vs. White	11.13	2.9 **	10.78	2.90 **
Other vs. White	0.32	3.89	0.13	3.90
Gifted and Talented Vs not	18.79	0.66 **	18.89	0.67 **
Years in Free or Reduced lunch	0.64	0.09 **	0.64	0.10 **
SWD vs. not SWD	-12.97	0.52 **	-12.94	0.53 **
Parent some College (or more vs. less)	1.27	0.49 **	1.26	0.50 **
Effect of Ever Retained	-5.44	0.37 **	-5.40	0.38 **
Track A vs. all other tracks	1	0.39 **	0.93	0.40 **

Table 29

Student Achievement Growth: Mathematics Sample 2

	Model 2		Model 3	
Variable	Parameter Estimate	S.E.	Parameter Estimate	S.E.
Average Annual Growth	0.57	0.53	0.65	0.54
Pct White at School			0.00	0.00
Ave volunteer hrs/month			0.01	0.00 *
Exposure (years of LA's BEST attend)			0.08	0.07
Intensity (Daily attendance (ln)			-0.12	0.06 **
Cohort (latter Vs earlier)	-0.04	0.07	-0.05	0.08
Female vs. males	-0.05	0.07	-0.06	0.08
Hispanic vs. White	-0.21	0.49	-0.19	0.50
African American vs. White	-0.22	0.53	-0.20	0.53
Asian vs. White	-0.99	0.62	-0.97	0.62
Other vs. White	0.27	0.87	0.29	0.88
Years in Free or Reduced lunch	0.08	0.02 **	0.08	0.02 **
Effect of Ever Retained	-0.43	0.07 **	-0.44	0.08 **
Time Varying Covariates				
Attend non-resident School	2.79	0.39 **	2.80	0.40 **
English Language Learner (vs. English only)	-0.91	0.56	-0.96	0.56 *
Redesignated ELL (vs. English only)	-0.16	0.55	-0.21	0.55

The result for the time varying student effect of travel was consistent across content area and sample comparison (AS1 and AS2). That is, students traveling to attend a school other than the assigned home school performed about 2.8 to 3.1 NCE's better than students attending the same schools in the home area ($p < .01$). This is an effect size of approximately 0.14.

The results for the time varying, language status variables are less consistent. ELL students generally caught up in Reading, but fell further behind in Mathematics, while re-designated students performed about as well as native English speakers except in AS2 where they fell slightly behind ($p < .10$).

In general, the effects of concomitant variables are consistent with expectations. That is, gender, race/ethnicity, disability status, GATE status, and SES indicators demonstrate effects in the posited direction. Although the student characteristics are not the focus of these analyses, they confirm that the model captures the achievement dynamic and provides evidence that the sampled students' academic achievement are representative of the district and patterns found in previous research.

We next focus on examining the effects of the treatment. We present the results by analysis sample and content.

Analysis Sample 1: **Reading**. In the sample of LA's BEST schools and for the Reading outcome, we first examine whether a simple indicator for program attendance captures any significant program effects. The average treatment effect is not a significant predictor of either beginning of middle school Reading achievement or linear achievement growth. To further refine the treatment and to take into account the potential between-student and between-site variability in exposure and intensity, we tested two attendance indicators. To measure exposure we used the number of years a student attended the program. We also included daily attendance in the program as a measure of intensity[29]. The coefficient of the exposure predictor is not significant for the beginning of middle school Reading achievement status, but is a significant predictor of linear achievement growth ($p <.10$). The daily attendance coefficient is not significant for Reading achievement status at the beginning of middle school, but is a significant predictor of linear achievement growth ($p <.10$). The results of these models are not presented.

[29] We used the natural log of daily attendance to eliminate the skewed distribution of attendance.

Both years of attendance (exposure) and daily attendance (intensity) may have accounted for between-student variability in treatment effects. We included both exposure and intensity in a single model to determine whether exposure and intensity had separate additive effects. Measured in this way, there is a significant amount of multi-colinearity that affects the results when both are included in the model. This results in neither exposure nor intensity being statistically significant.

We next focus on between-site variability of implementation, for which we had two potential proxies: average number of volunteer hours reported at each site and the number of workshops attended by staff. The number of training sessions attended by staff is not significant. Regardless of the treatment specification, however, the monthly number of volunteer hours is significant and positively related to linear Reading achievement growth ($p < 01$). This result indicates that students attending treatment schools that have greater monthly average of volunteer hours demonstrate faster Reading achievement growth than students attending schools with fewer volunteer hours per month. It is important to note that the impact of volunteer hours on growth is not limited to the time the student is enrolled in LA's BEST, but instead demonstrates statistically significant marginal impact on growth through 2003.

Placing the effect of volunteer hours into context demonstrates both the importance of implementation effects and the limitations of program effects on student Reading achievement in LA's BEST schools. At an average number of volunteer hours (33.7), a student is predicted to gain approximately 0.3 NCEs per year above that of a student not attending LA'S BEST, ceteris paribus. By way of comparison, girls are expected to demonstrate slightly lower Reading achievement growth than boys (-.19 NCEs per year). Hence, a girl attending LA's BEST could eliminate this achievement growth gap over time.

We also examine the relationship of traditional risk factors on student achievement status and growth. We first consider SES. When controlling for other student characteristics, students who are considered low SES by virtue of being eligible for FRL are approximately 3 NCEs ahead of their non-eligible classmates at the end of elementary school. The FRL indicator we included accumulates the number of years a student received services. The results indicate that students receiving more services perform better than students not receiving any services. These students are expected to demonstrate achievement growth of approximately 0.05 NCEs per year ($p < 0.05$) greater than students not receiving lunch services.

Parental education is perhaps better able to differentiate among LAUSD students because a significant majority of them (about 94%) are eligible for FRL. If we use parental education as a proxy for low SES, we find the low SES students are about 1 NCE behind non-low SES students at the end of elementary school, ceteris paribus.

While these effects are statistically significant and demonstrate substantive impact on the margin, other achievement gaps highlight the program's limitations. Schools that have monthly volunteer hours two *s.d.* above average would have another adult present each day after school – effectively cutting the adult/student ratio in half. This amount of additional help would be expected to produce about a 0.76 NCE gain per year. This is a growth effect size[30] of about 0.43, which is substantial for a large scale program. However, even with this effect, Hispanic and African American students with achievement gaps of -5 and -10 NCEs by the beginning of middle school would have a hard time closing the existing Reading achievement gaps.[31] In fact, Hispanics and African American students are expected to fall further behind at a rate of about 2.2 (p <.01) and 1.2 (p <.10) NCEs per year, respectively. Hence, the treatment did not outweigh several existing achievement gaps, nor did it provide enough help to eliminate increasing performance gaps.

Analysis Sample 1: Mathematics. As above, here we present the results for AS1, the 24 LA's BEST schools only. We use the same analysis procedures for Mathematics achievement growth as for Reading. The results are displayed in Tables 33 and 34. Again, the first test of program effectiveness is whether the simple treatment indicator is significant. For the Mathematics score, the average treatment indicator is not significant for the beginning of middle school status or for the achievement growth rate.

We next test the effect of exposure (i.e. years of attendance). Exposure is not significant for either achievement status or growth. Although exposure is not significant, we still test whether intensity of exposure had a significant effect on Mathematics achievement or achievement growth. Attendance is significant and positively related to end of early middle school status but is not significantly related to achievement growth.

We also test whether exposure and intensity, in combination, had a significant effect on Mathematics achievement and achievement growth. When exposure and

[30] We calculated the growth effect size in this case as the estimated effect of hours 2 s.d. above average (vs. no hours) divided by the estimated s.d. of the slope (year effect).

[31] This is even if they are otherwise demonstrating the achievement growth rates of their non- Hispanic/African American classmates, which they are not.

intensity are included simultaneously in the model, none of these treatment indicators are significant. Again, the multicolinearity of this formulation contributes to insignificant findings.

As with Reading achievement, we are particularly interested in whether monthly number of volunteer hours are significantly related to achievement growth. The results indicate that students attending schools with more volunteer hours are slightly behind in Mathematics achievement at the beginning of middle school ($p <.10$), and unlike Reading achievement growth, the average monthly volunteer hours do not have a significant effect on Mathematics achievement growth.

Analysis Sample 2: Reading. We now present results using the second analysis sample, which allow us to make comparisons against a different comparison group and examine effects of differing school context. We again report results for Reading and Mathematics.

We first test the average treatment effect for the beginning of middle school Reading achievement and achievement growth.

Results not presented in this report indicate that the average treatment effect is significant and positively related to results at the beginning of middle school. In other words, LA's BEST students tend to perform higher at the beginning of middle school on the Reading assessment than non-LA's BEST students from schools without LA's BEST programs. There is no effect on Reading achievement growth.

The number of years of attendance in LA's BEST, or exposure, is not significantly related to either the beginning of middle school achievement status or achievement growth. Separately testing for intensity, it reveals that the number of days attending is significantly and positively related to the beginning of middle school Reading achievement status, but not significantly related to Reading achievement growth. When we include both exposure and intensity, daily attendance remains a significant and positive predictor of the beginning of middle school achievement ($p < .01$); while exposure is negatively related to the beginning of middle school Reading achievement status ($p < .10$).

As with AS1, we examine the potential effect of program implementation, approximated by the number of volunteer hours and the number of training course completed by staff. The number of volunteer hours is not significant of the beginning of middle school status, while it is positively related to Reading achievement growth ($p < .01$).

The results of the number of volunteer hours imply a 0.43 NCE increase in average student growth as exposure increased. At a volunteer level of about one person per day (about 2 s.d. above average), the effect on growth is about 1 NCE. This is a growth effect size of about 0.60, which is moderate to large. These results are consistent with those generated with AS1. Again, the impact of LA's BEST with sufficient staff (as made up of volunteers and paid staff) can potentially mitigate some performance gaps but cannot alleviate large gaps,[32] especially where some student subgroups are predicted to fall further behind after elementary school.

Analysis Sample 2: Mathematics. In the case of Mathematics achievement and achievement growth, the simple treatment effect indicator is significantly and positively related to the beginning of middle school Mathematic achievement but not to achievement growth over time. In contrast, the exposure treatment indicator is not significant for either the beginning of middle school Mathematics achievement or achievement growth.

The intensity of treatment indicator (total days of attendance) is a positive and statistically significant predictor of the beginning of middle school Mathematics achievement. However, treatment intensity is not related to Mathematics achievement growth.

Using AS2 results in mixed effects for the treatment, introduces both exposure and intensity into the model. Exposure is significantly and negatively related to the Mathematics achievement at the beginning of middle school (p <.10), but is unrelated to Mathematics achievement growth. Intensity is significantly and positively related to beginning middle school Mathematics achievement (p <.01), but is significantly and negatively associated with achievement growth p <.05).

Finally, we include a measure of program fidelity into the model, exemplified by number of volunteer hours, to determine whether fidelity had an impact on beginning of middle school status and Mathematics achievement growth. The results indicate that volunteer hours are significantly and negatively associated with beginning of middle school Mathematics achievement (p < .05). The number of volunteer hours is

---

[32] there is no performance gap at the beginning of middle school between Hispanic and white students, but Hispanic students begin to fall behind White students at a rate of about 1.25 NCEs per year (p < .05)].

significantly and positively related to Mathematics achievement growth over time ($p < .10$).

Achievement Results Summary. The results of the four analyses described above provide some evidence for beneficial effects of LA's BEST on long-term achievement growth. The longitudinal models we project provide estimates for both achievement status and achievement growth. We are primarily interested in the effect of the treatment on achievement growth as this represents a potential lasting impact of the program. The analyses indicates that simply using a treatment indicator (i.e., splitting students into a treatment and non-treatment group) is insufficient to adequately capture two important program dynamics: student engagement and fidelity. We use the number of years of attendance as an indicator of exposure and the total number of days attended as an indicator of intensity. We then use the number of monthly volunteer hours and staff attending workshops as indicators of quality and program implementation fidelity.

The results summarized in Table 30 highlight several important patterns. Moving beyond simply using a yes/no indicator of program engagement and measuring program exposure demonstrates consistently positive effects for the first post-treatment year. However, these effects are not significant – likely due to the imprecision with which exposure captures treatment effects. This is corroborated by the consistently positive and generally significant results for intensity. This interpretation is further corroborated by the results derived from the models using both exposure and intensity. These imply that exposure as a single indicator confound potential program effects and the sorting of students with fewer after-school alternatives. This finding also indicates the complexity of producing supportive learning environments for students. As expected, one single factor alone cannot determine program quality, it is the interplay of contextual factors such as volunteer hours, student demographic backgrounds, supportive structure etc. that together could make a difference.

Table 30

Summary of Final Achievement Model Results

Models		Reading Sample I	Reading Sample II	Mathematics Sample I	Mathematics Sample II
M1A	Exposure Intercept	+	+	+	+
M1A	Exposure Slope	+ / *	+	-	-
M1B	Intensity Intercept	+	+ / ***	+ / *	+ / ***
M1B	Intensity Slope	+ / *	-	-	- / *
M3	Exposure Intercept &	-	- / *	-	- /*
M3	Intensity Intercept	+	+ / ***	+	+ / ***
M3	Exposure Slope &	+	+	+	+
M3	Intensity Slope	+	-	-	- / **
M3	Volunteer Hours Intercept	- / *	-	- / *	- /**
M3	Volunteer Hours Slope	+ / ***	+ / ***	+	+ /*

* p < 10, ** p < .05, *** p < .01.

Note: Estimations for models 1A and 1B are not presented in this report.

The model including both exposure and intensity is interpreted as the marginal effects of intensity accounting for exposure. That is, students attend the same number of years, but more often demonstrate benefit from positive achievement effects at the beginning of the post treatment time, ceteris paribus. Further, accounting for both exposure and intensity, the results consistently indicate a positive effect on achievement growth. Given the model specification, this result implies that students attending LA's BEST with a greater adult presence fair better throughout the grade span. This provides further evidence of the effects of program engagement.

The results in Table 30 are not entirely consistent, indicating a need for additional research that can more carefully collect implementation and student engagement data. This lends to the tradeoff of going back in time to build a dataset that can provide adequate longitudinal academic and social student outcomes. We gain multiple years of information, but are limited to data generated in the early 1990's that are gathered with this type of program evaluation in mind.

The results demonstrate stability with respect to the sign of the effects across the different samples and tests. Nevertheless, the results also suggest the existence of key, unobserved characteristics of the control students not restricted in the selection models. For example, it is possible that control students in treatment schools decide not to participate in LA's BEST because they chose instead to participate in other after-school activities. In this case, the treatment effect found in AS1 would represent a lower bound estimate for the program effect. More details on the analysis is available in Appendix D.

The mixed achievement results are consistent with previous research on academic achievements of after-school programs. However, the central reason for constructing a sample beginning in 1993-94 is to follow the students longitudinally to completely build a juvenile social history. The following section addresses the results on juvenile crime.

Juvenile Crime Results

Using the multilevel survival analysis methods described above, we estimate a series of models that examine the relationship between youth crime, concomitant student and school characteristics, and LA's BEST program effects[33]. We use any crime as the outcome as opposed to modeling felonies and misdemeanors separately[34]. We present single level survival models that examine felonies and misdemeanors separately[35]. In order to more fully describe crime committed by juveniles, we present descriptive analyses of the crime data.

Descriptive Results of the Criminal Offenses. The following tables and figures summarize findings focused on crime in relation to the treatment. As previously mentioned, we note that students in Control I are students attending the same schools as treatment students. Control II students attend matched schools.

[33] We censor the data by eliminating students once they reach adulthood (18 years old).

[34] Given the complexity of multilevel survival models and the low incidence of felonies and misdemeanors when considered separately as events, the models often do not converge and present unstable results.

[35] The student-level single level results for the any crime outcome are virtually identical to the multilevel model results.

Table 31

Number and Percentage of Juveniles in the Original and DOJ Data

	Sample sent to DOJ Office	Sample with arrest info in DOJ	% in DOJ Data
LA's BEST	2331	184	7.9%
Control I	2331	179	7.7%
Control II	1237	96	7.8%
Total	5898	459	7.8%

It is interesting to note that the proportions of juveniles who are arrested (through) 2006 are very similar across the 3 treatment groups. Table 32 summarizes the demographic characteristics by treatment and control groups and specifically for the students who committed crimes. While in the original sample the distribution of gender is very similar for the 3 treatment groups, in the arrested sample the percentage of males is almost 3 times higher than the percentage of females. In terms of the distribution by ethnicity, the proportion of Hispanics in the original sample is between 81% and 89%, and the percentage of African Americans range from 10% to 17%. In the arrested sample, the proportion of Hispanics and African Americans in control II remains very similar. However, in control I and the treatment group the proportion of Hispanics decrease to 66% and 71%, and the percentages of African Americans increase to 32% and 27%, respectively.

Table 32

Distribution of the Samples by Gender and Ethnicity

	Original Sample						Arrested Sample					
	Control II		Control I		LA's BEST		Control II		Control I		LA's BEST	
	N	%	N	%	N	%	N	%	N	%	N	%
Female	641	52%	1180	51%	1159	49.7%	20	21%	50	28%	52	28%
Male	596	48%	1151	49%	1172	50.3%	76	79%	129	72%	132	72%
Hispanic	1096	89%	1886	81%	1889	81%	80	83%	118	66%	131	71%
African American	125	10%	390	17%	383	16%	13	14%	57	32%	50	27%
Asian	7	1%	22	1%	31	1%	0	0%	0	0%	1	1%
Other	1	0%	10	0%	8	0%	1	1%	0	0%	0	0%
White	8	1%	23	1%	20	1%	2	2%	4	2%	2	1%
Total	1237	100%	2331	100%	2331	100%	96	100%	179	100%	184	100%

The sample of 459 students with arrest information committed a total of 990 offenses. We were able to code about 90% of these offenses. However, we were unable to correctly classify 98 of 990 arrest charges because of incomplete or unclear coding information. For all the cases, the offenses counted in the following tables represent the most severe crime –a felony or misdemeanor - recorded for each arrest. In the cases where the offender had several misdemeanors under the same offense code, only the first code was selected and reported in the following tables. Similarly, if the offender showed several felonies under the same arrest code, only the first one is presented.

The subsequent tables and figures present the number of offenses and the characteristics of the offenders. Note that since a single juvenile can commit several and different types of offenses, the same individual could be counted several times.

Table 33

Number and Percentage of Juveniles by Type of Offense

General crime category	Control II		Control I		LA's BEST	
	N	%	n	%	n	%
Felony	118	66%	218	59%	222	64%
Misdemeanor	62	34%	149	41%	125	36%
Total	180	100%	367	100%	347	100%

Table 34 indicates that among the main felony categories, violent crimes are more prevalent in control group I and the treatment group in comparison to the number of incidents in control group II. The number of other felonies is very similar across the 3 groups.

Table 34

Characteristics of the Sample by Type of Felony Offense

	Control II		Control I		LA's BEST	
Felony categories	N	%	n	%	n	%
Violent	30	26%	77	35%	75	34%
Property	51	44%	89	41%	98	44%
Drug offenses	16	14%	22	10%	24	11%
Sex offenses	5	4%	2	1%	5	2%
Other	15	13%	27	12%	20	9%
Total	117	100%	217	100%	222	100%

Table 35

Number and Percentage of Offenses by Arrest Offense Code

Arrest offenses codes	Control II		Control I		LA's BEST	
	N	%	n	%	n	%
F Murder		0%	3	1%		0%
F Forcible Rape	1	1%	1	0%	1	0%
F Robbery	8	4%	31	8%	35	10%
F Assault	21	12%	42	12%	39	11%
F Burglary	17	10%	26	7%	29	8%
F Theft	26	15%	41	11%	40	12%
F Motor Vehicle Theft	7	4%	21	6%	26	7%
F Arson	1	1%	1	0%		0%
F Forgery		0%		0%	3	1%
F Drug-Of Narcotics	6	3%	10	3%	11	3%
F Drug-Of Marijuana	2	1%	8	2%	6	2%
F Drug-Of Dangerous Drugs	8	4%	4	1%	7	2%
F Sex-Of Lewd or Lascivious	3	2%	2	1%	2	1%
F Sex-Of All Other	2	1%	2	1%	3	1%
F Weapons	15	8%	25	7%	20	6%
M Assault and Battery	11	6%	45	12%	28	8%
M Petty Theft	6	3%	12	3%	13	4%
M Drug-Of Marijuana	15	8%	35	10%	34	10%
M Drug-Of Other Drugs	3	2%	4	1%	10	3%
M Annoying Children		0%	2	1%		0%
M Prostitution	1	1%	4	1%	4	1%
M Liquor Laws		0%	1	0%	1	0%
M Disturbing the Peace		0%	2	1%		0%
M Disorderly Conduct	1	1%		0%		0%
M Malicious Mischief	1	1%		0%		0%
M Vandalism	20	11%	33	9%	32	9%
M Trespassing	2	1%	6	2%	1	0%
M Weapons		0%	3	1%	2	1%
M Driving under the Influence	1	1%		0%		0%
M Selected Traffic Violations		0%	1	0%		0%
Total	178	100%	365	100%	347	100%

The following figures show the cumulative percentage of arrests for two groups, cohort II and III. In this study, cohort II includes those students who were in grade 2 in 1994, while cohort III contains pupils who were attending grade 3 in the same year.

The purpose of the figures is to examine the timing of crime by youth. In general, Figures 18 through 21 demonstrate that the rate of crime began to increase dramatically around middle school.

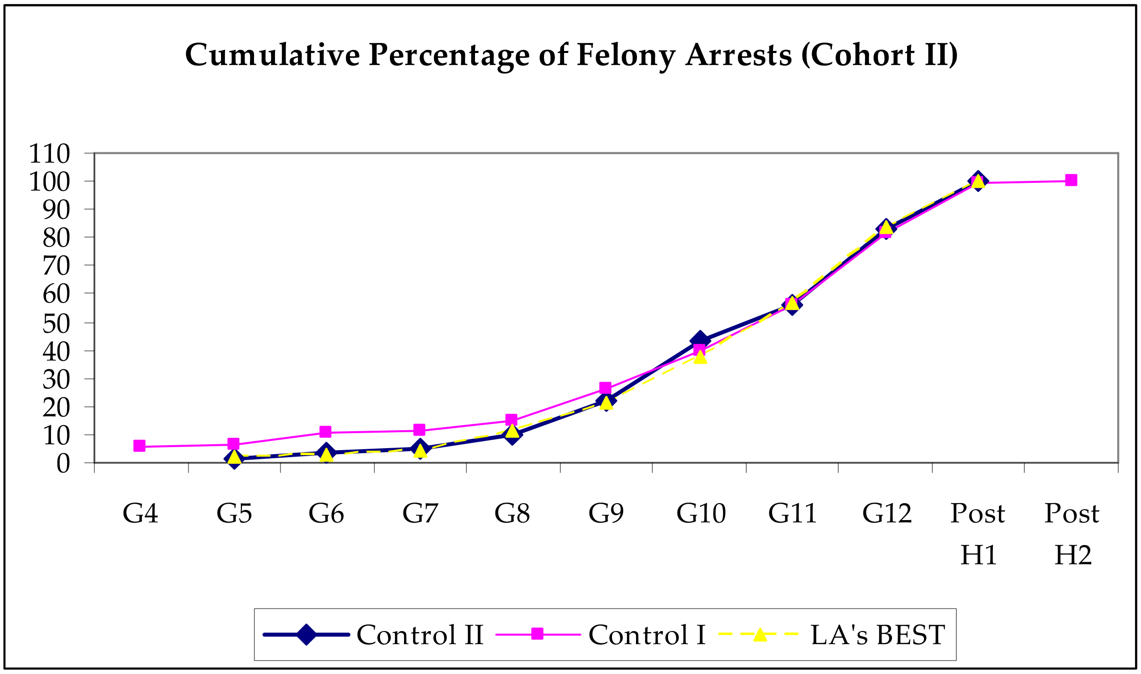


Figure 18. Cumulative percentage of felony arrests (cohort II)

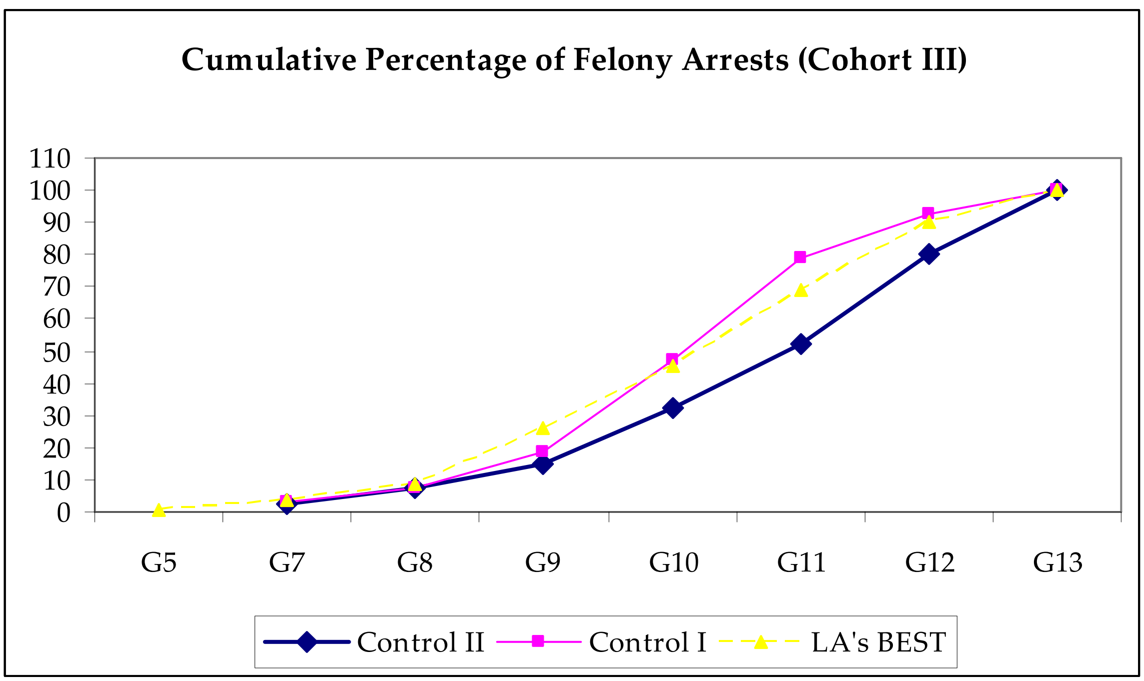


Figure 19. Cumulative percentage of felony arrests (cohort III)

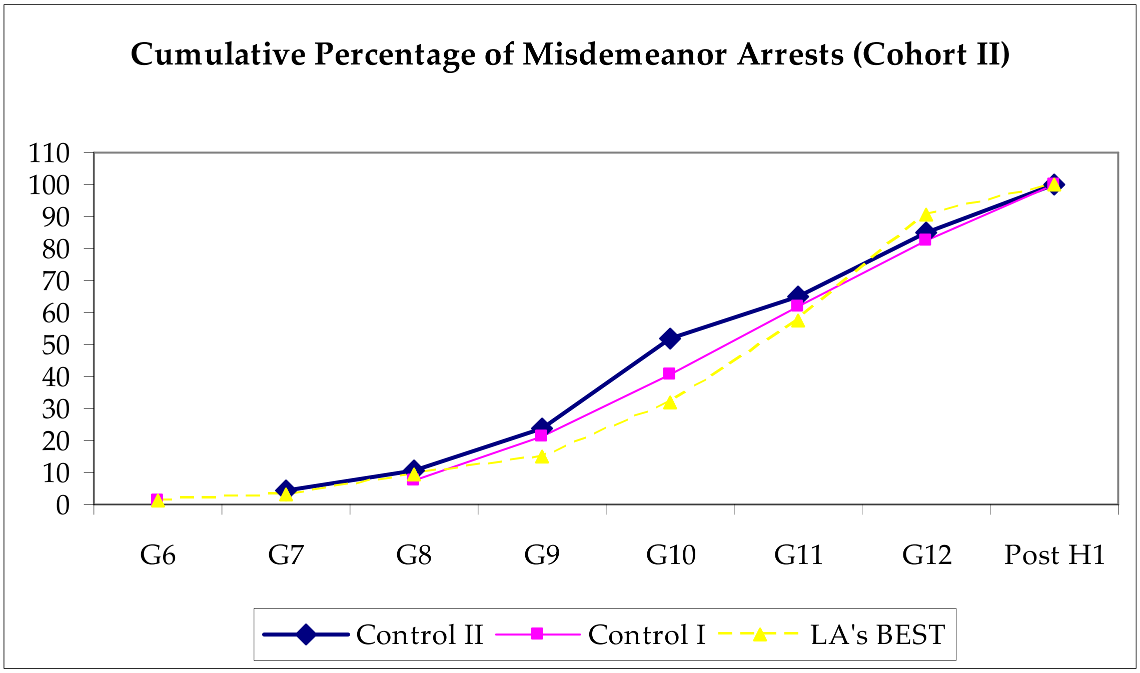


Figure 20. Cumulative percentage of misdemeanors arrest (cohort II)

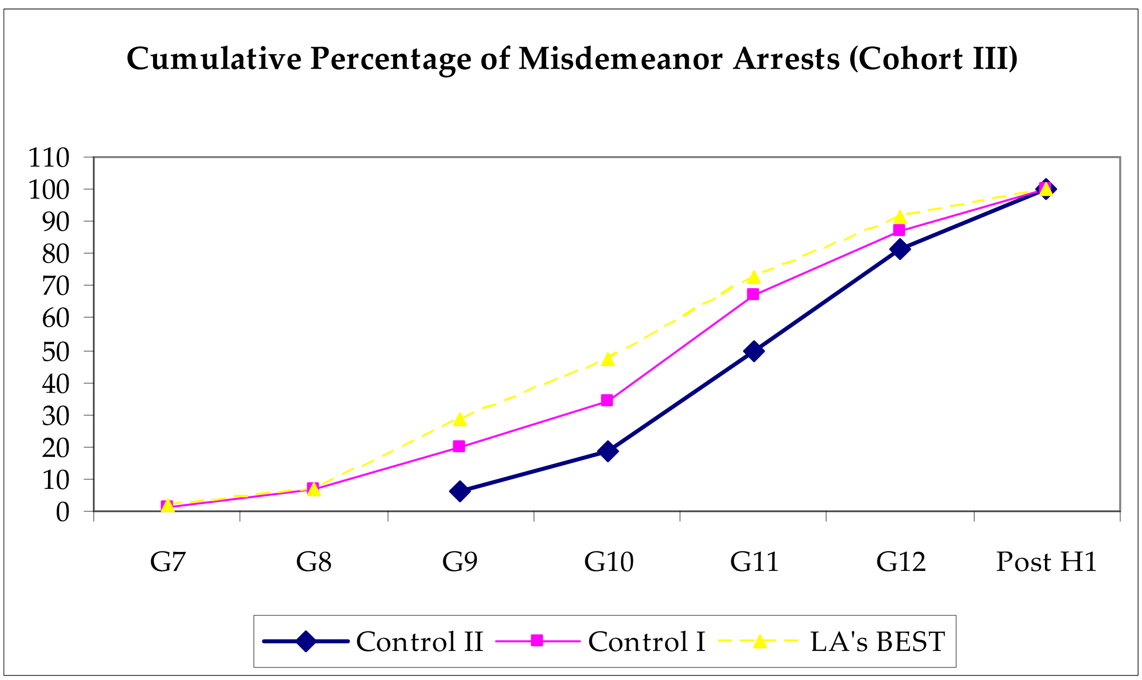


Figure 21. Cumulative percentage of misdemeanor arrests (cohort III)

The cumulative crime percentages are different for cohorts II and III. The percentages of felony offenses tend to increase with less magnitude for LA's BEST students from cohort II compared to controls I and II, until pupils reach grade 11. After grade 11, the three groups' curves overlap. Instead for cohort III, the percentage of felony offenses tend to increase faster for LA's BEST students until they reache grade 10. After Grade 10, control I pupils' rate of offenses exceed the ones from treatment students.

A comparable situation is observed in the case of misdemeanor offenses. For cohort II, LA's BEST students' rate of committing misdemeanor offenses over time tend to be below controls I and II until grade 11. After grade 11, LA's BEST students exceed the two control groups. In contrast, the treatment students' offense rates for cohort III surpass the curves of controls I and II in all the grade levels.

To interpret these results, information in Table 35 is presented. Pupils in cohort II attend the LA's BEST program for more years and more days in the year, than students in cohort III. In general, these results together with those presented previously suggest that the program is effective in capturing at-risk pupils and those who are more likely to commit a crime in the future. The figures presented above indicate a decreased rate of committing criminal offenses for treatment students.[36]After grade 11, the treatment and control offense rates tend to be similar, suggesting that the possible positive effects of the program vanish after this grade. Again, these results are investigated in more detail using multilevel survival models.

---

[36] These students attend the program for a greater number of years and/or days (i.e. cohort II students)

Table 36

Exposure and Intensity of LA's BEST Students' Attendance by Cohort

	Cohort II		Cohort III	
Years in the program (Exposure)				
	N	%	N	%
1	604	46.8	645	55.6
2	355	27.5	278	23.9
3	183	14.2	202	17.4
4	129	10.0	36	3.1
5	19	1.5		
Total	1,290	100	1,161	100
Days attended over a period of 5 years (Intensity)				
	Mean	*SD*	Mean	*SD*
	194.3	156.99	174.0	132.12

The subsequent table presents the distribution of the sample by treatment groups and the number of students not arrested in each of these groups.

Table 37

Number of Students Not Arrested by Treatment Groups

Duration	Original N	Not Arrested	
		N	%
Control II	1237	1141	92%
Control I	2331	2152	92%
Treatment:			
Low	1158	1076	93%
Med Low	624	564	90%
Med High	375	344	92%
High	174	163	94%
Total	5899	5440	92%

The treatment group is divided into four exposure sub-groups, based on the duration of the juveniles' attendance in the program. The category "low" corresponds to those students who attend LA's BEST program for only one year, "medium low" to those who attended for two years, "medium high" to those who attended three years, and "high" to students who attended four or five years during the period between 1993 and 1997. Tables 38 and 39 summarize the number and percentages of offenses for each of the treatment groups.

Table 38

Number of Offenses by Crime Categories and Treatment Groups

Groups	General crime categories		Felony categories				
	Misdemeanor	Felony	Violent	Property	Drug offenses	Sex offenses	Other
Control II	62	118	30	51	16	5	15
Control I	149	218	77	89	22	2	27
Treatment:							
Low	46	112	41	46	13	3	9
Med Low	54	63	18	33	7		5
Med High	19	30	12	12	3	2	1
High	6	17	4	7	1		5
Total	336	558	182	238	62	12	62

Table 39

Number and Percentage of Offenses by Crime Categories and Treatment Groups

Groups	General crime categories		Felony categories				
	Misdemeanor	Felony	Violent	Property	Drug offenses	Sex offenses	Other
Control II	62	118	25%	43%	14%	4%	13%
Control I	149	218	35%	41%	10%	1%	12%
Treatment:							
Low	46	112	37%	41%	12%	3%	8%
Med Low	54	63	29%	52%	11%	0%	8%
Med High	19	30	40%	40%	10%	7%	3%
High	6	17	24%	41%	6%	0%	29%
Total	336	558	33%	43%	11%	2%	11%

The results from Tables 38 and 39 suggest a relationship between the number of years of attendance in the program and the type of felony offense committed by the juvenile. In general, those who attend four or five years tend to commit fewer drug and sex-related crimes than those treatment pupils who attended fewer years. Since treatment students vary with regards to the number of years of attendance (exposure) and the number of days attended each year (intensity), the possible pattern observed in these tables needs to be explored in more detail, controlling for students engagement (average attendance per year).

The following figures attempt to present the differences in intensity of attendance between non-offenders and different types of offenders. Again, those offenders are counted in the data as many times as the number of times they are arrested while information of non-offenders is counted once.

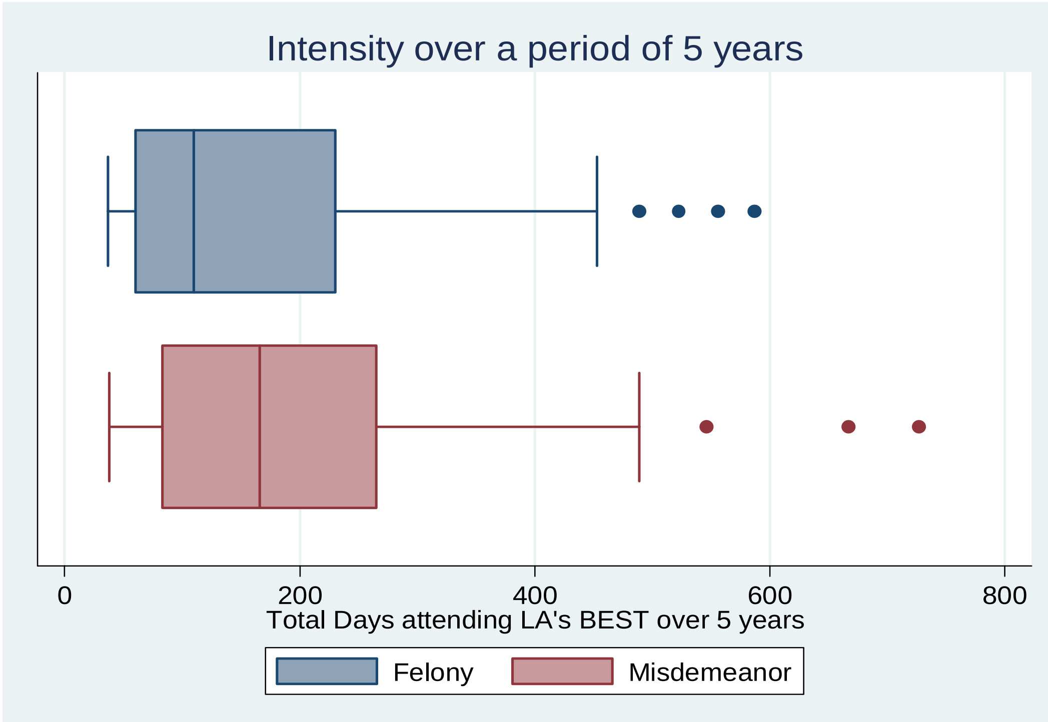


Figure 22. Boxplots of intensity of attendance by general crime category

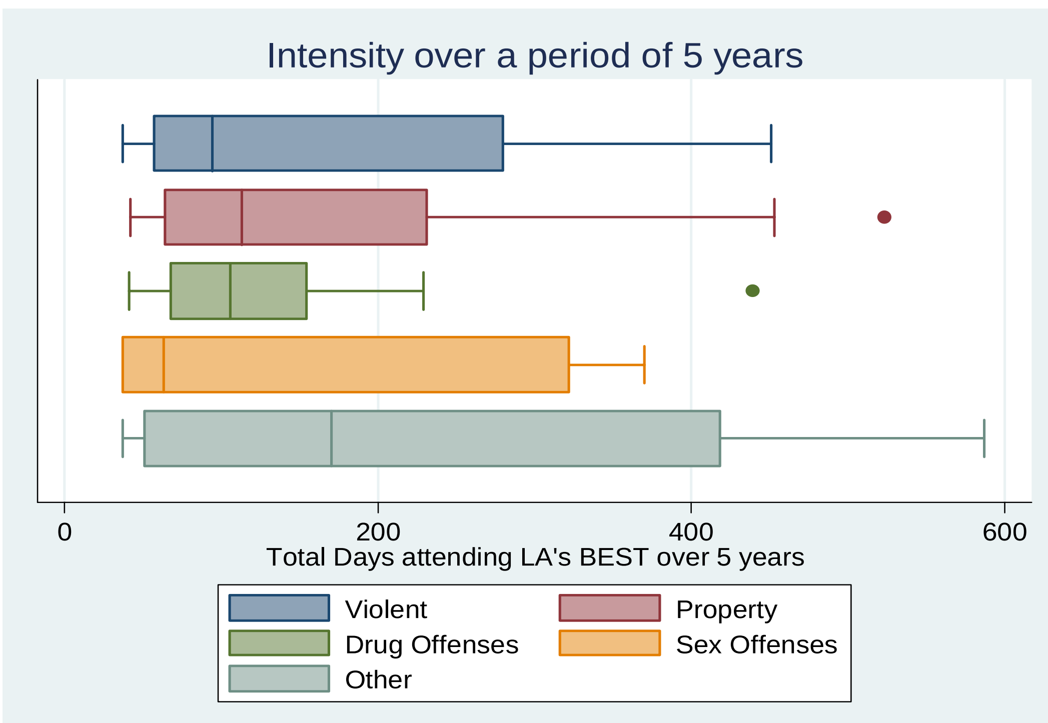


Figure 23. Boxplots of intensity of attendance by type of felony

Table 40

Descriptive Statistics of Days of Attendance Over a Period of 5 Years by Type of Arrest Offense

Arrest Offense Charge	N	Mean	*SD*	Min	Max
No Crime	2259	185.2	146.65	37	780
Felony	228	163.5	132.68	37	587
Misdemeanor	129	189.6	130.75	38	727
Missing	37	166.1	116.69	37	446
Type of Felony Offense					
Violent	77	162.4	129.31	37	451
Property	101	161.4	126.28	42	523
Drug Offenses	25	124.1	85.64	38	439
Sex Offenses	5	165.8	165.71	37	370
Other	20	226.6	195.44	37	587
Missing	166	184.4	127.79	37	727

*Missing included unclassified offenses.

Even though the boxplots for felony and misdemeanor offenders overlap to a large extent, it seems that treatment juveniles who committed misdemeanor offenses also tend to show more exposure to the program than those who committed felony offenses. Furthermore, treatment juveniles who commit sex or drug offenses show less exposure to the program than other felony offenders. Those who commit "other offenses" are those who exhibit more engagement to the program than other offenders.

Figures 24 through 35 illustrate school crime data for both the sample and the un-sampled schools in LAUSD. The absolute figures are not directly comparable because the sampled schools are elementary schools, while the remaining schools include middle and high schools, where there are generally higher incidences of crime. Overall, the trends over time were the same. We use the school crime data in an attempt to account for between-school variation in the crime rate.

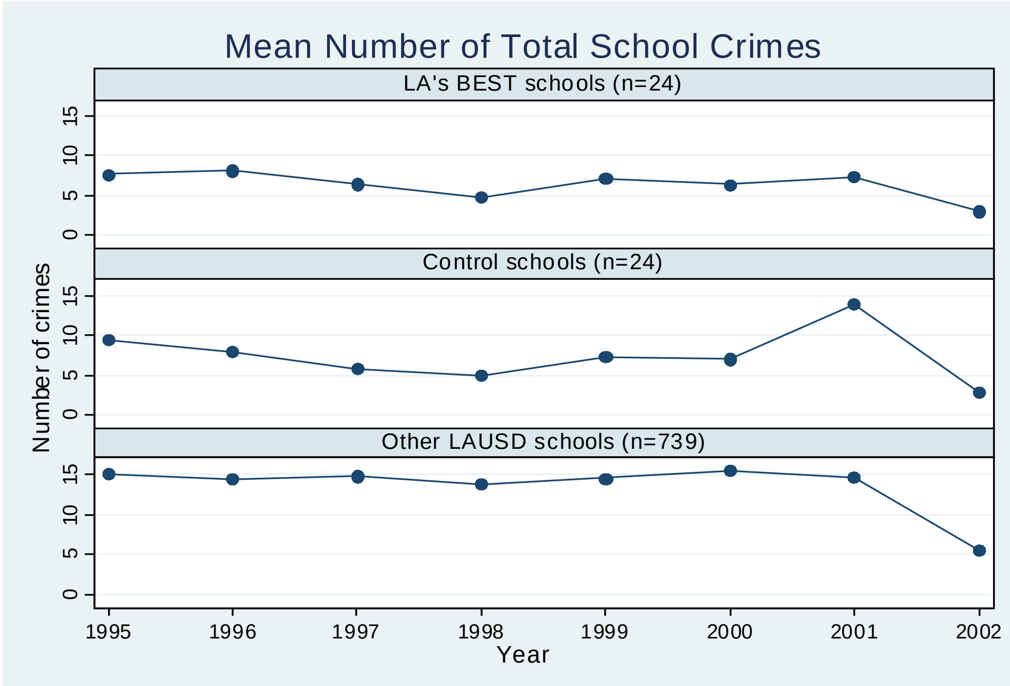


Figure 24

Note. Total school crimes include all the crime categories.

LA's BEST schools demonstrate a similar trend of data as the control schools, and exhibit lower school crimes than other LAUSD schools, although control schools experienced a sharp increase in crime in 2001, doubling that of LA's BEST. Moreover, LA's BEST schools demonstrate a relatively flatter rate of incidence over the eight-year period. Several literatures have cited that since the beginning of the operation of an afterschool program in the neighborhood, crime rates have declined (Newman, 2000).

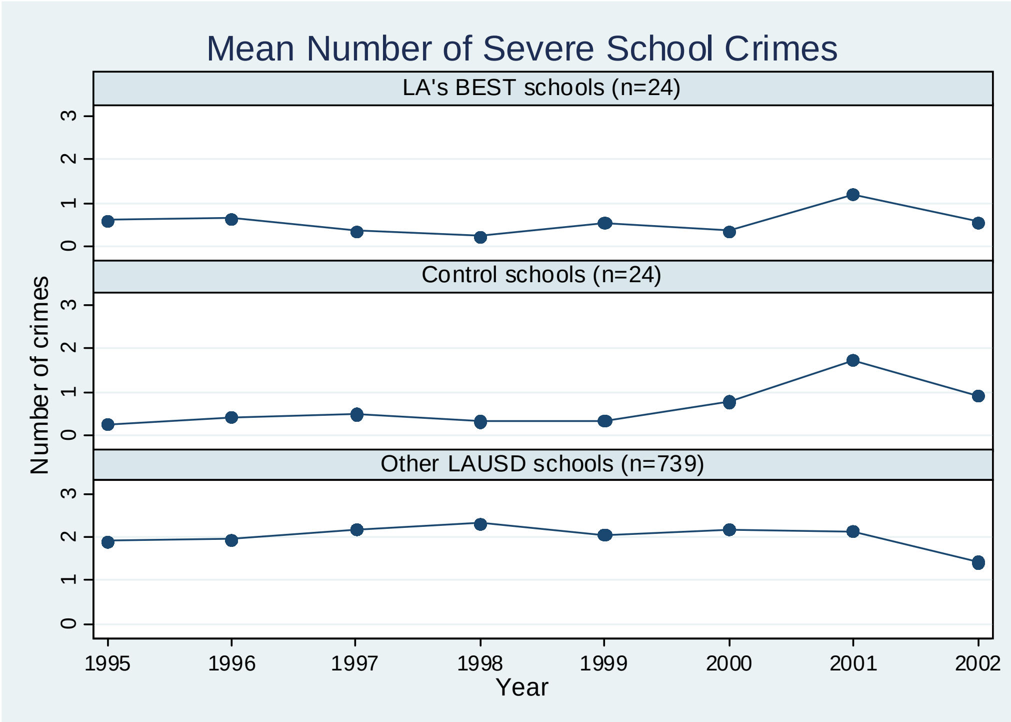


Figure 25

Note. Severe crimes include the following crime categories, adw (alcohol, drug or weapons), homicide, possession of weapons, robbery and sex offenses crimes.

LA's BEST schools had extremely similar results as control schools in the category of several school crimes, although the numbers were lower in the final two years. Both the intervention and control schools[37] had lower crimes rates than other LAUSD schools.

[37] Intervention and control schools included middle and high schools

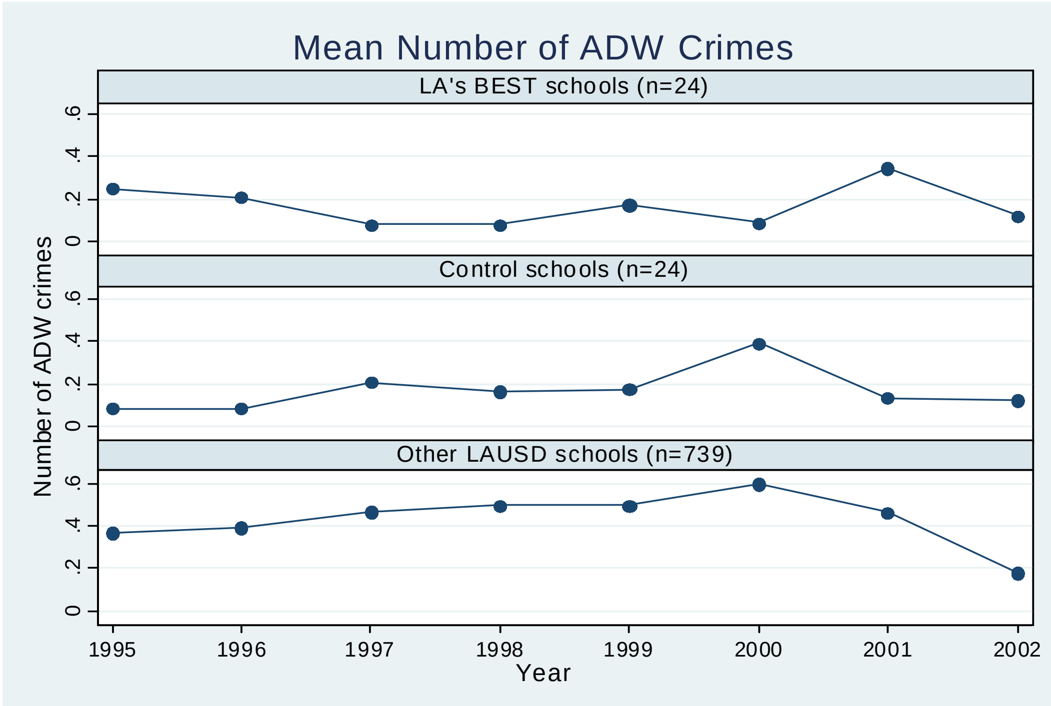


Figure 26

Note. ADW stands for alcohol, drug or weapon crimes.

LA's BEST schools had fewer alcohol, drug, and weapon-related crimes during 1997, 1998 and 2000, but equal or higher rates in other years. The trend data here diverges for the three categories and further analysis would be necessary to attempt to offer explanation for these anomalous findings.

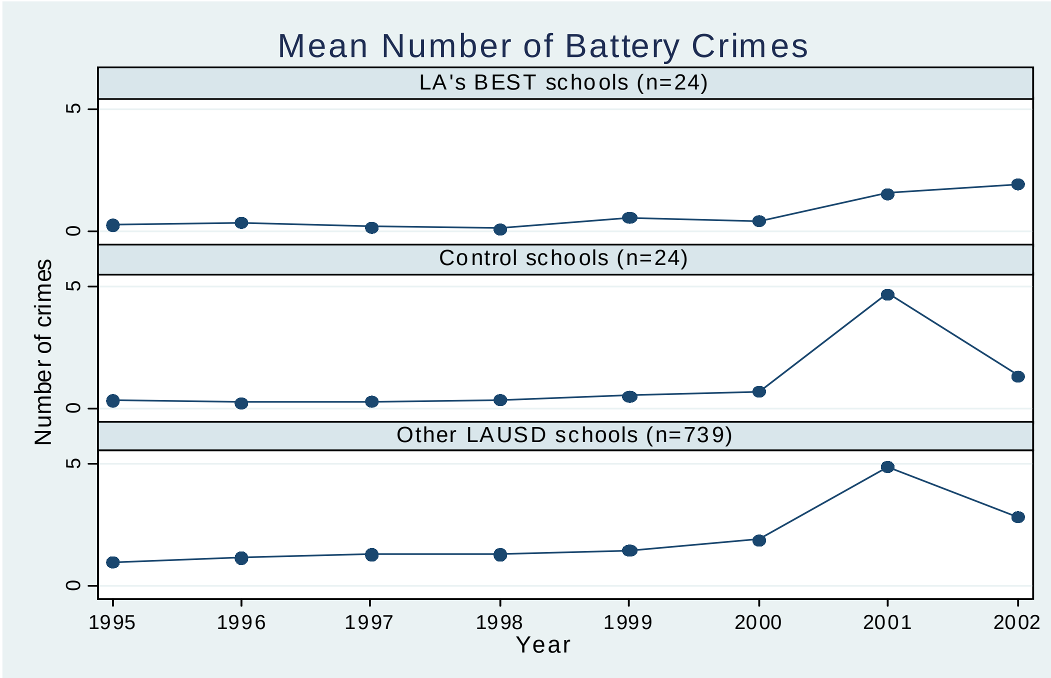


Figure 27

Battery crime data was again quite similar and very low for both LA's BEST and control schools, although control schools did experience an increase in crime in 2001, not experienced by LA's BEST. Both categories had lower battery crime incidents than the larger group of LAUSD schools.

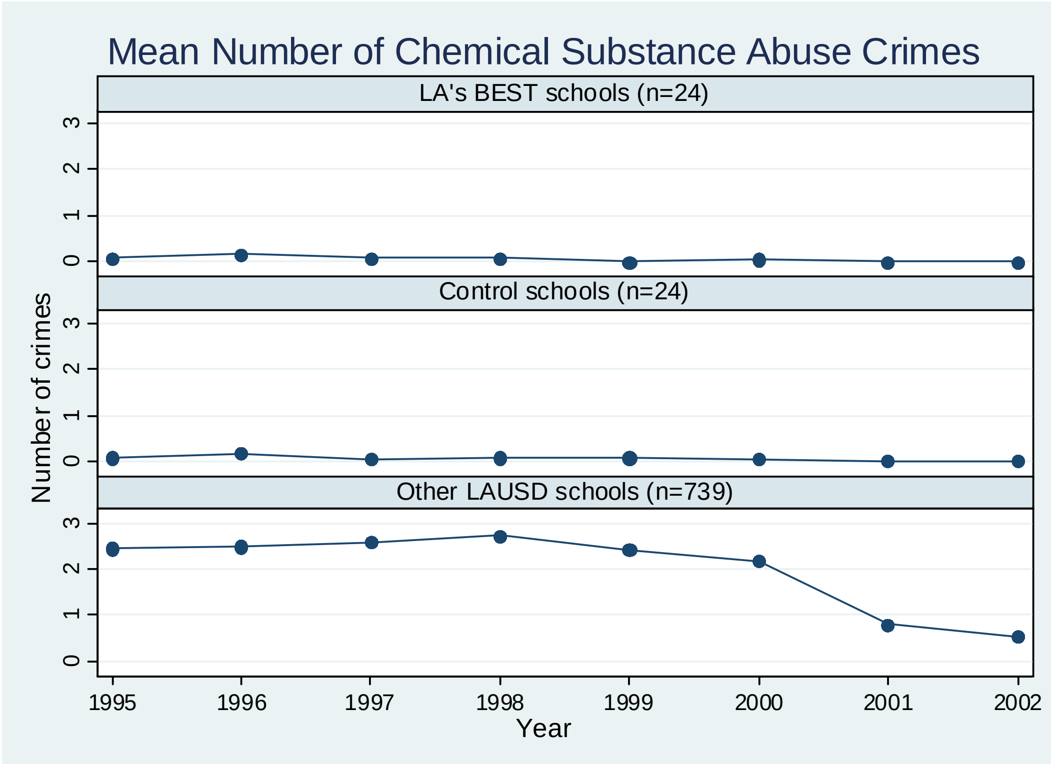


Figure 28

Chemical substance abuse crime rates are extremely low for LA's BEST and control schools, while higher for other LAUSD schools, though these figures are not surprising for elementary school students.

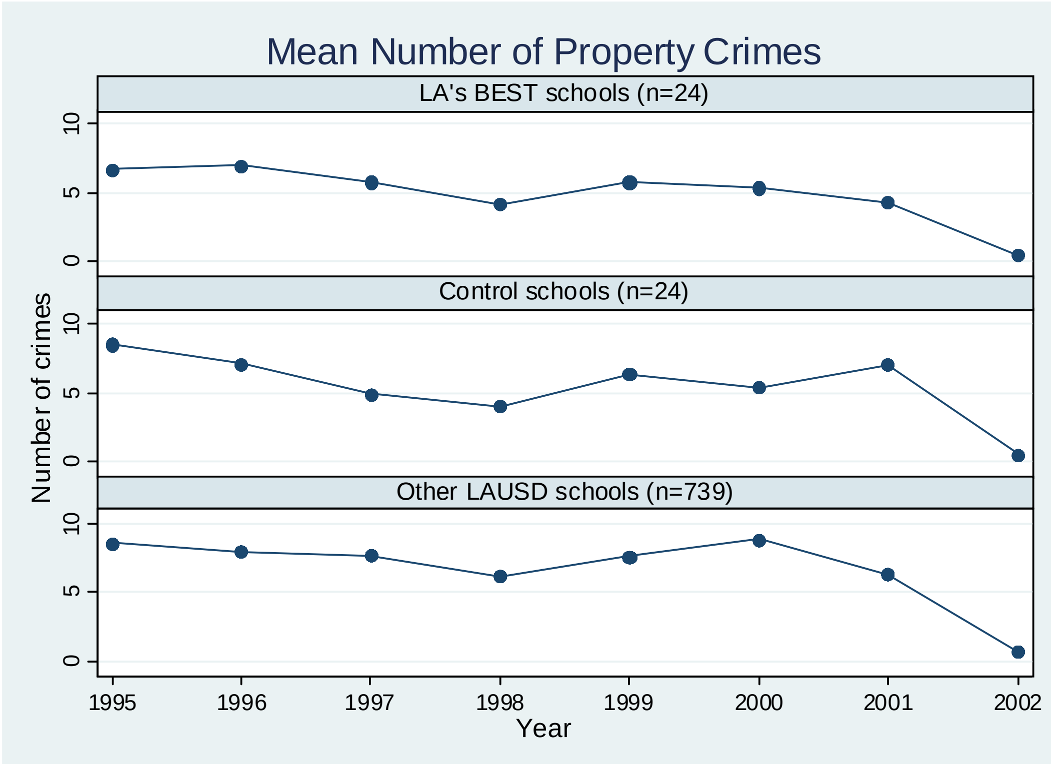


Figure 29

Property crime rates are also quite similar among the intervention and control schools, as well as other LAUSD schools. An interesting result here is the universal abatement of these crimes in 2002, which, however, may be attributable to an underreporting of crime.

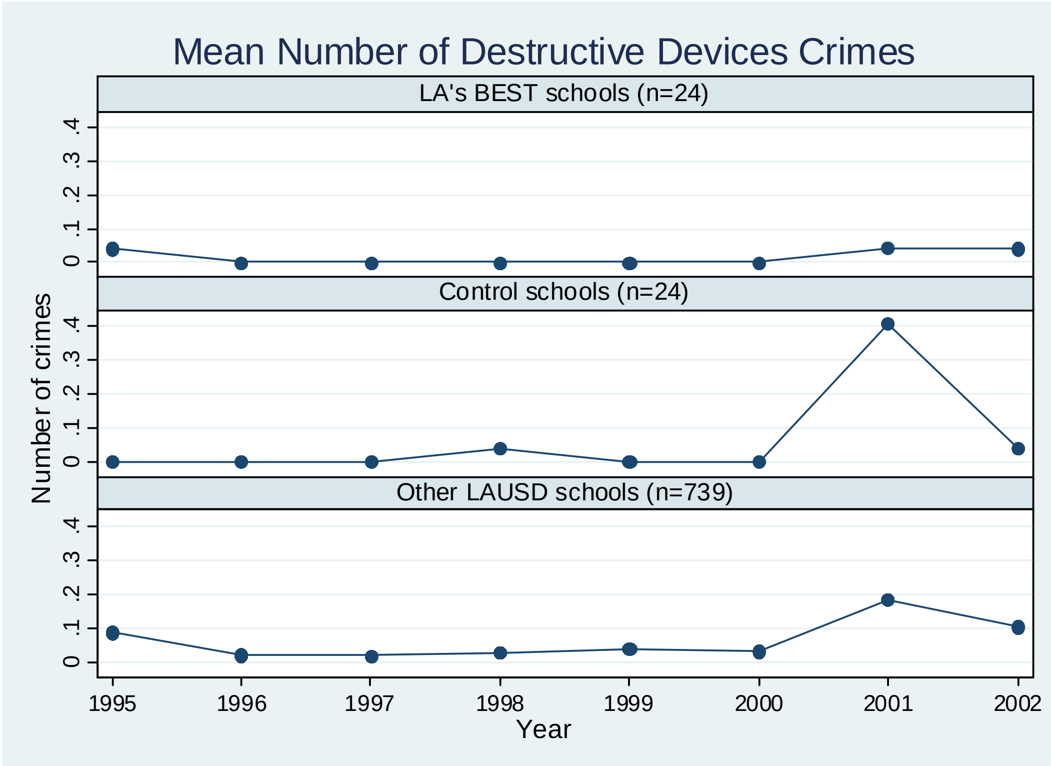


Figure 30

Destructive devices crimes are all but non-existent for LA's BEST and control schools, although control schools experienced a peak in 2001, as did other LAUSD schools.

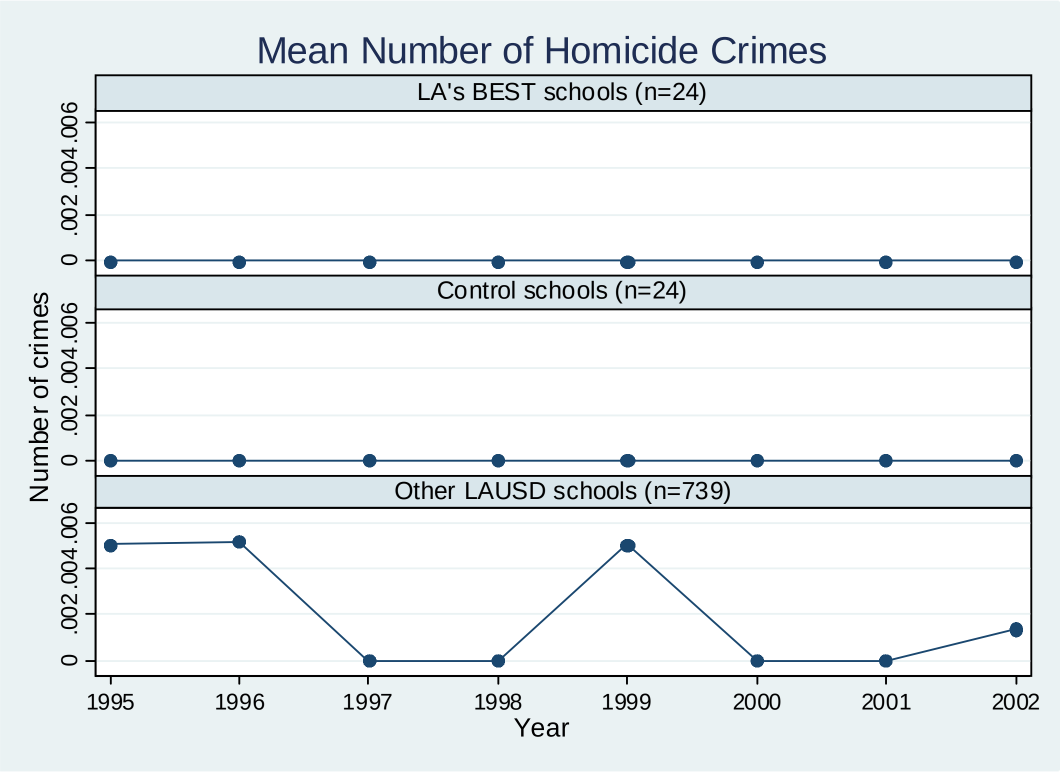


Figure 31

Homicide crimes are also non-existent among intervention and control schools with the larger LAUSD population showing increased incidents in some years, but no clear pattern. The variability could arguably be related to troubled youth that left the school after committing their crimes.

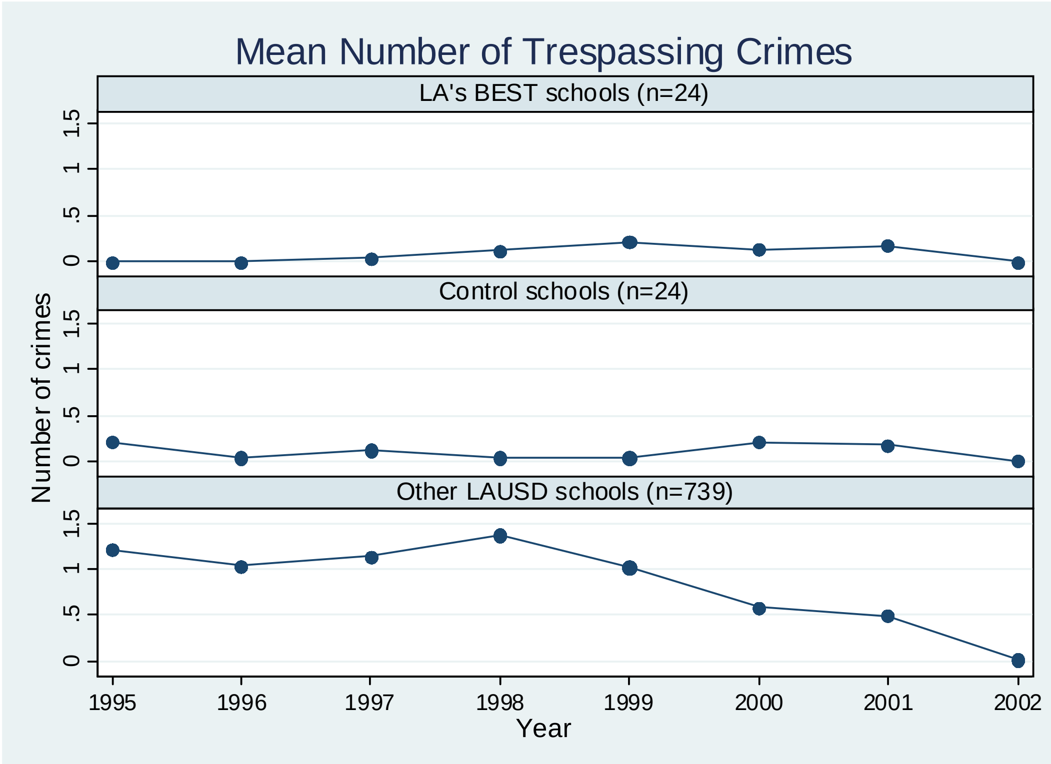


Figure 32

As with the other data, the mean number of trespassing crimes is low and flat over the eight-year period for the intervention and control schools, with other LAUSD school incidents trending downward starting in 1998.

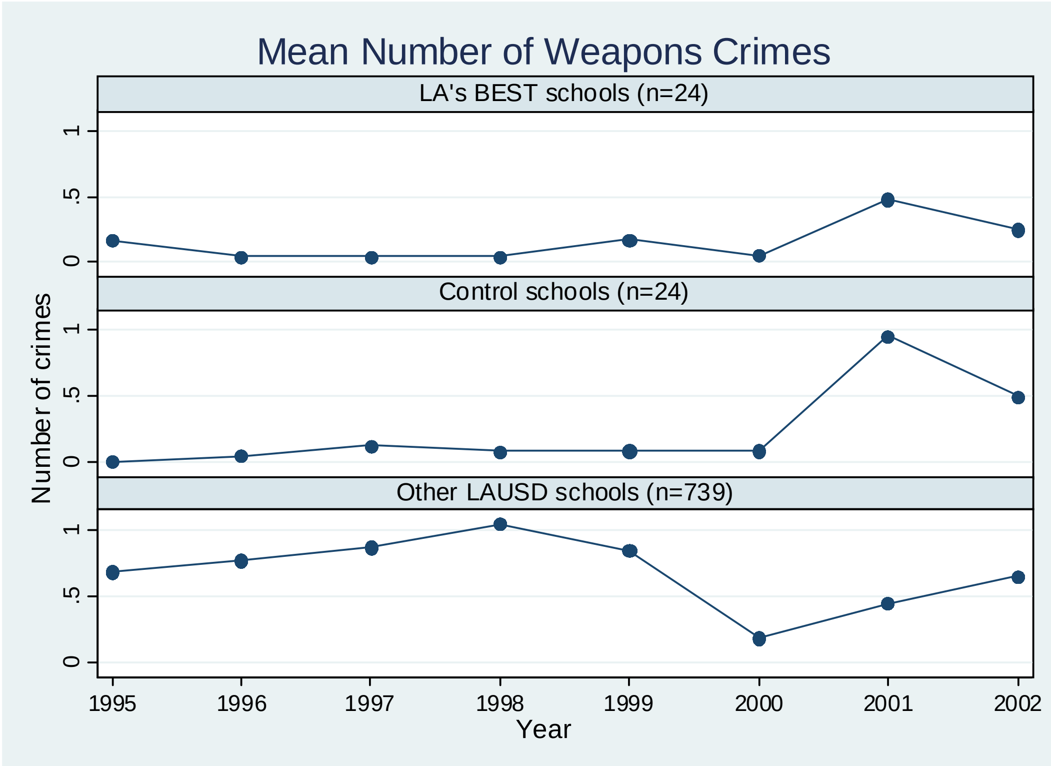


Figure 33

The mean number of weapons crimes is largely parallel among LA's BEST schools and the control schools; although LA's BEST schools have lower crime incident rates in 2001 and 2002. As expected, weapon crimes are substantially higher at other LAUSD schools, especially during the final two years of the period due to the older age of students and their access to fire arms.

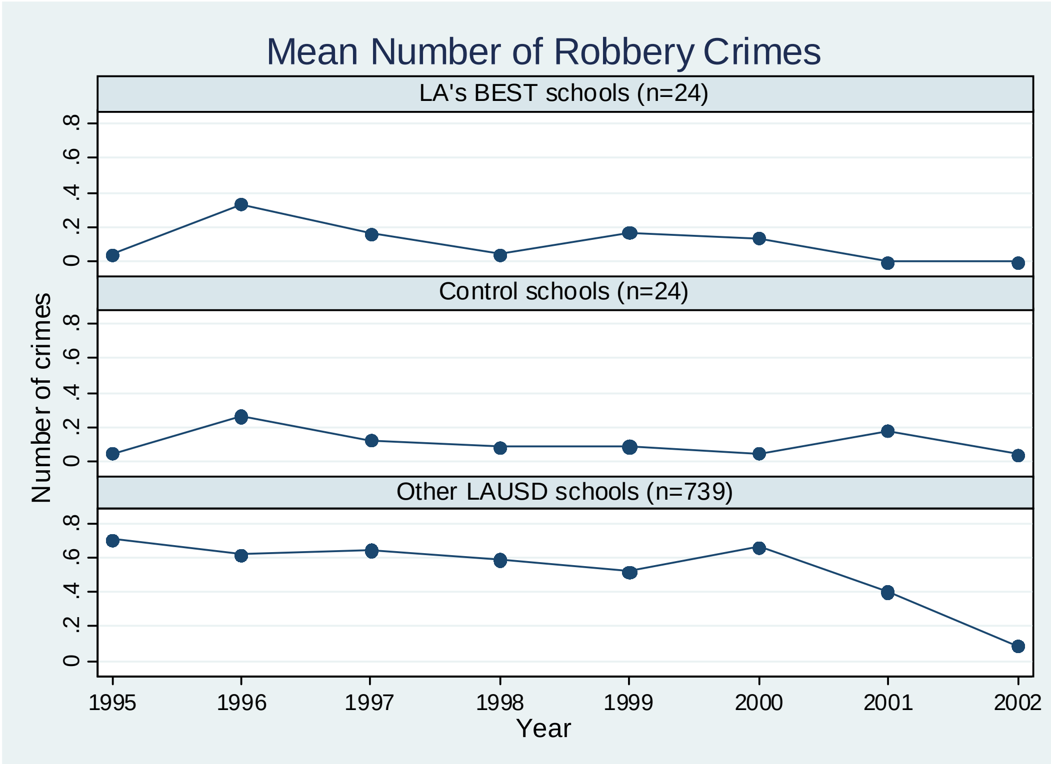


Figure 34

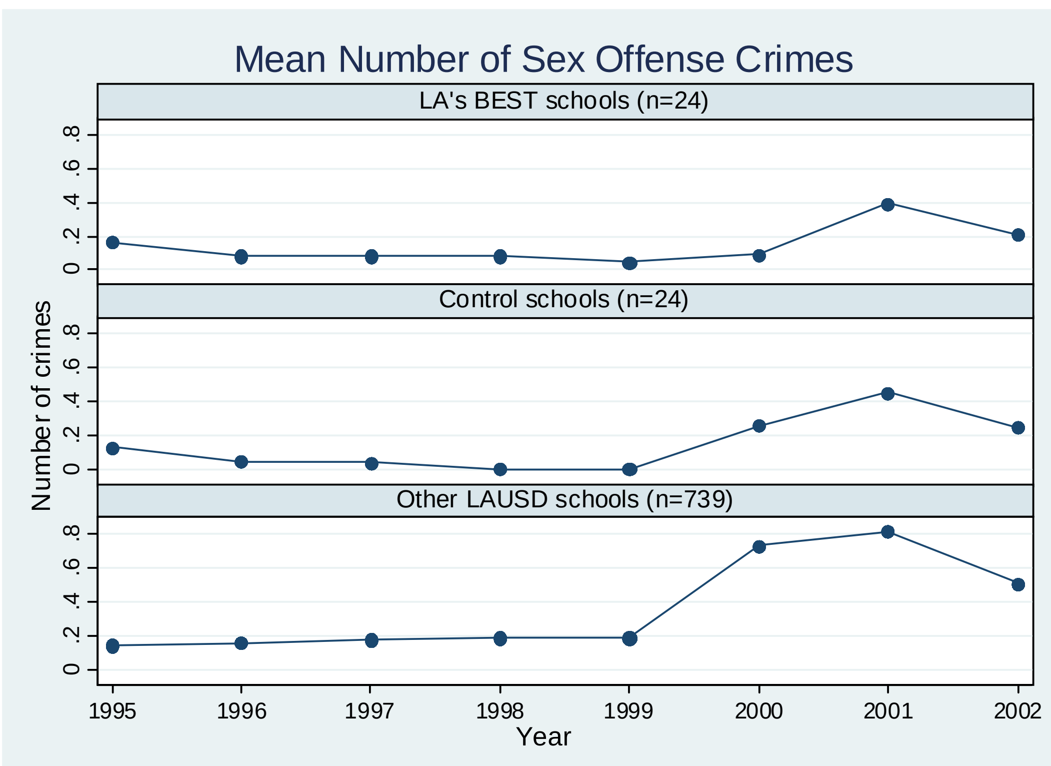


Figure 35

The Survival Analysis Results

Similar to the achievement models, we defined the time metric and reproduced the unconditional baseline hazard of committing a crime. Figure 36 displays the actual hazard that the model had to approximate.

Table 41

Base Hazard as Function of Time

Variable	Estimate	s.e.	Approx p-value	
Base rate (numerare)	-8.26	0.12	0.00	**
Annual change in rate	1.28	0.06	0.00	**
Quadratic effect of time	-0.10	0.01	0.00	**

$* p < .05$, $**p < .01$

There were several options for defining the time metric, but in order to balance a sufficiently fine-grained measure of time with an adequate number of events per time period, we used a yearly time metric. Figure 37 displays the unconditional hazard. The hazard displayed in Figure 37 is consistent with expectations; it displays an increasing hazard from elementary through early high school and a decreasing hazard from juvenile to adulthood. In order to accurately model this pattern we used linear and quadratic time indicators. The results of fitting the basic hazard model are displayed in Table 41. Consistent with the plotted hazard, we found both the linear and quadratic terms for time to be highly significant (p <.01). The results indicate that the maximum hazard is when students are in grades 9, 10, and 11.

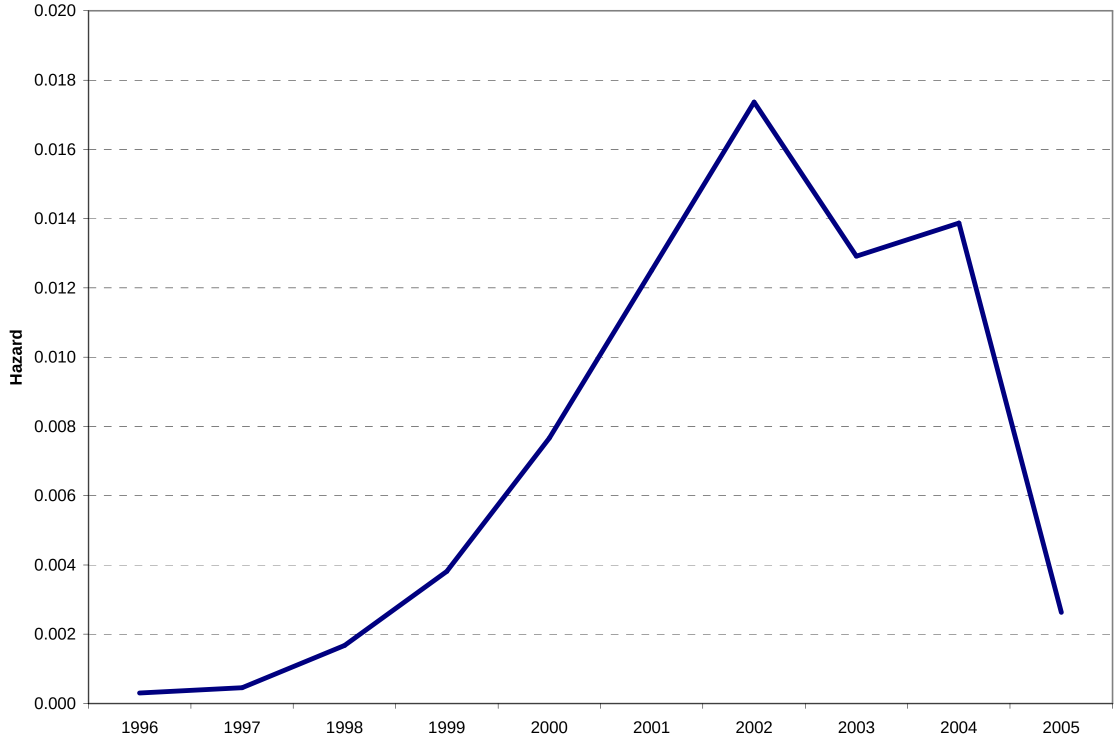


Figure 36. Actual hazard of juvenile crime over time

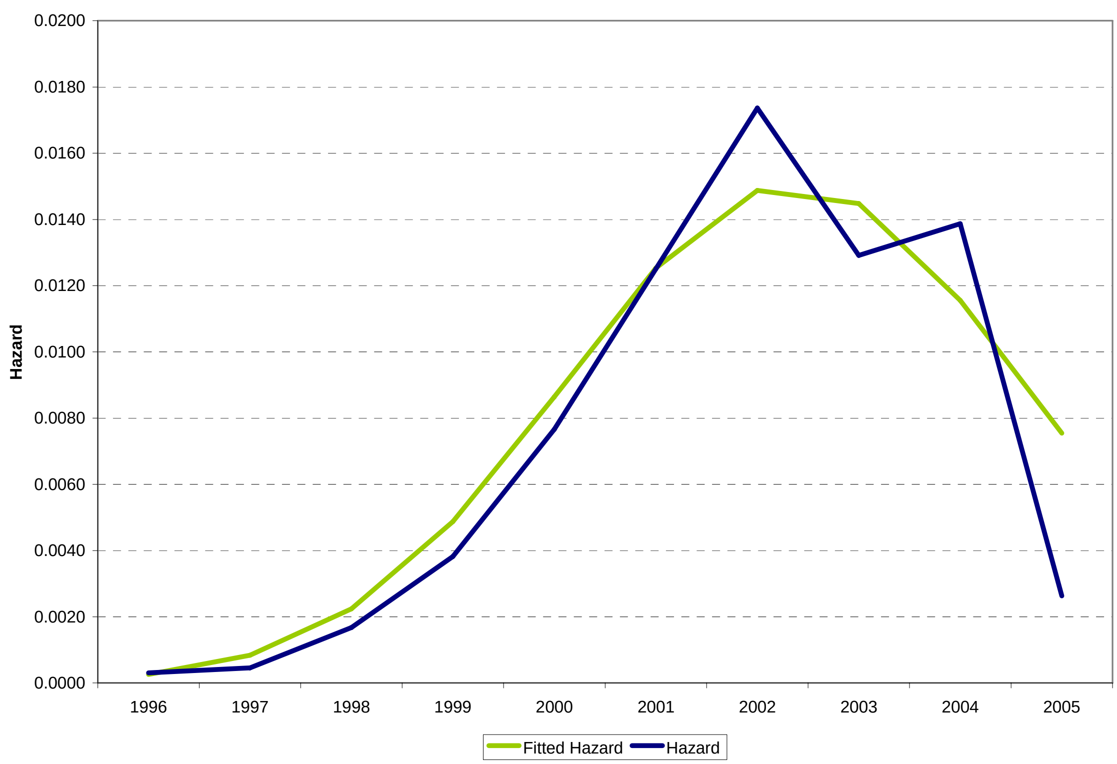


Figure 37. Fitted and actual Hazard of juvenile crime over time

Table 42 summarizes the multilevel survival analysis results. Preliminary analyses (not presented in this report) revealed that similar to achievement results, a simple dichotomous treatment indicator did not effectively capture the treatment effect. Consistent with the achievement analyses, we again examined exposure and engagement. Exposure was measured by years of attendance. Engagement was categorized into three levels: low engagement; medium engagement; and high engagement. Low engagement represented students who attended the after-school program between 4 and 9 days per month. Medium engagement represented students who attended between 10 and 14 days per month and high engagement consisted of students attending at least 15 days per month. The three coefficients of engagement were introduced simultaneously in the model; therefore, the reference group is students with "zero engagement".

It is important to reiterate that only those comparison students who were similar to the treatment individuals (based on the propensity score) were sampled and matched to treatment students. Those treatment students who did not find matched control students were eliminated from the sample. Therefore, each of the treatment students, in this model classified into three groups, has a matched control student in the comparison group. This condition allows the comparison between the treatment groups and the comparison students. However, comparisons across the treatment groups are more complicated because it is possible that bias may be introduced through self-selection. Student exposure and engagement in after-school programs may be a function of measured variables such as parental level of education. In addition, subsequent models introduce student characteristics to help address this issue.

Model 2 tested whether, unconditioned on concomitant variables, the treatment significantly impacts the probability that a student would commit a crime. The results indicate that student exposure has no marginal impact on the crime hazard once student engagement is taken into consideration. Model 2 results also indicate that students who were sporadic attendees (low engagement) did not benefit from the treatment. However, students who are engaged on a more consistent basis are significantly less likely to commit a crime. Students who are medium attendees are about 30% less likely to commit a crime ($p < .05$) and students who are high attendees are about half as likely to commit a crime ($p < .05$).

In order to more carefully isolate potential treatment effects we next examined the marginal impact of the treatment accounting for student characteristics. The results relate to Model 3. The treatment effects are quite robust with the inclusion of student background characteristics. The estimated treatment effects did not change substantively from Model 2 to Model 3. The effects of the concomitant student variables are generally consistent with expectations and we briefly summarize those next.

Table 42

	Model 2				Model 3				Model 4			
Variable	Estimate	s.e.	Aprox p-value		Estimate	s.e.	Aprox p-value		Estimate	s.e.	Aprox p-value	
Base rate (numerare)	-8.36	0.13	0.00	**	-7.80	0.37	0.00	**	-8.51	0.45	0.00	**
School Percent African American									0.04	0.01	0.00	**
School Percent Parents w/college									-0.15	0.06	0.02	*
LA's BEST school									0.02	0.02	0.30	
Later becomes LA's BEST school									0.23	0.16	0.15	
School's zipcode % HH in poverty									-0.33	0.11	0.00	**
Annual change in rate	1.29	0.06	0.00	**	1.39	0.07	0.00	**	1.58	0.09	0.00	**
School Percent African American									-0.01	0.00	0.00	**
School Percent Parents w/college									0.03	0.01	0.01	*
Later becomes LA's BEST school									-0.01	0.03	0.73	
School's zipcode % HH in poverty									0.07	0.02	0.00	**
Quadtratic effct of time	-0.10	0.01	0.00	**	-0.11	0.01	0.00	**	-0.13	0.01	0.00	**
Effect of low engagement	0.19	0.14	0.19		0.13	0.15	0.38		0.04	0.16	0.81	
School's zipcode % HH in poverty									-0.06	0.03	0.05	*
Effect of medium engagment	-0.36	0.14	0.01	**	-0.38	0.15	0.01	*	-0.38	0.15	0.01	*
School's zipcode % HH in poverty									-0.04	0.06	0.57	
Effect of high engagement	-0.66	0.23	0.00	**	-0.59	0.24	0.02	*	-0.60	0.25	0.02	*
School's zipcode % HH in poverty									-0.01	0.10	0.94	
Girls vs boys					-1.02	0.09	0.00	**	-1.02	0.09	0.00	**
Hispanics vs. Whites & other					-0.81	0.31	0.01	**	-0.81	0.34	0.02	*
African Americans vs. Whites & other					0.05	0.34	0.89		0.08	0.38	0.82	
Asian vs. Whites & other					-2.00	0.84	0.02	*	-2.03	0.88	0.02	*
SWD vs non-SWD					0.26	0.11	0.01	*	0.26	0.11	0.02	*
Parent Educ college vs less					-0.24	0.13	0.07		-0.26	0.14	0.06	
Years of Exposure	0.11	0.07	0.10		0.14	0.07	0.06		0.12	0.08	0.10	
Years of ELL					0.03	0.01	0.01	**	0.03	0.01	0.01	*

Consistent with expectations, the results in Model 3 indicate that girls are significantly less likely to commit a crime ($p < .01$). In fact, boys are about three times as likely to commit a crime as girls. The results for race/ethnicity are interesting in that the predicted marginal probabilities of committing crime are somewhat counter-intuitive at first glance. Consistent with expectations, Asians are predicted to commit crimes at a significantly lower rate than White students ($p < .01$), ceteris paribus. Hispanics are also estimated to be less likely to commit crimes than their White classmates ($p < .05$). African American students are estimated to commit crimes at about the same rate as their White classmates, all else being equal. It is important to bear in mind that African American students have a greater unconditional crime rate than their White classmates, but after controlling for concomitant factors, the rates are virtually identical. Accounting for the other student characteristics in the model, students with disabilities are estimated to commit crimes about 30% more often than their non-disabled classmates.

Another key aspect of Model 3 is the inclusion of a proxy for student SES. We use only parent education as the other common indicator for SES status, because FRL (Free and Reduced price Lunch) eligibility represents about 94% of the sample and does not differentiate students. In addition, students of college educated parents are 25% less likely to commit crimes than students whose parents do not have a college education.

Preliminary analyses tested several subsets of interactions as recommended by Singer and Willett (2003). We test whether there are any treatment-by-time effects. That is, we analyze whether the effect of LA's BEST wanes over time. Of course the effect of LA's BEST on social outcomes, such as juvenile crime, would have been negligible during the treatment period because the hazard in elementary is very low. No interaction effects are evident. However, the results in Figure 40[38] demonstrate the marginal effects of LA's BEST on the juvenile crime hazard. The most discernable impact is during the peak hazard years. This effect, while greatest during the peak hazard years, does have a significant effect on the survival probability. This can most readily be seen by the survival curves displayed in Figure 38. The survival curves demonstrate the cumulative effect of low period hazard rates. By the end of the analyses period the low engagement treatment group and the control group are equally likely to have committed a crime. We would have expected about 13%[39] of these two

[38] We plot the hazard functions and survival possibilities for the numeraire – this does not affect treatment effect interpretations.
[39] This estimate is for the numeraire

groups to have committed a crime. Yet crime rates decrease as engagement increases. Figure 38 highlights the lack of benefit to LA's BEST students, upon sporadic attendance. Benefits, however, increase, when engagement and attendance increase.

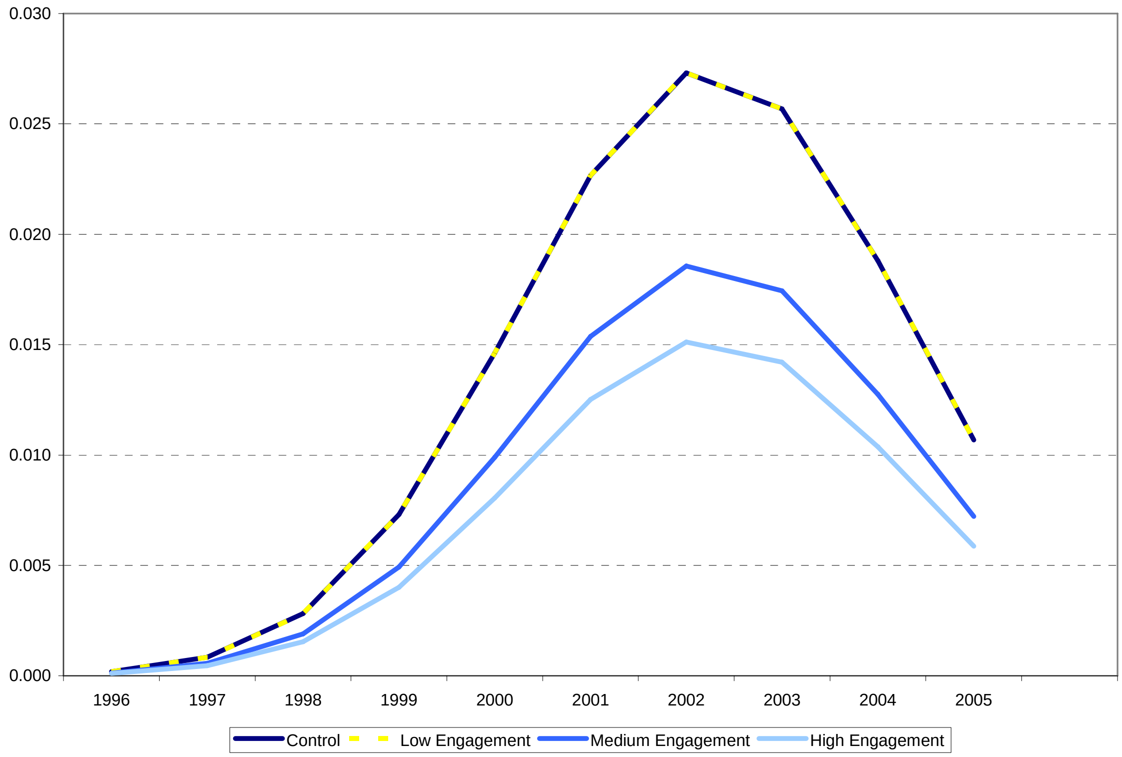


Figure 38. Hazard functions for treatment and control groups

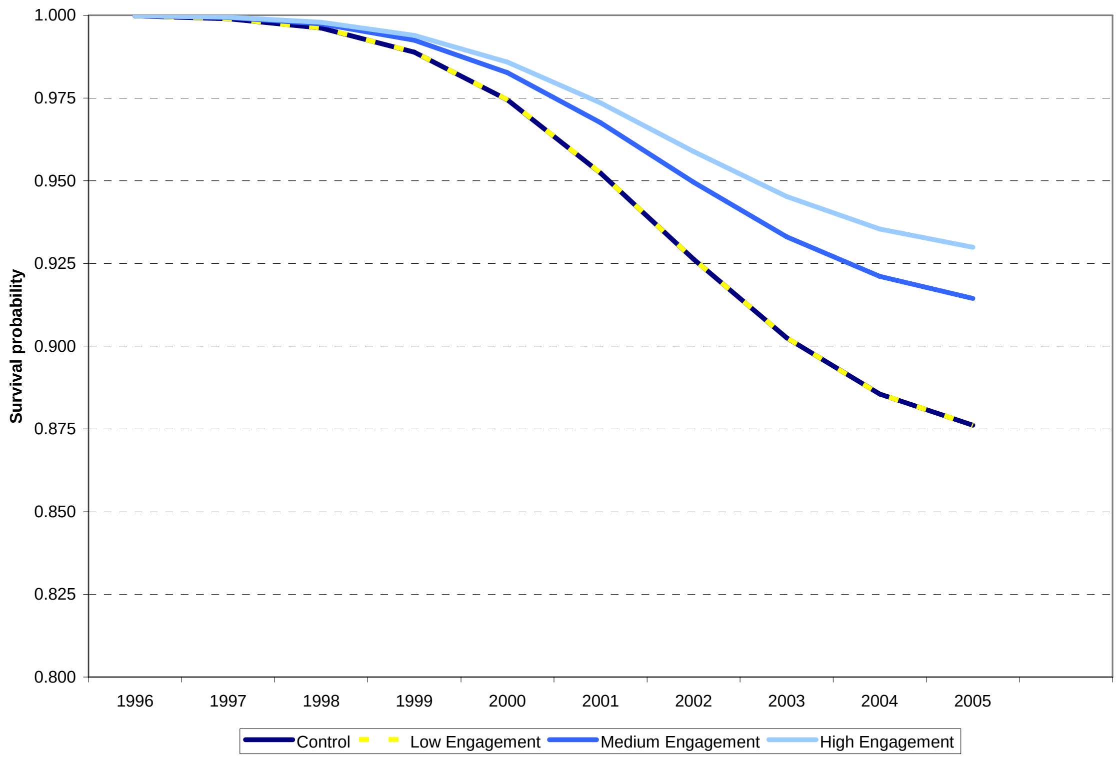


Figure 39. Survival probabilities for treatment and control groups

The survival curves in Figure 39 also highlight that students who are consistently engaged in the after-school program derive marginal benefits from the treatment. The results in Figure 39 clearly demonstrate the cumulative benefit of the treatment. By the end of the study period we would have expected about 9% of the medium engagement students and about 7% of the highly engaged students to have committed a crime.

Although the effect of exposure is not statistically significant at the 5% level, we need to carefully consider the impact of exposure. The model results suggest that the number of years a student attends LA's BEST is irrelevant and implies that as long as a student is engaged with the program for at least a year, benefits accrue. The benefit of engagement may then be as much a function of average weekly attendance (engagement) as it is the cumulative effect of attendance (intensity[40]).

We test this parameterization of intensity and found it is not significant. Still, plotting the marginal hazard functions, we assume all else is equal, overstating the potential effects of engagement because students who attend necessarily had exposure. Although we cannot reject the null hypotheses that the effect of duration is 0, we none-the-less must plot the survival curves as above and include the effect of exposure at the

[40] define intensity as the total number of days attended

average exposure (i.e., years of attendance) as a way to examine the sensitivity of treatment effects. The survival curves reveals that students with medium or

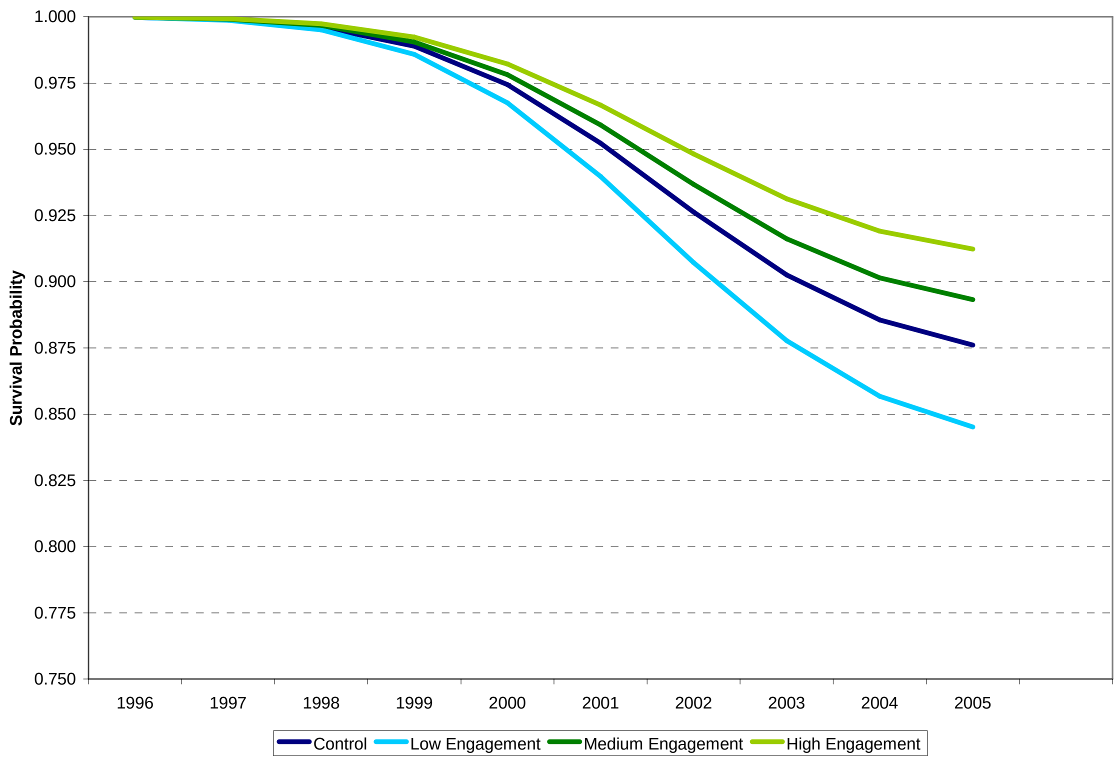


Figure 40. Survival Probabilities including the effect of exposure

high engagement are still expected to commit a crime at a substantively lower rate than students in the non-treatment, sporadic attendees group. Because the parameter estimate for exposure is positive, it reduces the effect of the treatment conditions. Given that the low engagement students do not have a reduced likelihood of committing crime, the effect of exposure causes the sporadically engaged students to have a predicted survival curve substantially below the non-treatment group and the other treatment conditions. The marginal effects of the medium and high treatment conditions demonstrate substantively meaningful increases on survival. The cumulative difference between the medium and high engagement groups and the control group is 1.8% and 3.7%, respectively. This reduction is associated with a 14% and a 29% increase in survival for the medium and high engagement groups, respectively.

Another set of interactions we tested are interactions between-student concomitant variables and treatment conditions. For example, it could be the case that low SES students benefit more from the LA's BEST program than non-low SES students. Preliminary analyses reveal that there are no significant interactions between treatment conditions and student covariates.

We next take advantage of the nested nature of the data and examine the potential between-school and neighborhood effects that potentially mediate the hazard functions. Figure 41 highlights the variation among schools in base hazard rates as well as the annual year effect on the hazard function. The axes in Figure 41 are in logits, which must be transformed in order to be meaningful. In order to place Figure 41 into context we note that as we construct hazard rates from the base and year pairs[41] the hazard functions tend to shift to the left as we use pairs from left to right. Figure 42 highlights this phenomenon.

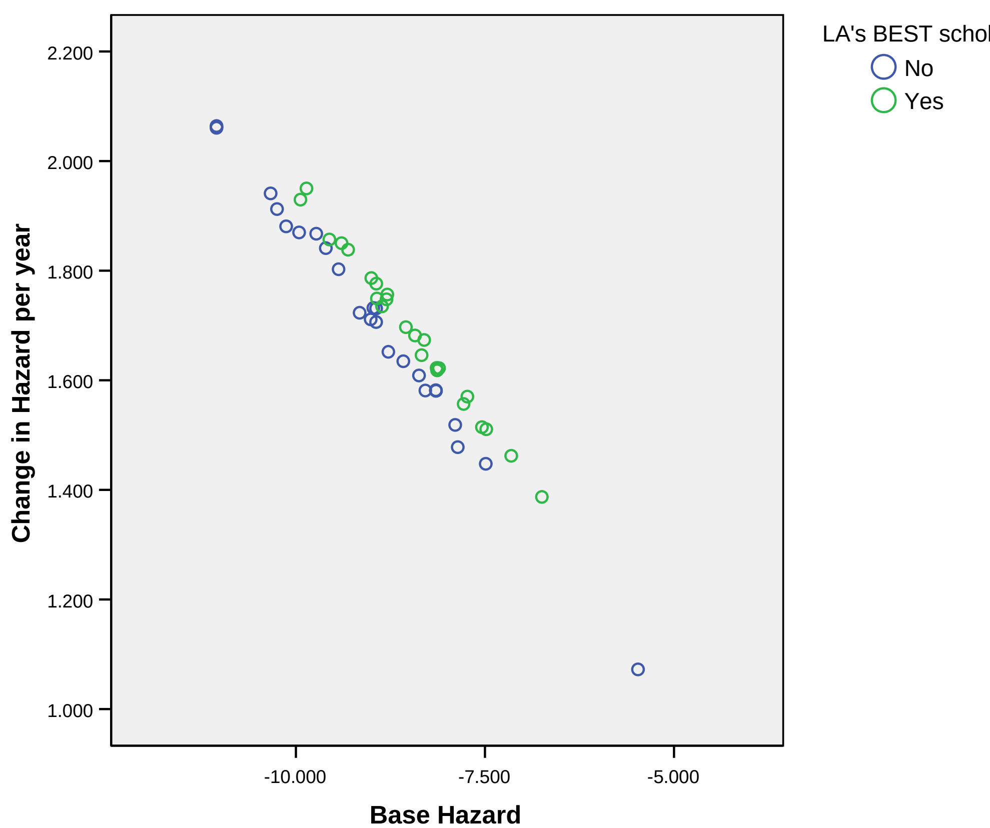


Figure 41. Estimated base Hazards and annual hazard effect by school

[41] The year cubed parameter is fixed across schools.

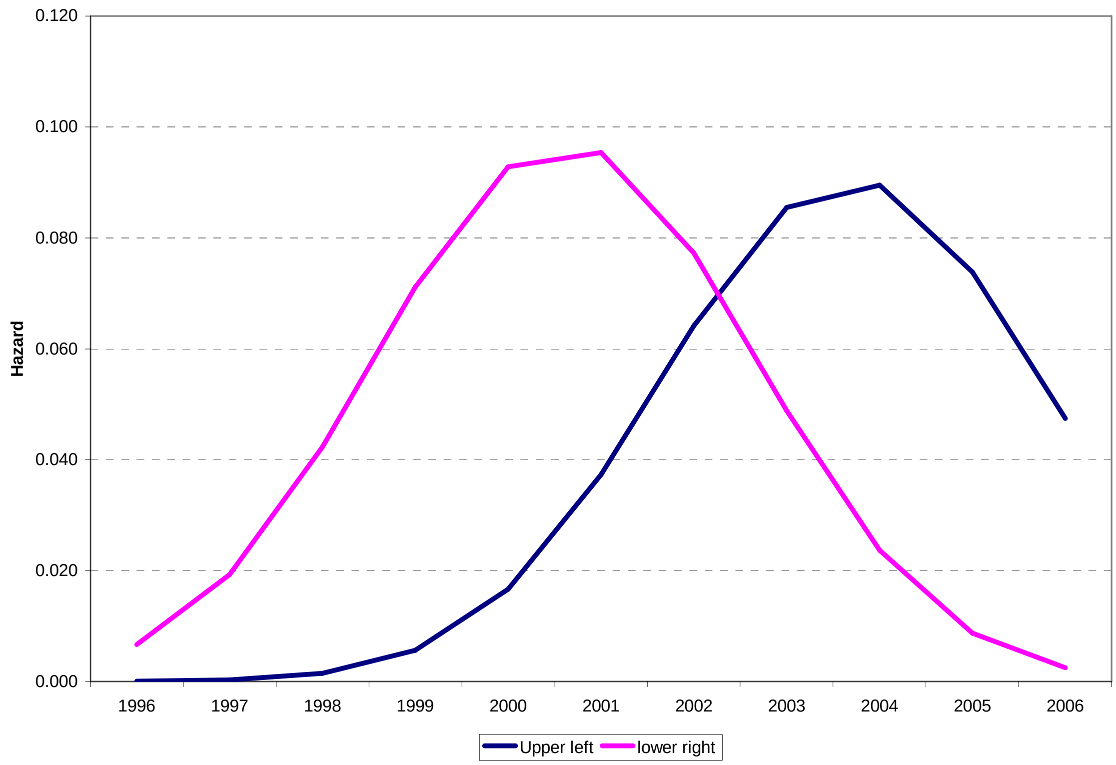


Figure 42. Two school specific hazard functions

This indicates that at each school, there is a different dynamic associated with student crimes; it is not accounted for by whether or not a school participates in the LA's BEST program. Given the multilevel propensity scores used to match students and schools, we would not expect substantive differences merely due to being an LA's BEST school.

The results of Model 4 use the same set of treatment indicators and individual student characteristics as in Model 3. The variables carried over from Model 3 remain consistent in the expanded specification presented in Model 4. As Figure 41 highlights, LA's BEST schools have somewhat higher base rates (though not at the 5% significance level) and not significantly different year effects. This implies that the treatment effect we observe for students is not due to school level effects associated with LA's BEST systematically selecting schools. We also include a school level test of whether schools that later become part of LA's BEST schools are systematically different than schools that are not part of LA's BEST schools. Contrary to expectations, schools that later become a part of LA's BEST schools have a lower base hazard than schools that do not become apart of LA's BEST schools.

We also consider the potential impact of school context. Schools with a higher percentage of minority students as well as parents with less than a college education have systematically higher crime hazards. After accounting for individual student characteristics, treatment conditions, and other school context indicators, there is a substantive effect of neighborhood poverty on juvenile crime. The results in Model 4 indicate that although the average effect of LA's BEST on students who attended sporadically (low engagement) is 0, this effect is mediated by neighborhood poverty. Consistent with expectations, the results imply that survival probabilities are lower in high poverty neighborhoods; yet the results also imply that poverty has an inverse relationship with the estimated effect of the treatment for the low engagement group. This effect can be seen in Figure 45. The solid lines represent low poverty neighborhoods and the dashed lines represent high poverty neighborhoods. The difference in survival probabilities between the low poverty, low engagement treatment and control groups is minimal. However, the difference in survival probabilities between the high poverty, low engagement treatment, and control groups is substantively large – approximately 12 percentage points. Control group students in high poverty neighborhoods are substantially less likely to survive without committing a crime.

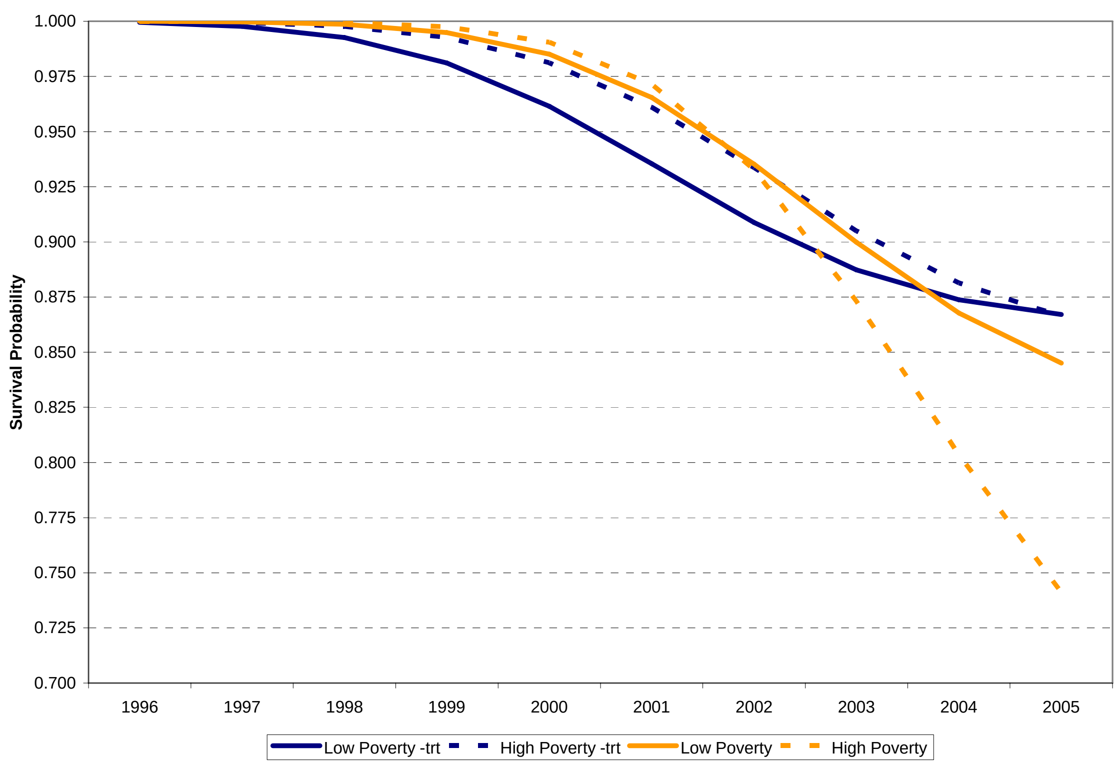


Figure 43. Effect of neighborhood poverty and low treatment engagement on survival probability

We next present single level survival models that were suggestive of the different patterns that existed between felonies and misdemeanors as outcomes, when modeled separately. We summarize the salient results in Figures 44 and 45. Overall, the concomitant variables have the same relationship to the outcomes as they did in the multilevel survival model using any crime as an outcome[42]. In general, the hazard functions for the medium and high treatment conditions are (statistically) significantly different from the control group hazard function when we use felony as the outcome. Further, the low engagement condition tends to be similar to the control condition, although the parameter estimates make the hazard functions in the figures look substantively different, they are not due to large standard errors of the estimate.

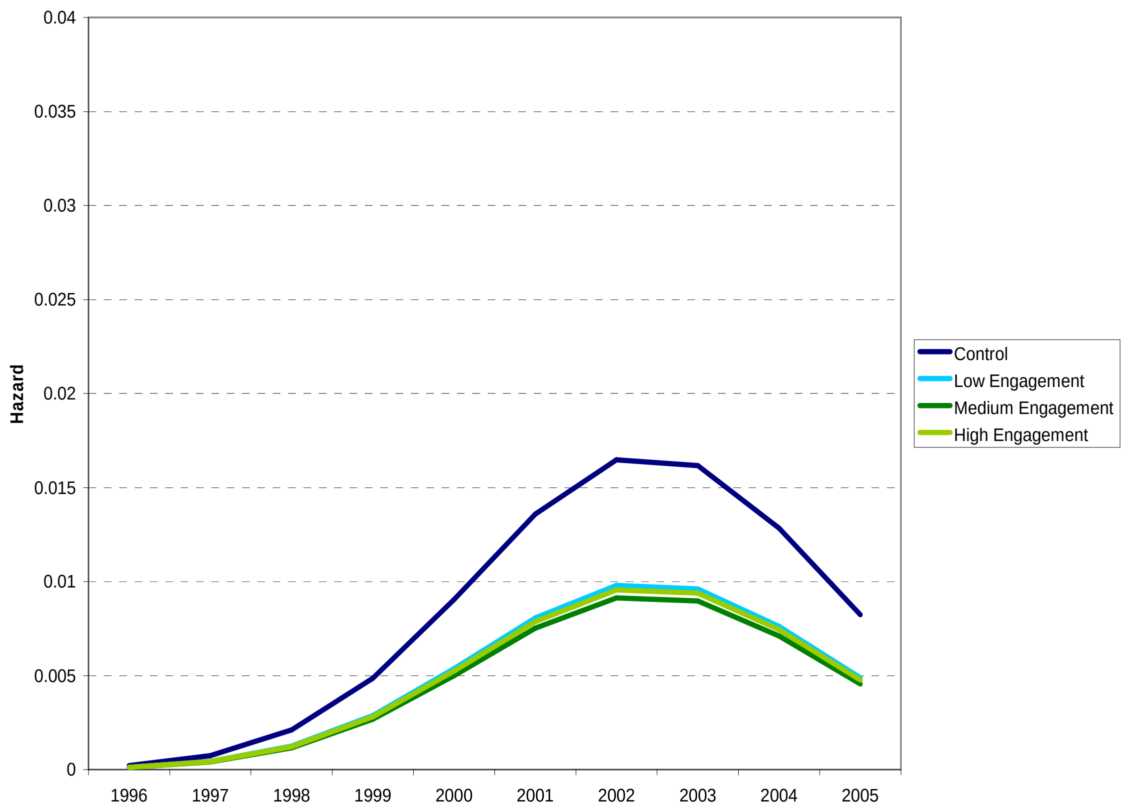


Figure 44. Fitted hazard for felonies

[42] The single level survival model using any crime as an outcome yielded results that were virtually identical to the multilevel survival model – however, single level models do not allow us to adequately test between-school and neighborhood effects on juvenile crime.

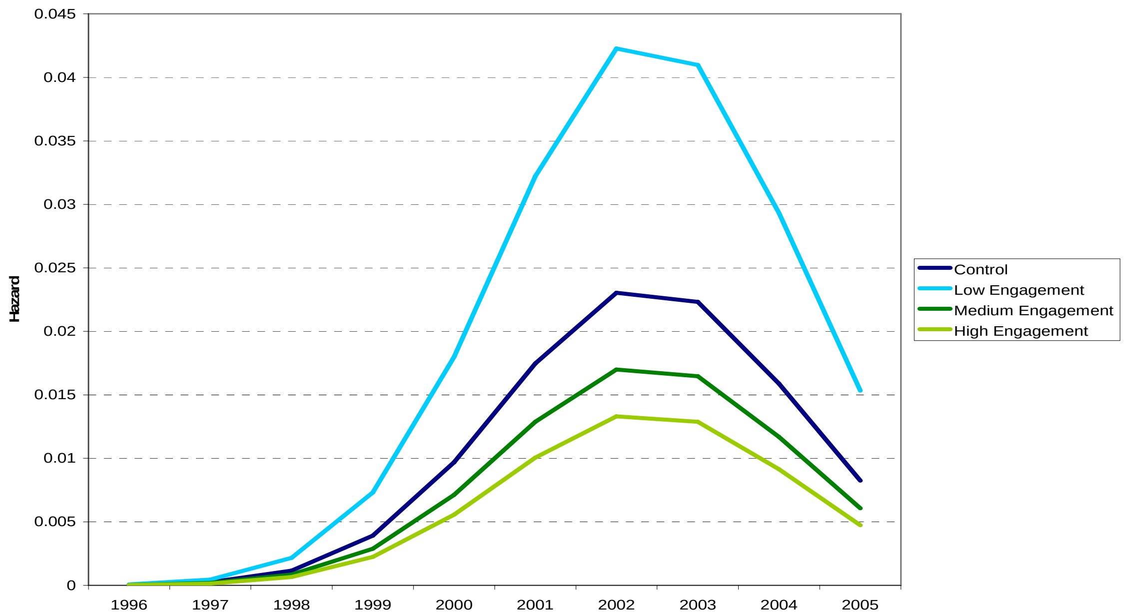


Figure 45. Fitted hazard for misdemeanors

The results for misdemeanors appear suggestive of treatment effects, but due to large standard errors, none of the treatment conditions' hazard functions differ significantly from the control group's hazard. The pattern is consistent with the pattern for felonies.

Summary of Juvenile Crime Results. The results from the multilevel survival analyses indicate that LA's BEST positively impacts juvenile crime probabilities. This is not the result of differential crime hazards between LA's BEST and non-LA's BEST schools, but is related directly to individual participation in the program. We find that students who are actively and intensely engaged benefit most from LA's BEST, while those who are moderately engaged also benefit. In general, we find that students who only sporadically attend do not benefit from the program unless we consider mediating circumstances.

In this case, an important mediating factor is the percentage of households (per neighborhood population) living below the poverty threshold. Our model implies that the treatment have some potential positive reduction in crime hazards in high poverty neighborhoods, which is arguably where LA's BEST focuses its attention. However, it is important to note that we are modeling this effect among a non-random sample of neighborhoods. This means that we cannot be certain whether this effect would be visible if we used the entire spectrum of neighborhood poverty across large districts or states.

Benefit-Cost Results

Using the data and findings previously discussed in this report, we provide a benefit-cost analysis of LA's BEST as it relates to crime abatement. We follow the example of previous benefit-cost studies and the specific guidelines set forth in Levin (1983). To start, we determine who the stakeholders are and from whose perspective we would conduct the benefit-cost analysis (Levin, 1983; Cohen, 2000). Generally, the stakeholders are participants (private costs and benefits), tax payers, victims, and funding agencies (social costs).

The general idea of a benefit-cost analysis is to determine whether the present value of benefits accrue to program participants and society at large are greater than the program's costs. Mathematically, the benefits outweigh the costs if the ratio of benefits to costs is greater than 1. The ratio compares benefits and costs in constant dollar terms and discounts cost and benefit streams to a single present value amount.

We follow the lead of previous evaluations in using cost estimates associated with specific crimes and juvenile court costs provided by Cohen and his collaborators (2000, 1998). Although we are not gathering information from specific victims, estimates of tangible costs to victims for specific types of crimes have been used and are based on the National Crime Victimization Survey (NCVS), published by the Bureau of Justice Statistics.

Benefit-Cost Analysis

In this section we present the benefit-cost analysis of the LA's BEST effects on juvenile crime. We use the results developed in the previous section as the basis for developing various benefit-cost ratios. We present these ratios because they simplify the comparison of both cost and benefit streams that occur over time. We explicitly examine benefits, which in this case are derived from avoided costs and costs that potentially occurred over the entire study period. The benefit estimates are based on Cohen (1998) and subsequent research that has updated those figures[43]. We develop

[43] The costs of crime were obtained from: http://www.neshaminy.k12.pa.us/attendance_policy/cost_juvenile_crime.htm. Estimates based on Cohen, (1998 and 1995) prevailed through the literature in terms of the costs of crime. The lifelong criminal estimates presented by Cohen assumed a youth continues to commit crimes as an adult. Often the literature presumes that juveniles commit 1-4 crimes between the ages of about 11 and 18 (although the evidence for this is not consistent). In our sample, 459 students (7.8%) committed 990 offenses, or about 2.16 offenses per offender.

cost estimates based on actual program costs. We use 1998 (the end of the treatment period) dollars for both costs and benefits, with benefit estimates displayed in the present value of 1998 dollars. We use the CPI to adjust 1994 annual program costs to 1998 dollars.

Given the complex nature of the results, we present three sets of benefit-cost ratios: based on a single year's participation, based on the average exposure (i.e., years attended) of the sample, and based on each year of exposure separately. Each scenario presents ratios for low, high, and lifetime crime estimates. The low and high estimates are based on juvenile crime costs only, while the lifetime crime estimate is based on the assumption of adult crime[44]. Previous research indicates that between 4% and 16% of juveniles continue committing crimes after becoming adults (Garcia, 1990).

Costs

Program cost estimates are based on actual incurred costs as well as opportunity costs associated with adult volunteers assisting LA's BEST programs. The costs are presented in Table 42. It is important to reiterate that the we consider students to be participating in LA's BEST only if they attended a minimum of 36 days per school year, or about once per week. Only students meeting this criterion are counted in the treatment group. Clearly the per -student costs would be significantly lower if we included students with no minimum attendance threshold. Per-student program costs would be approximately $351. However, in order for the costs to be in accordance with assumptions used to derive benefits, we use the attendance requirement that yields a per-student cost of about $568 (in 1998 dollars). The program financial costs are detailed in the Appendix A. The per-student cost we use also include the per-student cost of volunteers based on the hourly compensation of LA's BEST field staff. We do not include facilities or start-up costs. These costs are consistent with per-student cost estimates of other after-school programs. We also exclude participant opportunity costs, as well as incremental costs associated with parents picking up their children as we assume that the marginal cost is 0, given parents would likely pick up their children after school with or without program participation. The $568 amount becomes the

[44] We note that much of the literature assumes lifetime crime for all juvenile offenders. However, this is not consistent with studies that indicate that juvenile crime peaks and that there is not a 100% recidivism rate.

annual figure used in the denominator of the benefit-cost ratio. Scenarios two and three adjust this amount as detailed within each scenario.

Table 43

Annual Costs Associated with LA's BEST After-School Program

Costs	FY 1994 $	in 1998 $
Direct costs	1,774,680	1,951,909
Administrative	383,859	422,193
Total financial	2,158,539	2,374,102
Opportunity (volunteers[1])	126,715	111,771
Total Economic cost	2,285,254	2,485,873
Number of participants[2]	4,380	
Cost per participant	521.75	567.55

Notes. 1) Volunteer hours were estimated at the hourly staff rate

2) Students were considered LA's BEST participants if they attend a minimum of 36 days per year.

Benefits

In this analysis, we focus explicitly on the benefits of LA's BEST associated with reduced juvenile crime. The achievement results presented above provide some positive results but are inconsistent; thus warranting caution when trying to establish potential systematic academic benefits or post-secondary schooling benefits. The results, based on the multilevel survival analyses, allow us to build three scenarios related to the benefits.

We use published estimates of benefits related to avoiding juvenile crime (Cohen, 1998) to measure the benefits associated with avoided costs. We present an estimated average cost based on the distribution of crimes in our sample. Cohen's (1988) estimates are based on juvenile crimes committed 1 to 4 times per year. The average in our sample is 2.15 over the sample period. The costs consist of victim costs, direct costs of adjudication, and probation. These costs are presented in Table 44. As noted above, many studies use costs associated with lifelong criminals, which we report as a separate category. Given the nature of the analysis, we use expected survival probabilities as the basis for the benefit-cost analyses. In each case we also include the likelihood of a student falling into a particular outcome set. We follow previous research in using a range for the benefit-cost analysis (Karoly et al, 1988; Cohen, 1988). The outcome sets are defined by each of the scenarios presented below. Hence, we need to combine the probability of a student being in a particular outcome cell within a set and the survival probability. This result is combined with the potential benefit and is used as numerator for the benefit-cost ratio.

Table 44

Present Value Costs of Juvenile Crime

Cost	Low	High	Lifelong	Estimated[1] Sample Ave.
Victim	62,000	250,000		42,470
Adjudication	21,000	84,000		14,385
Total	83,000	334,000		122,238
Adult			1,100,000	

1) Based on the sample distribution of misdemeanors and Felonies and misdemeanor costs estimated at 0.1 of Felony costs.

Scenario 1 – Effect of Annual LA's BEST Attendance. Given the results of Table 41, we assume that exposure has no impact on survival probabilities; the effect is statistically insignificant at $\alpha = .05$. Hence, we compare the benefits and costs across the three levels of student engagement, (i.e., low, medium and high), ignoring the estimated effect of exposure[45]. The results in Table 44 indicate that the expected total crime rate over the study period decrease by level of engagement, except for students sporadically engaged (low), whose estimated rate is somewhat higher than the control group's rate. Given these results, we consider the low engagement group as part of the intent to treat (ITT) group. We base this approach on the common practice of providing separate estimates that excluded low engagement or non-compliant students. In this way, we are able to examine three treatment conditions against the control, as well as two treatment conditions against two controls.

[45] We include the effect of a single year of exposure since it would be impossible to receive any treatment without any exposure.

Table 45

Summary of Results - Annual Exposure

Treatment condition	Number of participants	Probability	Survival probability	Total crime rate
Control			94.1%	5.9%
Low engagement	1,225	49.8%	93.3%	6.7%
Medium engagement	793	32.3%	95.4%	4.6%
High engagement	440	17.9%	96.3%	3.7%
Total	2,458			

Table 46 presents the expected crime costs per student, based on the dollar amounts presented in Table 44 and the total crime rates presented in Table 45. For example, students in the control group are estimated to commit crimes at a rate of 5.9% and the total cost of a lifelong criminal is $1.1 million; hence the expected value in terms of crime of a student in the study would be 0.059 X $1.1 million = $64,900[46].

Table 46

Expected Crime Cost Per Student

	Cost Assumption			
Treatment condition	Low	High	Life	Sample Ave.
Control	4,888	19,668	64,776	7,212
Low engagement	5,588	22,485	74,053	8,190
Medium engagement	3,782	15,219	50,121	5,623
High engagement	3,085	12,416	40,891	4,523

We use the estimated crime costs per student to derive the expected avoided crime cost per student presented in Table 47. For example, the expected avoided crime cost per student when comparing a student in the high engagement treatment condition to the control group, assume low crime costs, is $4,888-$3,058 = $1,802.

[46] Difference in table due to rounding in text.

Table 47

Net Expected Avoided Crime Cost Per Student

Treatment condition	Cost Assumption Low	High	Life	Sample Ave.
(vs. Control)				
Low engagement	-700	-2,817	-9,277	-1,031
Medium engagement	1,106	4,450	14,655	1,629
High engagement	1,802	7,252	23,885	2,654
(vs. Low engagement)				
Medium engagement	1,806	7,267	23,932	2,659
High engagement	2,502	10,069	33,162	3,685

We next calculate the expected value of avoided costs per student using the likelihood of being in a particular treatment group. We report the results in Table 48 as comparisons using three different groups as the numeraire. The first set of results compare the three treatment conditions to the control group. In this case, the potential expected value of benefits vary from $331 to $4,380 per student under the low and lifelong crime assumptions, respectively. The second set of results compare the two treatment conditions that are estimated to have a positive impact against the control group (benefits range from $1,029 to $13,648). The final set of comparisons, under the assumption that the low engagement group comprises a valid counterfactual group, compares the medium and high engagement treatment conditions against the low engagement treatment condition. The expected value of per student benefits is significantly higher when the comparisons were against the low engagement group. The estimates for based on the sample average crime costs naturally fall between the low and high crime cost figures.

Table 48

Expected Value of Avoided Costs

Treatment condition	Cost assumption Low	High	Life	Sample Ave
(vs. Control)				
Low engagement	-349	-1,404	-4,623	-514
Medium engagement	357	1,436	4,728	525
High engagement	323	1,298	4,276	475
Expected value vs. control	331	1,330	4,380	487
(vs. Control)				
Medium engagement	711	2,862	9,425	1,047
High engagement	319	1,282	4,223	469
Expected value vs. control	1,030	4,144	13,648	1,517
(vs. Low engagement)				
Medium engagement	1,161	4,673	15,392	1,710
High engagement	893	3,593	11,834	1,315
Expected value vs. low eng.	2,054	8,267	27,226	3,025

Finally, we use the estimated program costs per student presented in Table 43 and figures above to calculate the benefit-cost ratio. We present these results in Table 49. The benefit-cost ratios we report in Table 49 follow the same set of comparisons we generated in Table 48. The benefit-cost ratios vary quite substantially. The results indicate that under the low crime cost assumption, each dollar invested in LA's BEST returns only $0.58 (when comparing the expected value of three treatment conditions to the control group). Using the lifelong crime cost assumption each dollar invested in LA's BEST would return $7.72. Benefit-cost ratios derive from comparisons excluding the low engagement treatment condition yielded higher benefit-cost ratios. The second set of benefit-cost ratios ranges from 1.81 to 24.05, and the third set of ratios ranges from 3.62 to 47.97. Again, the ratios based on the sample average crime costs fall between the low cost and high cost estimate. It is important to reiterate that these estimates are based on the assumption that exposure is irrelevant, so calculations held exposure constant at one year. Another scenario is to consider benefit-cost ratios at the sample average exposure. This is discussed in the following section.

Table 49

Benefit/Cost Ratios by Cost Assumption

Treatment condition	Cost Assumption Low	High	Life	Sample Ave.
(vs. Control)				
Low engagement	-0.61	-2.47	-8.15	-0.91
Medium engagement	0.63	2.53	8.33	0.93
High engagement	0.57	2.29	7.53	0.84
Expected value vs. control	0.58	2.34	7.72	0.86
(vs. Control)				
Medium engagement	1.25	5.04	16.61	1.85
High engagement	0.56	2.26	7.44	0.83
Expected value vs. control	1.81	7.30	24.05	2.67
(vs. Low engagement)				
Medium engagement	2.05	8.23	27.12	3.01
High engagement	1.57	6.33	20.85	2.32
Expected value vs. low eng.	3.62	14.57	47.97	5.33

Scenario II – Effect of Exposure at the Sample Average Exposure Level. In this analysis, we recalculated the benefit-cost ratios using all of the same assumptions as in Scenario I, except that we changed the exposure level to 1.8 years, which was the sample average. Given that there was a positive coefficient (i.e., increased hazard associated with students who attended more years of LA's BEST), increased exposure actually decreased all of the benefit-cost ratios. Changing the exposure would affect the results presented in all of the tables in Scenario I, included in Appendix B, but here we present only the final table.

The results in Table 50 indicate that when we compare the three treatment conditions to the control group, the weighted benefit-cost ratios are negative (-.027 to -3.56). This is the result of the combined ineffectiveness of low engagement and the use of average exposure. However, among the first set of weighted estimates it is clear that both the medium and high engagement treatment conditions have benefit-cost estimates greater than 1 when the high crime cost assumption is used. If we eliminate the low engagement treatment condition from the weighted average, as was done in the second set of estimates, the benefit-cost ratio under all cost avoidance assumptions is greater than 1. Again, the most optimistic results stem from comparisons between the medium and high treatment conditions against the low treatment condition and the ratios based on the sample average crime costs fall between the low crime and high crime cost estimates. The final column is again based on the sample average distribution of crime and the ratios range from -0.40 to 5.88.

Table 50

Benefit/Cost Ratios by Cost Assumption

	Cost Assumption			
Treatment condition	Low	High	Life	Sample Ave.
(vs. Control)				
Low engagement	-1.13	-4.55	-14.99	-1.67
Medium engagement	0.40	1.61	5.29	0.59
High engagement	0.46	1.87	6.15	0.68
Expected value vs. control	-0.27	-1.08	-3.56	-0.40
(vs. Control)				
Medium engagement	0.80	3.20	10.54	1.17
High engagement	0.62	2.49	8.20	0.91
Expected value vs. control	1.41	5.69	18.74	2.08
(vs. Low engagement)				
Medium engagement	2.25	9.07	29.88	3.32
High engagement	1.73	6.98	22.99	2.55
Expected value vs. low eng.	3.99	16.05	52.87	5.88

Scenario III – Plausible Benefit-Cost Ratios Under All Possible Exposure Levels.

In Scenario III we again use all of the assumptions used in the previous two scenarios, except we allow exposure to vary from one to four years, thus taking all plausible values in our data. That is, rather than estimate benefit cost ratios at various exposure levels and various engagement levels we can use the sample distribution of attendance to weight the results. For example about 25% of students attended LA's BEST for 2 years. Further, about 33% of students are expected to exhibit medium engagement. We use these results with the estimated costs and benefits to derive the results in Table 51.

It is important to reiterate that LA's BEST is designed for students to attend 5 days per week. On average students attend significantly less. If we consider full attendance as the high engagement group then we can use those estimates to compare fully implemented program effects. Further, the low engagement or sporadic engagement (four to nine days per month) group clearly receive no benefits (irrespective of exposure). Hence, Table 51[47] summarizes results as in scenario II where the first set of benefit-cost ratios is based on all levels of treatment against the control group. The second set of benefit-cost ratios compares those students with more than sporadic attendance (in some sense minimal compliers) to the original control group. The final set of benefit-cost ratios considers the sporadically engaged students as receiving no treatment and includes them into the control group for comparison purposes. Using the benefits based on the sample average distribution of costs avoided, we posit that the benefit-cost ratio of 2.46 is most applicable.

Table 51

Benefit/Cost ratios using sample average exposure and condition

	Cost Assumption			
	Low	High	Life	Sample Ave.
Expected value vs. control (includes low, medium, and high engagement)	-0.33	-1.32	-4.35	-0.48
Expected value vs. control (includes medium, and high engagement)	1.67	6.73	22.16	2.46
Expected value vs. low engagement (includes medium, and high engagement)	2.91	11.69	38.50	4.28

[47] A detailed table showing different years of exposure and treatment condition is shown in Appendix C

Discussion and Conclusions

This study set out to evaluate the long-term effects of after-school programming on educational adjustment and juvenile crime. It extends the literature on the impact of after-school program effects on academic achievement and juvenile crime in two key ways. First, the analyses explicitly modeled individual achievement and crime trajectories longitudinally for ten years; and second, it used a large sample of over 6,000 students as a basis.

The retrospective data do not allow us to randomly assign students to treatment and control conditions. However, extensive care is taken to apply advanced multilevel propensity scores methods to establish study samples from which valid inferences could be generated. The comparison group (i.e. students not attending LA's BEST) is assumed to not include participants in other after-school activities. However, it is possible that control students from non-LA's BEST schools attend alternative and comparable after-school programs. We consider this scenario to be very unlikely between years of 1994-1996, when we sample the students to participate in this study. During this period, the supply of after-school programs in the area under study was not as extensive as it is now, and comparable after-school programs (e.g. same foci) were even less likely to exist. However, in the improbable event that a large percentage of control students were attending other comparable after-school programs, the treatment effects find lower bound estimate of program effects for LA'S BEST.

Multilevel longitudinal models are used to model student academic achievement and event occurrence over time. The multilevel modeling is statistically necessary to properly account for the nested structure of the data, but also provides a tool with which important between-school variation in program implementation can be examined. It is important to note that the long-term impacts of the program occur one to seven years after the intervention took place. In addition, given that LA's BEST primarily serves at-risk students in a large urban area, the study results can be generalized to other large urban settings as well.

The results imply marginal, positive program effects on student academic achievement, consistent positive effects on juvenile crime, and generally positive benefit-cost ratios. This study also highlights that simple indicators of program participation are inadequate to fully capture program effects. Results indicate that exposure (years of participation), intensity (total days of attendance), engagement (average weekly attendance), and contact with additional adults during the day all impact program effectiveness. The following paragraphs discuss each of these findings.

Student Academic Achievement

Reading and Mathematics achievement are examined separately over a ten-year period from 1993 to 2003. To more readily isolate treatment effects, the analyses are based on two analysis samples: one comparing treatment students to non-treatment students in the *same* 24 schools; and the other, comparing treatment students against non-treatment students attending 24 *different* schools. The comparison schools in the second analysis did not have LA's BEST programs during the 5 years that LA's BEST students receive services[48].

The results of the four analyses of academic achievement (i.e., two samples by two content areas) provide some evidence for beneficial effects of LA's BEST on long-term achievement growth. The longitudinal models we utilize provide estimates for both achievement status and achievement growth. We are primarily interested in the effect of the treatment on achievement growth as this represents the potential lasting impact of the program. The analyses indicate that simply using a treatment indicator (i.e., splitting students into a treatment and non-treatment group) is insufficient to adequately capture two important program dynamics: student exposure and intensity. We use the number of years of attendance as an indicator of exposure and the total number of days attended as an indicator of intensity. We then used the number of monthly volunteer hours and staff attending workshops as indicators of quality and program implementation fidelity. Further, the analyses included both time-varying student characteristics and time-invariant student characteristics. The time varying student characteristics include student travel (i.e., whether the student was attending his school based on home residence or not), English Language Learner (ELL) status and Re-designated English Proficient (REP) status. Time invariant student characteristics include student demographic characteristics such as gender and race/ethnicity as well as other student background data such

[48] Although, some of the 24 control schools subsequently received LA's BEST programs.

as disability status, GATE (Gifted and Talented) status, SES (parent education), and cohort.

Given the use of multilevel longitudinal models, the following are included as between-school predictors: school contextual effects, program fidelity, and implementation. Results from the analyses of the three HLM achievement models indicate that concomitant variables such as gender, race/ethnicity, disability status, GATE status, and SES indicators demonstrate effects in the posited direction, confirming that the model is capturing the achievement dynamic as well as providing evidence that the sample students' academic achievement is representative of the district as well as patterns found in previous research.

The results (as summarized in Table 39) underline several important patterns. Measuring program exposure demonstrates consistently positive effects for the first post-treatment year. Nevertheless, these effects are not significant – likely due to the imprecision with which exposure captures treatment effects. This is corroborated by the consistently positive and generally significant results for intensity. This interpretation is supported by the results derived from the models using both exposure and intensity. These models imply that exposure as a single indicator confound potential program effects and the sorting of students with the least after-school alternatives. That is, these students attend multiple years because there is a lack of after-school choice available to parents.

Upon further analysis, the results from models include both exposure and intensity are interpreted as the marginal effects of intensity accounting for exposure. That is, with everything being equal, students attending the same number of years more often demonstrate positive achievement effects at the beginning of the post treatment period. This indicates that achievement effects do not necessarily last beyond the first year after students receive the intervention. Instead, program quality is an important indicator for the potential of the program to have lasting effects on student academic achievement growth. Given the model specification, results imply that students attending LA's BEST with a greater adult presence fair better throughout the grade span. This provides further evidence of the effects of program quality.

The results demonstrate stability with respect to the sign of the effects across the different samples and tests. Treatment effects are not significant when treated students are compared to untreated classmates within the same schools. The treatment effects, however, become significant when treated students are compared to untreated students in different schools. This finding implies that there may be other potential explanations, besides LA's BEST attendance that account for such effects. One potential reason could be the existence of a positive *diffusion-effect* on control students attending the treatment schools. The control students who attend schools offer the LA's BEST program may have benefited from the program by means of sharing their classrooms with treatment students. The mechanisms on how the diffusion effect might operate warrants further investigation. Another potential reason could be the existence of key unobserved characteristics of the control students that are not explicitly accounted for in the selection models.

The mixed achievement results are consistent with previous research on academic effects of after-school programs. However, the potential effects of participations in after-school on academic attitudes, work habits, and social development should not be overlooked. As previously reviewed in the literature, it is clear that these students process the risk factors associated with low retention in school and juvenile delinquency (Mayer, 2001; Hawkins et al, 2000; Carr & Vandiver, 2001). In order to counter academic failure and juvenile delinquency, these students need to have access to protective buffers that will decrease the likelihood of them engaging in problematic, antisocial and anti-school behaviors, and increase the likelihood of them developing into competent and successful adolescents. Studies have shown that resiliency in youth is developed by affirming personal relationships that teach about the importance of education and provide a sense of well-being (Durlak & Weissberg, 2007; Reisner, Vandell, Pechman, Pierce, Brown & Bolt, 2007). In their program mission, LA's BEST has stated "relationship building" and "academic readiness" as among their central goals for "developing the whole child". Accordingly, previous studies of LA's BEST consistently demonstrate that participating students develop better attitudes towards reading (Huang & Lin, 2000; Huang et al, 2001, 2002), have better attitudes towards school (Huang, Choi, Davis, Henderson, Kim, Lin, & Waite, 2003; Huang, Choi, Henderson, Howie, Kim, Vogel, Waite, & Yoo, 2004), enjoy school more (Huang, Choi, Henderson, Howie, Kim, Vogel, Yoo, & Waite, 2005), and have increase their school attendance (Huang, Choi, Davis, Henderson, Kim, Lin, & Waite, 2003; Huang, Choi, Henderson, Howie, Kim, Vogel, Yoo, & Waite, 2004). School teachers report that students improved their attitudes towards learning, their work habits, and homework completion after they attended LA's BEST (Huang, Miyoshi, La Torre, Marshall, Perez, & Peterson, 2006; Huang, Coordt, La Torres, Leon, Miyoshi, Perez, & Peterson, 2007). Both afterschool staff and school teachers perceive that students have higher self-confident, better participation in class, and higher self-efficacy in learning after they participated in the program (Huang, Choi, Davis, Henderson, Kim, Lin, & Waite, 2003; Huang, Choi, Henderson, Howie, Kim, Vogel, Yoo, & Waite, 2004; Huang, Choi, Henderson, Howie, Kim, Vogel, Yoo, & Waite, 2005).

As for fostering a sense of well-being, the recreational beat focuses on developing the social competency of the students, team activities such as softball, drill team, gymnastic team, health and nutrition programs, conflict resolution skills classes, leadership training in debate teams, etc. are all geared towards developing self-efficacy, cooperation, and collaborative skills of the students. These activities stress the importance of effort and teamwork, and provide a collaborative environment and opportunities for students to foster their creativity and critical thinking. Past studies on LA's BEST indicated that participating students develop better relationships with adults and peers, have better conflict resolution skills, have higher aspirations towards their future and indicate a desire to finish high school and go on to college (Huang, Choi, Davis, Henderson, Kim, Lin, & Waite, 2003; Huang, Choi, Henderson, Howie, Kim, Vogel, Yoo, & Waite, 2004; Huang, Choi, Henderson, Howie, Kim, Vogel, Yoo, & Waite, 2005).

Furthermore, results in the model specification imply that program quality as indicated by staff/student ratio and engagement are key ingredients for positive outcomes. Students attending LA's BEST with greater adult presence fair better throughout the grade span. These results support the findings of an earlier report on LA's BEST *The After-school Hour* (Huang et al, 2007), which states that students and staff develop close relationships that build up resiliency for the students. This is transpired through high expectations that are placed upon students, encouragement provided by staff, and support in program resources that help student achieve their goals. In addition, program quality in LA's BEST is expressed by the consistent findings throughout the years as parents consistently remain satisfied with the program in terms of equipment, materials, management, and staff quality (Huang, Choi, Davis, Henderson, Kim, Lin, & Waite, 2003; Huang, Choi, Henderson, Howie, Kim, Vogel, Yoo, & Waite, 2004; Huang, Miyoshi, La Torre, Marshall, Perez, & Peterson, 2006). Parents and teachers also consistently perceived that students have become more interested in school, and are benefited both socially and academically (Huang, Choi, Davis, Henderson, Kim, Lin, & Waite, 2003; Huang, Choi, Henderson, Howie, Kim, Vogel, Yoo, & Waite, 2004). Students also reported enjoying the program, feeling safer after attending the program, and feeling genuinely cared for by the staff (Huang, Choi, Davis, Henderson, Kim, Lin, & Waite, 2003; Huang, Choi, Henderson, Howie, Kim, Vogel, Yoo, & Waite, 2004; Huang, Miyoshi, La Torre, Marshall, Perez, & Peterson, 2006). All these positive outcomes are essential in cultivating youth attitudes and behaviors that will lead them to resiliency and become successful citizens in the future.

This study also suggests a positive *diffusion-effect* on control students attending the treatment schools. Part of the LA's BEST training for their Site Coordinators is on how to establish a solid relationship with the school principal, close connection between the day school staff and the after school staff, and how to communicate with the school teachers to enhance linkages between school and afterschool curriculum. At the same time, LA's BEST staff are also trained to establish relations with the community and bring in additional resources to the program, that include: 1) inviting dentists to visit the program to demonstrate dental hygiene; 2) inviting book authors to come and read for the students; 3) a community project where students go out to pick up street trash for a day; 4) or volunteer in the neighborhood senior home. As a result, both the human capital and the resources of the school and surrounding community are highly leveraged. These interactions may have benefited the entire community and the non-participating students as suggested with the diffusion effect. Further study is needed to examine this phenomenon.

At the same time, it should be noted that while some effects are statistically significant and substantively important, other results suggest limitations. Results show that Hispanics and African American students not only have initial achievement gaps but are also expected to fall further behind at a rate of about 2.2 (p <.01) and 1.2 (p <.10) NCEs per year, respectively. At this rate, LA's BEST cannot have a significant impact in closing these widening achievement gaps.

This also indicates the need for additional research to more carefully identify mechanisms through which program elements can have higher impacts on specific subgroups.

Long-Term Impacts on Juvenile Crime

The results from the multilevel survival analyses indicate that LA's BEST positively impacts juvenile crime survival probabilities. Moreover, the result of differential crime hazards is not found between LA's BEST and non-LA's BEST schools, but directly relates to individual student participation in the program. This indicates that the relationship between LA's BEST programs and juvenile crime hazards do not result from a selection process. In essence, LA's BEST does not select the "best" schools to place their programs.

Similar to achievement outcomes, a simple program participation indicator fails to adequately capture program effects related to juvenile crime. The results indicate that program quality, exposure, and engagement, need to be considered together in order to identify program effects. After engagement and exposure are properly parameterized, the results are extremely robust across alternative specifications and modeling choices. That is, program effects remain consistent, irrespective of other concomitant student factors or school and neighborhood context effects included in the model. Further, the results are consistent whether the survival models are single level models, multilevel models, or multilevel frailty models.

Specifically, model results are consistent with expectations regarding student level effects. For example, boys are estimated to be three times as likely to commit a crime as girls. The results also demonstrate the importance of considering multiple characteristics simultaneously. For instance, African Americans do not have distinguishable[49] crime rates in comparison to their classmates, when student level characteristics and parental level of education are controlled. It is also interesting to note that student classification bears some relationship with juvenile crime. Students with disabilities are estimated to have a crime rate that was 30% higher than non-disabled students. The interplay of

[49] African Americans appear to have distinguishable crime rates in comparison to other groups when factors are not controlled for

these factors combined warrants further study in its relationship with juvenile delinquency and crime.

We also test several potential interactions. That is, we attempt to identify effects of moderating student factors. For example, are treatment effects more or less pronounce for students whose parents had less education? It is found that while parental education is significantly related to juvenile crime rate, it has no impact on program effects. The program benefits all students equally. Participating in the program reduces the hazard in committing crime for both students from homes of better educated parents and students from less educated parents. This also implies that the program could not mitigate all existing differences in crime hazards.

As noted, the key parameterization correctly specifies how students receive the treatment. Again, for juvenile crime this consists of including exposure and engagement. Student exposure is one to four or more years and engagement is classified into three levels: low (four to nine days of attendance per month); medium (10 to 14 days of attendance per month); and high (at least 15 days per month). The results indicate that few benefits accrue to students who only sporadically attend (low engagement), but that benefits increase as engagement increases (although not linearly – rather as a step function). In other words, it is found that students who are intensely engaged benefited most from LA's BEST, while those who are moderately engaged also benefit.

Similar to the achievement models, we take advantage of the multilevel models and examine between-school differences in program effects. Two key between-school effects emerged. First, controlling for individual student SES, school average SES plays a significant role in mediating crime rates. That is, students who attend higher SES schools (whether or not the student was classified as low SES) demonstrate reduced crime hazards. Second, for students who sporadically attend, an important mediating factor is the percentage of households (per neighborhood population) living below the poverty threshold. The model results imply that even sporadic participation in the program lead to some reduction in crime hazards for students living in very poor neighborhoods. This provides further validation for LA's BEST effects as these neighborhoods are a focus of the intervention.

As mentioned previously, after-school programs are beneficial to student resiliency and the prevention of juvenile delinquency in three critical ways. First, they provide children with supervision during a time when they might normally fall prey to deviant or antisocial behaviors[50]. At the same time, LA's BEST also increases students' feelings of attachment to school and provides them with skills needed to avoid delinquent behaviors. Secondly, after-school programs provide experiences that may benefit students' social skills and classroom conduct. As indicated in the previous LA's BEST studies, student participants tend to exhibit better behavior in school and higher academic interest, better social skills and self-control, and improved self-confidence through the development of positive relationships with adults and peers. Furthermore, students can also benefit from the extra-curricular activities that LA's BEST offer. According to the Carolina Longitudinal Study (Cairns & Cairns, 1994), extracurricular activity participation is associated with low rates of early school drop-out (Mahoney & Cairns, 1997, Huang et al, 2005) and low rates of criminal arrest in young adulthood (Mahoney, 2000). Finally, after-school programs may help improve academic achievement (Fashola, 1998). LA's BEST students who participate in these programs are more positive about school and their own schoolwork, and are more likely to have ambitions to graduate from high school and attend college (Huang et al, 2006). These students who have bright out look for their futures are less likely to commit crime (Mahoney & Cairns, 1997).

Moreover, the model results of this study imply that even sporadic participation in LA's BEST lead to some reduction in crime hazards for students living in very poor neighborhoods. This finding affirms that aversive or punitive environments in the community and neighborhood such as poverty, community disorganization, and exposure to drugs, criminal adults, violence, and racial prejudice all contribute to antisocial behaviors (Hawkins et al, 2000). For these students, protective buffers (i.e. providing a safe place to go to after school, and receiving mentorship and encouragement from adults) are especially important in dissuading them from delinquent involvement.

[50] Research shows that the rates for both violent juvenile crimes and victimization of juveniles peak between 3 and 6 p.m. on school days (Newman et al., 2000; Richardson et al, 1993; U.S. Department of Education & U.S. Department of Justice, 2000).

Benefit-Cost of LA's BEST on Juvenile Crime Results

Benefit-cost analysis of the LA's BEST program demonstrates that students and the larger society benefit from the program. Our benefit-cost analysis is based on whether the present value of benefits accrued to program participants and society at large are greater than the actual program costs. We calculate benefit-cost ratios that summarize the value of benefits that accrue to each dollar spent on the after-school program. A benefit-cost ratio greater than 1 indicates that the benefits outweigh the costs. Benefits and costs are calculated in constant dollar terms and the present values of benefit and cost streams are generated by discounting by the CPI.

The benefit-cost analysis focuses explicitly on the benefits of LA's BEST associated with reduced juvenile crime.[51] We follow the lead of previous evaluations in using cost estimates[52] associated with specific crimes and juvenile court costs as provided by Cohen and his collaborators (2000, 1998). Three scenarios are used in measuring the ratios: holding exposure constant at one year; using the sample average exposure (years attended); and using each year of participation separately. The benefit-cost ratios are calculated for low, high, lifelong, and sample based crime-avoidance cost estimates.

As noted, the results indicate that the expected crime rates decrease as engagement increases-except for students sporadically engaged. Given these results, we compare benefits and costs of the entire intent to treat group against the controls; we compare the effects of the treatment on the treated (i.e., medium and high engagement) against the control group and the low engagement group.

The benefit-cost results of the three scenarios highlight the importance of assumptions when deriving estimates. Discounted, expected benefits-cost ratios demonstrate extreme variability depending on the assumptions. The ratios range from about $-40.76 to $68.81. In sorting through the myriad of expected values, it is important to focus on those that are most plausible. If we include all exposure

---

[51] Although achievement results presented some positive findings, they were not consistent enough to confidently establish potential systematic academic benefits for this analysis.

[52] Estimates of tangible costs to victims for specific types of crimes are based on the National Crime Victimization Survey (NCVS).

and engagement levels and use the no-treatment condition as the sole comparison, benefit-cost ratios range from $0.57 to $7.53 for the high engagement students; from $-4.79 to $8.33 for the medium engagement students, and the low engagement student have negative benefit-cost ratios. However, when only medium and high engagement students are compared against the control, the benefit-cost ratio is significantly increased[53]. In this case, expected total benefit-cost ratios[54] range from $0.09 to $24.05. Ultimately, the most plausible combination of exposure (sample average) and engagement (medium and high), using the sample average cost avoidance (from Table 43) yields a benefit-cost ratio of $2.50.

The above analyses highlight the importance of proper identification and categorization of the treatment and control conditions. In recognizing that participation in a program is more than a binary supposition, we discover that the frequency and quality of instruction are essential contributors to program success. As a result, the benefit-cost ratios vary dramatically based on the group considered as the correct counterfactual and the base level from which students actually receive treatment. Under most assumptions and scenarios, the medium and high treatment conditions yield benefit-cost ratios greater than 1. This means that a dollar invested in the LA's BEST after-school program returns more than a dollar in benefits. The findings clearly suggest that a sporadic level of participation is insufficient to reap program benefits. Future studies need to consider selection, program implementation and participation very carefully. We next turn to limitations of this study as well as implications for after-school programs and future program evaluations.

Limitations

The study demonstrates the application of rigorous sampling and verification procedures. It also illustrates the use of advanced statistical modeling techniques with existing data to examine the long term and longitudinal effects of an after-school intervention. While the methodology is strong, it is subject to the limitations of existing data. First, the nature of program implementation and

---

[53] This occurs because the exclusion of low engagement students would help to take away any negative effects of exposure and low engagement from the calculations for treatment effects

[54] Benefit-cost ratios are based on comparisons of the medium and high engagement

available data disallows the possibility of a randomized, experimental design - the preferred method through which potential confounding factors can be ruled out as alternative explanations. Thus, we are forced to rely on multilevel propensity scores to build a counterfactual, control group. Using propensity scores reduces effects of unmeasured factors, to some extent, but we are not able to rule out selection entirely. That is, we can not completely rule out a common cause for participating in the after-school program and committing crimes. Second, the unavailability of some data in the early years disallows the inclusion of some potential mediating factors in the analysis. Finally, there is the possibility of self-selection bias in the findings for the medium and high engagement groups. Although we assume that this second selection stage (stage one was enrolling in LA's BEST or not) is based largely on program quality.

Implications for Evaluating After-School Programs. Study results also argue for more sensitive indicators of program implementation in order to provide cleaner estimates of program effects and give program sponsors a clearer picture of what constitutes best practices. It is highly likely that gross indicators miss important program effects, as demonstrated in the findings on the relationships between the intensity of participation and engagement. For example, indicators of staff development in this study are unrelated to student performance. Why there is this lack of relationship is unclear. For example, whether it is due to poor staff development, or poor staff development implementation, or simply just due to significant noise in the indicator is unknown. It is suggested that after-school programs need to regularly collect data that monitor indicators of implementation quality. In order to critically and objectively evaluate programs, programs must monitor specific elements that are hypothesized to relate to their effectiveness. This implies that programs need to carefully consider their theory of action; they need to monitor and collect data to provide information whether the theory of action is opertionalized and to what extent. Also, programs need to monitor student attendance carefully in order to develop precise indicators of exposure, intensity and engagement.

One interesting finding in this study is that the low participation group actually performs the worst in all areas. Ecocultural theory emphasizes that a major adaptive task for each family is the construction and maintenance of a daily routine through which families organize and shape their children's activity and development (Rogoff, 1990; Gallimore & Goldenberg, 1993). Similarly, child developmental theorists also stress consistency as very important in developing children's self-discipline and self-regulation (Caprara et al, 2002). In turn, these characteristics are important corner stones for building resiliency (Newman et al, 2000). In this study, the low participation group included students who attended a minimum of one day a week per year, whether this pattern of inconsistency in the students' life have a negative effect on their intellectual and social development needs to be investigated further.

Another important note for the afterschool evaluators/ researchers is that consensus need to be reach in order to establish a uniform cutting point (for days of attendance per year) for students to be considered as participants. Since intensity and engagement have significant impact on outcomes, how we include/exclude participants will define the severity of the impact. This uniformity can be extremely helpful for audiences in interpreting findings across studies.

Furthermore, precise data that explain the mechanism (that operates between treatment and control students) on selection needs to be collected. For example, it is possible that control students in treatment schools decide not to participate in LA's BEST because they choose instead to participate in other after-school care activities. Furthermore, it is possible that unmeasured family characteristics are potentially different between these two groups. Studies intending to examine program effects should have access to such data. School selection mechanisms also need to be considered as well.

Since it is often difficult to measure within-site variation in program quality, multilevel models are particularly useful as they can partition variation into within and between site pieces. The between-site portion of the model can examine between-site differences in program quality and implementation (assuming such data have been collected). These site or school level factors can then be combined with school context (e.g. percentage of low SES population at a school) and neighborhood effects to determine whether they mediate program effectiveness.

It should also be noted that after-school programs are intended to benefit students beyond simply the actual time that the student is participating. The only way we can accurately reflect student outcomes both during and post treatment is to apply longitudinal models to examine effects over time. This is particularly relevant for youth crime as juvenile crime is very low in elementary school and only begins to increase after students enter middle and high school.

In addition, benefit-cost ratios are an important extension of traditional program evaluations as they provide results that allow stakeholders to determine whether the resources are placed into a program that has sufficient returns to warrant continued investment in the program. Benefit-cost studies require several assumptions to be made and some previous studies focus on the most encompassing results to demonstrate after-school program effects. It is incorrect to ascribe all potential benefits from multiple programs to a single program and sum the benefits. Also, in the case of juvenile crime, many studies assume that students will become life-long criminals. However, evidence suggests this is unlikely even for juveniles who commit crimes. Benefit-cost estimates are based on life-long crime significantly over-estimate avoided costs.

Implications for the Implementation of After-School Programs.

In summary, results indicate that after-school programs are potentially a powerful resource that can help reduce juvenile delinquency rates. Quality after-school programs such as LA's BEST teach students the academic and social skills they need to avoid the anti-school behaviors and attitudes that contribute to juvenile delinquency. The study results have several implications for the implementation of after-school programs so that participating students can reap maximum benefits. First, the traditional use of attendance as a key measure of engagement may be weak; instead, the results clearly demonstrate that the programs need to engage students and that this is accomplished with consistent attendance and through the use of additional adults (e.g. volunteers). Therefore, programs need to focus on engaging students, ensuring a minimum of 10 days of attendance per month, plus recruit and maintain a regular flow of volunteers to enhance program benefits. As noted above, programs must carefully state their theory of action and explicitly attempt to operationalize it (and collect data on this). Simply filling out student rosters year after year will not benefit students unless they are consistent and engaged participants.

The results also imply that neighborhood poverty is at least as important as school context. Hence, programs must improve their outreach efforts. Although the effects of combined individual and neighborhood effects warrant further study, our findings offer support for the potential positive effects of establishing after-school programs in the most at-risk and underserved neighborhoods and communities.

Concluding Statement

LA's BEST after-school program demonstrates statistically and substantively positive effects on youth crime abatement, especially for students who attend at least 10 days per month. Although the results for achievement are less consistent, however, it demonstrates a positive relationship between achievement scores and attendance. Benefit-cost ratios also vary substantially depending on assumptions, but the most plausible estimate indicates that each dollar spent on LA's BEST returns a benefit of $2.50 to society. This study highlights key issues about causal claims of program effects and isolates specific elements related to effects. It also suggests that a well implemented program consistently engages students, and thus promotes important benefits toward educational adjustment and juvenile crime.

References

Aitkin, M., & Longford, N. (1986). Statistical modeling issues in school effectiveness studies. *Journal of the Royal Statistical Society. Series A (General), 149*(1), 1-43.

Barber, J., Murphy, S., Azinn, W., & Maples, J. (2000). Discrete-time multilevel hazard analysis. *Sociological Methodology*, 30, 210-235.

Barnett, W. S. (1995). Benefits of compensatory preschool education. *Journal of Human Resources*, 27, 279-312.

Barnett, W. S. (1996). *Lives in the balance: age 27 benefit-cost analysis of the High/Scope Perry Preschool Program*. Monographs of the High/Scope Educational Research Foundation, no. 11. Ypsilanti, Mich: High/Scope Press.

Bloomquist, N. (1977). On the relation between change and initial value. *Journal of the American Statistical Association*, *72,* 746-749.

Brown, W. O., Frates, S. B., Rudge, I. S., & Tradewell, R. L. (2002). *The costs and benefits of after school programs: The estimated effects of the After School Education and Safety Program Act of 2002*. Claremont, CA: The Rose Institute of Claremont-McKenna College.

Burstein, L. (1980). The analysis of multi-level data in educational research and evaluation. *Review of Research in Education*, *4,* 158-233.

Cairns, R. B., & Cairns, B. D. (1994). *Lifelines and risks: pathways of youth in our time*. Cambridge: Cambridge University Press.

Caldwell, M., Vitacco, M., & Van Rybroek, G. (2006). Are Violent Delinquents Worth Treating? *Journal of Research in Crime and Delinquency*, *43*(2), 148-168.

Callens, M., & Croux, C. (2005). The impact of education on third births. A multilevel discrete-time hazard analysis. *Journal of Applied Statistics*, *32*(10), 1035-1050.

Campbell, D. T., & Stanley, J. (1963). *Experimental and quasi experimental designs for research*. Chicago: Rand McNally.

Caprara, G.V., Regalia, C., & Bandura, A. (2002). *Longitudinal impact of perceived self-regulatory efficacy on violent conduct*. *European Psychologist*, *7*, 63-69

Carr, M., & Vandiver, T. (2001). Risk and protective factors among youth offenders. *Adolescence*, *36*, 409-426.

Cassell, R., Chow, P., Demoulin, D., & Reiger, R. (2000). Extracurricular activities in high school produces honesty and fair play needed to prevent delinquency and crime. *Education*, *121*(2), 247-251.

Chinen, M., & Goldschmidt, P. (2006, April). Identifying counterfactual groups in evaluation studies using observational data in nested longitudinal settings. *Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.*

Cohen, M.A. (1998). The monetary value of saving a high-risk youth. *Journal of Quantitative Criminology*, *14*(1), 5-33.

Cohen, M.A. (2000). To treat or not to treat? A financial perspective. In C. R. Hollin (Ed.), *Handbook of Offender Assessment and Treatment*. New York: Wiley.

DeKalb, J. (1999). *Student truancy*. Eugene, OR: ERIC Clearinghouse on Educational Management. (ERIC Document Reproduction Service No. ED 429 334)

Fashola, O. S. (1998). *Review of extended-day and after-school programs and their effectiveness.* (Report No. 24). Baltimore MD: Center for Research on the Education of Students Placed at Risk. (ERIC Document Reproduction Service No. ED 424 343)

Freeman, K., & Redding, B. (1999). *LA's BEST after school enrichment program: Replication manual.* Los Angeles: LA's BEST.

Freedman, D., Pisani, R., & Purves, R. (1978). *Statistics*. New York: W. W. Norton & Co.

Gallimore, R., & Goldenberg, C. (1993). Activity settings of early literacy: Home and school factors in children's emergent literacy. In E. Forman, N. Minick, & C.A. Stone (Eds.), *Context for learning; Sociocultural dynamics in children's development.* 315-335. Oxford: Oxford University Press.

Garcia, E. (1990). An integrated county court system. In Proceedings of a Bureau of Justice Statistics/SEARCH Conference, *Juvenile and adult records: One system, one record.* Washington, DC: Department of Justice.

Garmezy, N. (1991). Resiliency and vulnerability to adverse developmental outcomes associated with poverty. *American Behavioral Scientist*, *34*(4), 416-430.

Goldschmidt, P. (1997). *The transition from high school to work: A longitudinal multilevel analysis*. Unpublished doctoral dissertation, University of California, Los Angeles.

Goldschmidt, P., Choi, K., & Martinez, F. (2003). *Using hierarchical growth models to monitor school performance over time: Comparing NCE to scale score results*. CSE Technical Report. Los Angeles, CA: Center for the Research on Evaluation, Standards, and Student Testing, University of California, Los Angeles.

Goldschmidt, P., & Wang, J. (1999, Winter). When can schools affect dropout behavior: A longitudinal multilevel analysis. *American Educational Research Journal, 36*(4), 715-738.

Goldstein, H. (1999). *Multilevel statistical models*. London: Institute of Education, Multilevel Models Project.

Gordon, K. A. (1996). Resilient Hispanic youths' self-concept and motivational patterns. *Hispanic Journal of Behavioral Sciences, 18*(1), 63-73.

Greenwood, P., Model, K., Rydell, C., & Chiesa, J. (1998). *Diverting children from a life of crime.* Washington DC: RAND.

Hanushek, E. & Lockeed, M. (1988). Improving educational efficiency in developing countries: What do we know? *Compare, 18*(1): 21-38.

Hawkins, J., Herrenkohl, T., Farrington, D., Brewer, D., Catalano, R., Harachi, T., & Cothern, L. (2000). Predictors of youth violence. *Juvenile Justice Bulletin*. Washington DC: Department of Justice.

Hodgkinson, H. (2006). *The whole child in a fractured world.* For the commission on the Whole child convened by the association for Supervision and Curriculum Development. Alexandria, VA.

Hox, J. J. (2002). *Multilevel analysis techniques and applications*. Quantitative methodology series. Mahwah, N.J.: Lawrence Erlbaum Associates.

Huang, D., Choi, K.C., Henderson, T., Howie, J., Kim, K., Vogel, M., Waite, P. & Yoo, S. (2004). *Exploring the long-term impact of LA's BEST on students' social and academic development.* Los Angeles: UCLA Center for the Study of Evaluation.

Huang D., Choi K. C., Henderson T., Howie, J., Kim, K., Vogel, M., Waite, P., & Yoo, S. (2005). *Examining the effects of academic skills and academic enablers* *taught at LA's BEST on the achievement of student participants.* Los Angeles: UCLA Center for the Study of Evaluation.

Huang, D., Choi, K., Davis, D., Henderson, T., Kim, K., Lin, S., & Waite, P. (2003). *Evaluating the impact of LA's BEST on students' social and academic development: Study of 74 LA's BEST sites 2001–2002 final report*. Los Angeles: UCLA Center for the Study of Evaluation.

Huang, D., Choi K.C., Henderson T., Howie, J., Kim, K., Vogel, M., Waite, P., & Yoo, S. (2005). *Examining the effects of academic skills and academic enablers taught at LA's BEST on the achievement of student participants.* Los Angeles: UCLA Center for the Study of Evaluation.

Huang D., Choi K., Henderson T., Howie, J., Kim, K., Vogel, M., Waite, P., & Yoo, S. (2005). *Examining the effects of academic skills and academic enablers taught at LA's BEST on the achievement of student participants.* Los Angeles: UCLA Center for the Study of Evaluation.

Huang, D., Coordt, A., La Torre, D., Leon, S., Miyoshi, J., Pérez, P., & Peterson, C. (2007). *The after-school hours: Examining the relationship between afterschool staff-based social capital and student engagement in LA's BEST.* Los Angeles: UCLA Center for the Study of Evaluation.

Huang, D, Davis, D., Golestani, C., Henderson, T., Kin, K., & Kosman, G. (2002) *The Exploratory evaluation of Literacy Loop: An after-school tutoring program.* Los Angeles: UCLA Center for the Study of Evaluation.

Huang, D, Devries, L., Golestani, C., Howie, J., Henderson, T., Kin, K., & Waite, P. (2003) *The Literacy Loop program in LA's BEST: The second year of implementation.* Los Angeles: UCLA Center for the Study of Evaluation.

Huang, D. & Lin, s. (2000). *The after-school literature project in the LA's BEST program.* Los Angeles: UCLA Center for the Study of Evaluation.

Huang, D., Kim, K. S., Marshall, A., & Pérez, P. (2005). *Keeping kids in school: An LA's BEST example study: A study examining the long-term impact of LA's BEST on students' dropout rates.* Los Angeles: UCLA Center for the Study of Evaluation.

Huang, D., Miyoshi, J., La Torre, D., Marshall, A., Pérez, P., & Peterson, C. (2006). *Exploring the intellectual, social and organizational capitals at LA's BEST.* Los Angeles: UCLA Center for the Study of Evaluation.

Karoly, L. A., Greenwood, P. W., Everingham, S. S., Hoube, J., Kilburn, M. R., Rydell, C. P., Sanders, M., & Chiesa, J. (1998). *Investing in our children: What we know and don't know about the costs and benefits of early childhood interventions*. Santa Monica, CA: Rand.

Kleinbaum, D. (1996). *Survival Analysis: A Self-Learning Text*. New York: Springer.

Kleingartner, D. G. (1997). *Survival analysis*, New York: Springer Verlag.

Krovetz, M. L. (1999). Fostering resiliency. *Thrust for Educational Leadership, 28*(5), 28-31.

Levin, H. M. (1983). *Cost-effectiveness: a primer*. New perspectives in evaluation, v. 4. Beverly Hills: Sage Publications.

Linn, R. L. (2000). Assessments and accountability. *Educational Researcher, 29*(2), 4-16.

Mahoney, J. (2000). Strategies of causal inference in small-N analysis*, Sociological Methods & Research, 28*(4), 387-424.

Mahoney, J. L., & Cairns, R. B. (1997). Do extracurricular activities protect against early school dropout? *Developmental Psychology, 33*, 241-253.

Masten, A. S., Hubbard, J. J., Gest, S. D., Tellegen, A., Garmezy, N., & Ramirez, M. (1999). Competence in the context of adversity: Pathways to resilience and maladaptation from childhood to late adolescence. *Development and Psychopathology, 11*, 143-169.

Mayer, G. (2001). Antisocial behavior: Its causes and prevention within our schools. *Education and Treatment of Children*, *24*(4), 412-429.

Murnane, R., Willet, J. & Levy, F. (1995). The growing importance of skills in wage determination. *The Review of Economics and Statistics, 77*, 251-266.

Newman, S., Fox, A., Flynn, E., & Christeson, W. (2000). *America's after-school choice: The prime time for juvenile crime, or youth enrichment and activity*. Washington DC: Fight Crime: Invest in Kids.

O'Donnell, J., Hawkins, J., & Abbott, R. (1995). Predicting serious delinquency and substance use among aggressive boys. *Journal of Consulting and Clinical Psychology*, *63*, 529-537.

Osgood, W., & Smith, G. (1995). Applying hierarchical linear modeling to extended longitudinal evaluations: The boys town follow-up study. *Evaluation Review*, *19*(1), 3-38.

Ramey, C.T. and Ramey, S.L. (1998). Early intervention and early experience. *American Psychologist, 84*, 1024-1032.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Reardon, S., Brennan, R., Buka, S. (2002). Estimating multi-level discrete-time hazard models using cross-sectional data: Neighborhood effects on the onset of adolescent cigarette use. *Multivariate Behavioral Research, 37*(3), 297-330.

Resiner, E., Vandell, D.L., Pechman, E.M., Pierce, K.M., Browm, B.B., & Bolt, D. (2007). *Chartering the benefits of high quality after-school program experience: Evidence from new research on improving after-school opportunities for disadvantage youth*. Washington, D.C.: Policy Studies Associates, Inc.

Reynolds, A. J. (1998). Resilience among Black urban youth: Prevalence, intervention effects, and mechanisms of influence. *American Journal of Orthopsychiatry, 68*(1), 84-100.

Reynolds, A. J., Temple, J. A., Robertson, D. L., & Mann, E. A. (2002). Age 21 Cost-Benefit Analysis of the Title I Chicago Child-Parent Centers. *Educational Evaluation and Policy Analysis*. *24*(4), 267-303.

Richardson, J., Radziszweska, B., Dent, C., & Flay, B. (1993). Relationship between after-school care of adolescents and substance use, risk taking, depressed mood, and academic achievement. *Pediatrics*, *92*, 32-38.

Richman, J. M., & Fraser, M. W. (2001). Resilience in childhood: The role of risk and protection. In J.M. Richman & M.W. Fraser (Eds.), *The Context of Youth Violence* (pp.1-12). Westport, CT: Praeger.

Rogoff, B. (1990). *Appreticeship in thinking: Cognitive development in social context.* Oxford: Oxford University Press.

Rogosa, D. R., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin*, *92,* 726-74.

Rosenbaum, P. R. (1986). Dropping out of high school in the United States: An observational study. *Journal of Educational Statistics*, *11*(3), 207-224.

Rosenbaum, P. R. (1987). The role of a second control group in an observational study. *Statistical Science, 2*(3), 292-316.

Rumberger, R., & Lamb, S. (2003). The early employment and further education experiences of high school dropouts: A comparative study of the United States and Australia. *Economics of Education Review, 22*, 353-366.

Schaps, E. (2006). Educating the whole child. For the Commission on the Whole-Child convened by the Association for Supervision and Curriculum Development. Alexandria, VA.

Schweinhart, L. J., Barnes, H. V., & Weikart, D. P. (1993). *Significant benefits: the High-Scope Perry preschool study through age 27*. Monographs of the High/Scope Educational Research Foundation, no. 10. Ypsilanti, MI: High/Scope Press.

Scott-Little, C., Hamann, M., & Jurs, S. (2002). Evaluations of after-school programs: A meta-evaluation of methodologies and narrative synthesis of findings. *American Journal of Evaluation*, *23*(4), 387-419.

Seltzer, M. (2004). The use of hierarchical models in analyzing data from experiments and quasi-experiments conducted in field settings. In D. Kaplan (Ed.), *The Handbook of Quantitative Methods for Social Science*. Thousand Oaks, CA: Sage Publications.

Shumow, L., Vandell, D. L., & Posner, J. (1999). Risk and resilience in the urban neighborhood: Predictors of academic performance among low-income elementary school children. *Merrill-Palmer Quarterly, 45*(2), 309-331.

Singer, J. D., & Willet, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford: University Press.

Smith, M., Hill, G. C., & Bandera, M. (1997). Success of high-risk students after completion of an elementary school intervention program: A longitudinal study. Tucson, AZ: *Proceedings of the Annual NREA Convention*. (ERIC Document Reproduction Service No. ED413151)

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.

Thum, Y. M. (2002). *Measuring Student and School Progress with the California API.* CSE Technical Report 578. Los Angeles: Center for Research on Evaluation, Standards, and Student Testing, UCLA.

U. S. Department of Education. (2002, February 4). *Fiscal year 2003 budget summary*. Retrieved July 10, 2003 from http://www.ed.gov/offices/OUS/Budget03/Summary/SectionII/A.html#9

U.S. Department of Education & U.S. Department of Justice. (2000). *Working for children and families: Safe and smart after-school programs.* Washington, D. C.: U. S. Government Printing Office. (ERIC Document Reproduction Service No. ED 441 579)

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1998). *Educational resilience.* Philadelphia, PA: National Research Center on Education in the Inner Cities.

Werner, E., & Smith, R. (1992*). Overcoming the odds: High-risk children from birth to adulthood*. Ithaca, NY: Cornell University Press.

Westfall, A., & Pisapia, J. (1994). *Students who defy the odds: A study of resilient at-risk students*. Richmond, VA: Metropolitan Educational Research Consortium.

Willett, J. B., Singer, J. D., & Martin, N. C. (1998). The design and analysis of longitudinal studies of development and psychopathology in context: Statistical models and methodological recommendations. *Development and Psychopathology*, *10*, 395-426.

Appendix A

LA's BEST Statement of Operations and Fund Balance

	1994 (Ending June 30, 1994)		
	Plan	Actual	Variance
Beginning Fund Balance		169,294	
Renevues and Support			
CRA Funds	2,327,876	1,896,336	(431,540)
Donations - Unrestricted	178,837	106,450	(72,387)
Donations - Restricted	40,017	53,343	13,326
Special Events (Net)	12,500	33,916	21,416
Convention Center GALA (Net)		556,120	556,120
Interest Income	2,500	4,254	1,754
Investment Income			
Total Revenues and Support	2,561,730	2,650,419	88,689
Expenditure			
Direct Cost			
Personnel	1,538,602	1,382,227	156,375
Benefits	162,270	144,365	17,905
Nutrition	81,876	40,128	41,748
Program Supplies	111,378	94,438	16,940
Program Equipment	17,164	10,087	7,077
Bus Services	31,104	3,077	28,027
Alteration/Improvements	11,751	4,944	6,807
Milage	0	0	0
Telephone	2,330	2,124	206
Program Evaluation	0	0	0
Special Event	5,000	4,729	271
Performing Arts Support	62,515	35,218	27,297
Restricted Donation Project	10,212	53,343	(43,131)
Contingency	171,876	0	171,876
Direct Cost Subtotal	2,206,078	1,774,680	431,398
Administration			
Personnel	368,868	303,891	64,977
Benefits	49,009	46,778	2,231
Accounting Services	6,000	4,800	1,200
Office Expense	5,775	4,145	1,630
Communications	22,000	20,930	1,070
Milage & Telephone	12,000	1,908	10,092
Legal	2,500	0	2,500
Miscellaneous	5,000	1,407	3,593
Consultants			
Restricted Donation Project	0	0	0
Administration Subtotal	471,152	383,859	87,293
Total Expenses	2,677,230	2,158,539	518,691
Surplus (Deficit)	(115,500)	491,880	607,380
Ending Fund Balance		661,174	

Appendix B

Table B1

Scenario II - Annual Exposure

Treatment condition	Number of participants	Probability	Survival probability	Total crime rate
Control			94.1%	5.9%
Low engagement	1,225	49.8%	92.6%	7.4%
Medium engagement	793	32.3%	95.0%	5.0%
High engagement	440	17.9%	95.9%	4.1%
Total	2,458			

Table B2

Expected Crime Cost Per Student

Treatment condition	Low	High	Life
Control	4,888	19,668	64,776
Low engagement	6,176	24,852	81,847
Medium engagement	4,186	16,844	55,475
High engagement	3,417	13,750	45,284

Table B3

Net Expected Avoided Crime Cost Per Student

Treatment condition	Low	High	Life
(vs. Control)			
Low engagement	-1,288	-5,183	-17,071
Medium engagement	702	2,824	9,300
High engagement	1,471	5,918	19,492
(vs. Low engagement)			
Medium engagement	1,990	8,007	26,371
High engagement	2,759	11,102	36,563

Table B4

Expected Value of Avoided Costs

Treatment condition	Low	High	Life
(vs. Control)			
Low engagement	-642	-2,583	-8,508
Medium engagement	226	911	3,000
High engagement	263	1,059	3,489
Expected value vs. control	-152	-613	-2,018
(vs. Control)			
Medium engagement	451.3	1,816.2	5,981.5
High engagement	351.3	1,413.8	4,656.1
Expected value vs. control	802.7	3,230.0	10,637.6
(vs. Low engagement)			
Medium engagement	1,279.8	5,149.9	16,960.7
High engagement	984.5	3,961.7	13,047.6
Expected value vs. low eng.	2,264.3	9,111.6	30,008.4

Appendix C

Benefit Cost Table 51 in detail

Benefit/Cost Ratios by Cost Assumption

Treatment condition	Exposure = 1 year			Exposure = 2 year			Exposure = 3 year			Exposure = 4 year		
	Low	High	Life	Low	High	Life	Low	High	Life	Low	High	Life
(vs. Control)												
Low engagement	-0.61	-2.47	-8.15	-1.27	-5.10	-16.81	-2.00	-8.06	-26.56	-2.83	-11.39	-37.52
Medium engagement	0.63	2.53	8.33	0.34	1.36	4.48	0.01	0.04	0.12	-0.36	-1.46	-4.79
High engagement	0.57	2.29	7.53	0.44	1.75	5.78	0.29	1.15	3.79	0.12	0.47	1.55
Expected value vs. control	0.58	2.34	7.72	-0.49	-1.99	-6.55	-1.71	-6.87	-22.64	-3.08	-12.38	-40.76
(vs. Control)												
Medium engagement	1.25	5.04	16.61	0.67	2.71	8.93	0.02	0.07	0.25	-0.72	-2.90	-9.56
High engagement	0.56	2.26	7.44	0.63	2.55	8.41	0.72	2.88	9.48	0.81	3.25	10.69
Expected value vs. control	1.81	7.30	24.05	1.31	5.26	17.33	0.73	2.95	9.73	0.09	0.34	1.13
(vs. Low engagement)												
Medium engagement	2.05	8.23	27.12	2.31	9.30	30.61	2.60	10.48	34.52	2.93	11.80	38.86
High engagement	1.57	6.33	20.85	1.78	7.15	23.55	2.01	8.07	26.58	2.26	9.09	29.95
Expected value vs. low engagement	3.62	14.57	47.97	4.09	16.45	54.17	4.61	18.55	61.09	5.19	20.89	68.81

Appendix D

HLM models for Achievement, using samples 1 and 2

Level 1 Model Specification

Model (1) Basic model: Controlling by the tests' indicators and polynomial covariates

Level-1 Model
Y = P0 + P1*(YEAR2) + P2*(YEAR22) + P3*(YEAR222) + P4*(CTBS) + P5*(CAT6) + E

Level-2 Model
P0 = B00 + R0
P1 = B10
P2 = B20
P3 = B30
P4 = B40
P5 = B50

Level-3 Model
B00 = G000 + U00
B10 = G100
B20 = G200
B30 = G300
B40 = G400
B50 = G500

Where:
"year2" is centered around year 1998 and captures the linear effect. "Year22" captures the quadratic term, and finally "year222" the cubic trend.

Level-1 Predictors are uncentered.
The Intercept represents the expected SAT9 achievement score in 1998.

Model (2): Including the tests' dummies, the polynomial indicators, and the time variant covariates (LEP, RFEP –reference group EO, and TRAVEL)

Level-1 Model
Y = P0 + P1*(YEAR2) + P2*(YEAR22) + P3*(YEAR222) + P4*(CTBS) + P5*(CAT6) + P6*(TRAVEL) + P7*(LEP) + P8*(RFEP) + E

Level-2 Model
P0 = B00 + R0
P1 = B10

P2 = B20
P3 = B30
P4 = B40
P5 = B50
P6 = B60
P7 = B70
P8 = B80

Level-3 Model
B00 = G000 + U00
B10 = G100
B20 = G200
B30 = G300
B40 = G400
B50 = G500
B60 = G600
B70 = G700
B80 = G800

The Intercept represents the expected SAT9 achievement score in 1998, for students who did not participate in the travel program and who are EO students.

Intermediate Model: Testing Random Coefficients

Level-1 Model
Y = P0 + P1*(YEAR2) + P2*(YEAR22) + P3*(YEAR222) + P4*(CTBS)
+ P5*(CAT6) + P6*(TRAVEL) + P7*(LEP) + P8*(RFEP) + E

Level-2 Model
P0 = B00 + R0
P1 = B10 + R1
P2 = B20
P3 = B30
P4 = B40
P5 = B50
P6 = B60
P7 = B70
P8 = B80

Level-3 Model
B00 = G000 + U00
B10 = G100 + U10
B20 = G200
B30 = G300
B40 = G400

B50 = G500
B60 = G600
B70 = G700 + U70
B80 = G800 + U80

Model 3a: Testing level 2 covariates on the intercept

Level-1 Model
Y = P0 + P1*(YEAR2) + P2*(YEAR22) + P3*(YEAR222) + P4*(CTBS)
+ P5*(CAT6) + P6*(TRAVEL) + P7*(LEP) + P8*(RFEP) + E

Level-2 Model
P0 = B00 + B01*(COHORT2) + B02*(FEMALE) + B03*(HISPANIC) + B04*(BLACK)
+ B05*(ASIAN) + B06*(OTHER) + B07*(EVERGATE) + B08*(YEARSLUN)
+ B09*(EVERDSP) + B010*(PEDUHI2) + B011*(EVERRET) + B012*(TRACK_A) + R0

P1 = B10 + B11*(COHORT2) + R1
P2 = B20
P3 = B30
P4 = B40
P5 = B50
P6 = B60
P7 = B70
P8 = B80

Level-3 Model
B00 = G000 + U00
B01 = G010
B02 = G020
B03 = G030
B04 = G040
B05 = G050
B06 = G060
B07 = G070
B08 = G080
B09 = G090
B010 = G0100
B011 = G0110
B012 = G0120
B10 = G100 + U10
B11 = G110
B20 = G200
B30 = G300
B40 = G400
B50 = G500

B60 = G600
B70 = G700 + U70
B80 = G800 + U80

MODEL 3B: Testing level-2 covariates on the intercept and slope of year (FINAL MODEL 3)
Original model omitted.

Final Model:
Level-1 Model
Y = P0 + P1*(YEAR2) + P2*(YEAR22) + P3*(YEAR222) + P4*(CTBS)
+ P5*(CAT6) + P6*(TRAVEL) + P7*(LEP) + P8*(RFEP) + E

Level-2 Model
P0 = B00 + B01*(COHORT2) + B02*(FEMALE) + B03*(HISPANIC) + B04*(BLACK)
+ B05*(ASIAN) + B06*(OTHER) + B07*(EVERGATE) + B08*(YEARSLUN)
+ B09*(EVERDSP) + B010*(PEDUHI2) + B011*(EVERRET) + B012*(TRACK_A) + R0

P1 = B10 + B11*(COHORT2) + B12*(FEMALE) + B13*(HISPANIC) + B14*(BLACK)
+ B15*(ASIAN) + B16*(OTHER) + B17*(YEARSLUN) + B18*(EVERRET) + R1
P2 = B20
P3 = B30
P4 = B40
P5 = B50
P6 = B60
P7 = B70
P8 = B80

Level-3 Model
B00 = G000 + U00
B01 = G010
B02 = G020
B03 = G030
B04 = G040
B05 = G050
B06 = G060
B07 = G070
B08 = G080
B09 = G090
B010 = G0100
B011 = G0110
B012 = G0120
B10 = G100 + U10
B11 = G110
B12 = G120
B13 = G130

B14 = G140
B15 = G150
B16 = G160
B17 = G170
B18 = G180
B20 = G200
B30 = G300
B40 = G400
B50 = G500
B60 = G600
B70 = G700 + U70
B80 = G800 + U80

Table D1

Deviance Statistics

	Model 1 Basic level 1	Model 2 Time variant covariates	Model 3 Level-2 covariates	Model 4 Level-3 covariates
Read lb1	259470.372993	220834.255777	217281.654300	217266.6307
Math lb1	259574.809986	220150.749069	216545.104758	216538.444403
Read lb2	245724.885409	213072.495180	209694.542547	209684.721223
Math lb2	245874.480142	212559.787836	209016.237800	209015.520686
Parameters	9	13	43	46

Model 4: Including Level 3 predictors

Level-3 Model
B00 = G000 + G001(MOVOLUHO) + U00
B01 = G010
B02 = G020
B03 = G030
B04 = G040
B05 = G050
B06 = G060
B07 = G070
B08 = G080
B09 = G090
B010 = G0100
B011 = G0110
B012 = G0120
B10 = G100 + G101(PCTWHITE) + G102(MOVOLUHO) + U10
B11 = G110
B12 = G120
B13 = G130
B14 = G140
B15 = G150
B16 = G160
B17 = G170
B18 = G180

Model 5: Including the treatment predictors at level 2

1. The average treatment effect is included
2. Duration of attendance to the program is included
3. Daily attendance (in log form) in included
4. Finally duration and intensity are included in the model simultaneously to explore the effect of attending the same number of days over different number of years and vice versa.

Level-1 Model: same
Level-2 Model
P0 = B00 + B01*(LABEST1) + B02*(COHORT2) + B03*(FEMALE) + B04*(HISPANIC)
+ B05*(BLACK) + B06*(ASIAN) + B07*(OTHER) + B08*(EVERGATE)
+ B09*(YEARSLUN) + B010*(EVERDSP) + B011*(PEDUHI2) + B012*(EVERRET)
+ B013*(TRACK_A) + R0
P1 = B10 + B11*(LABEST1) + B12*(COHORT2) + B13*(FEMALE) + B14*(HISPANIC)
+ B15*(BLACK) + B16*(ASIAN) + B17*(OTHER) + B18*(YEARSLUN)
+ B19*(EVERRET) + R1
P2 = B20
P3 = B30
P4 = B40
P5 = B50
P6 = B60
P7 = B70
P8 = B80

Level-3 Model:
B00 = G000 + G001(MOVOLUHO) + U00
B01 = G010
B02 = G020
B03 = G030
B04 = G040
B05 = G050
B06 = G060
B07 = G070
B08 = G080
B09 = G090
B010 = G0100
B011 = G0110
B012 = G0120
B013 = G0130
B10 = G100 + G101(PCTWHITE) + G102(MOVOLUHO) + U10
B11 = G110
B12 = G120
B13 = G130
B14 = G140

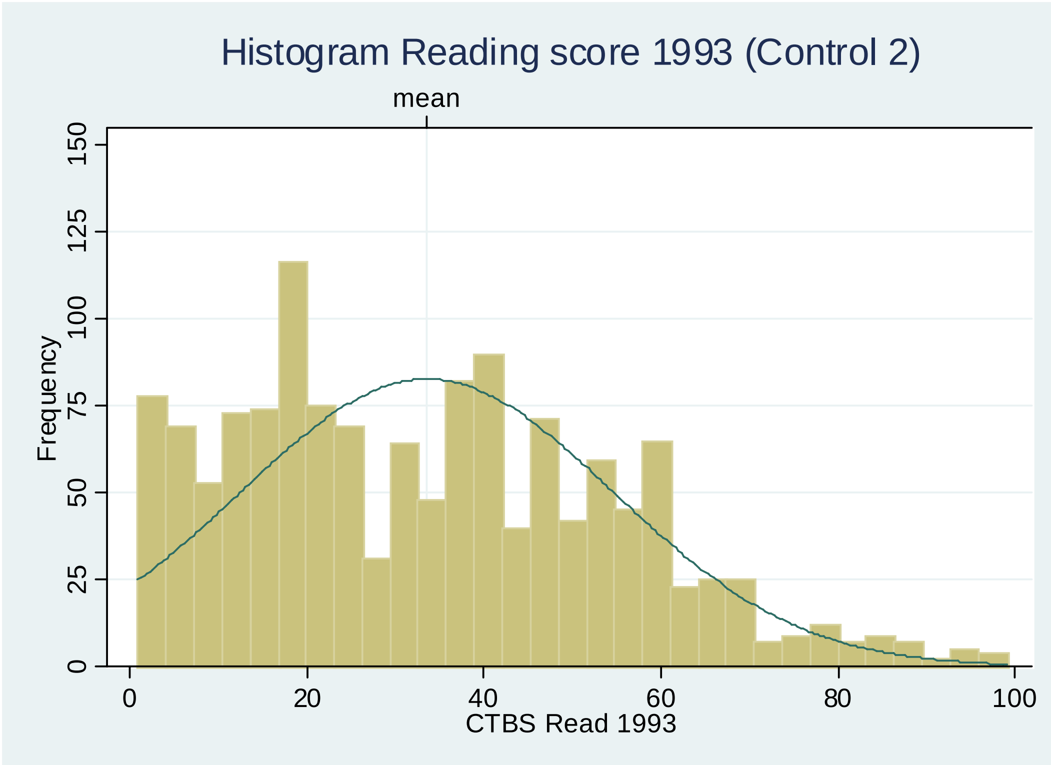
B15 = G150
B16 = G160
B17 = G170
B18 = G180
B19 = G190
B20 = G200
B30 = G300
B40 = G400
B50 = G500
B60 = G600
B70 = G700 + U70
B80 = G800 + U80

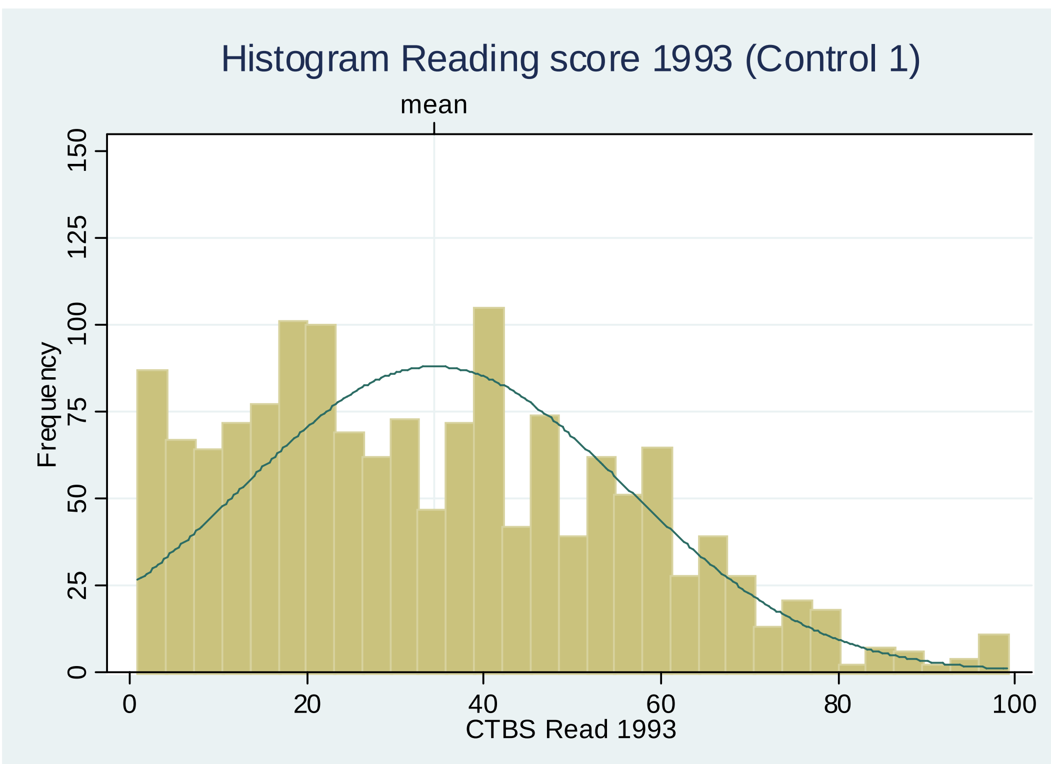
Table D2

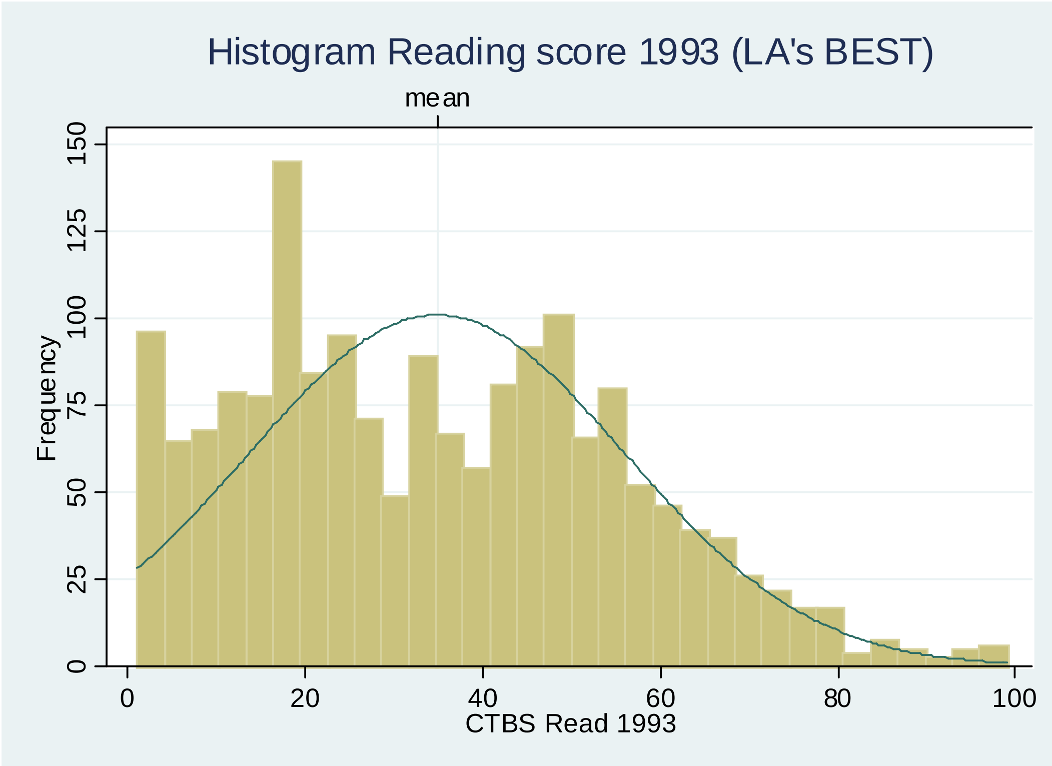
Deviance Statistics of the Models that Include the Treatment Predictors

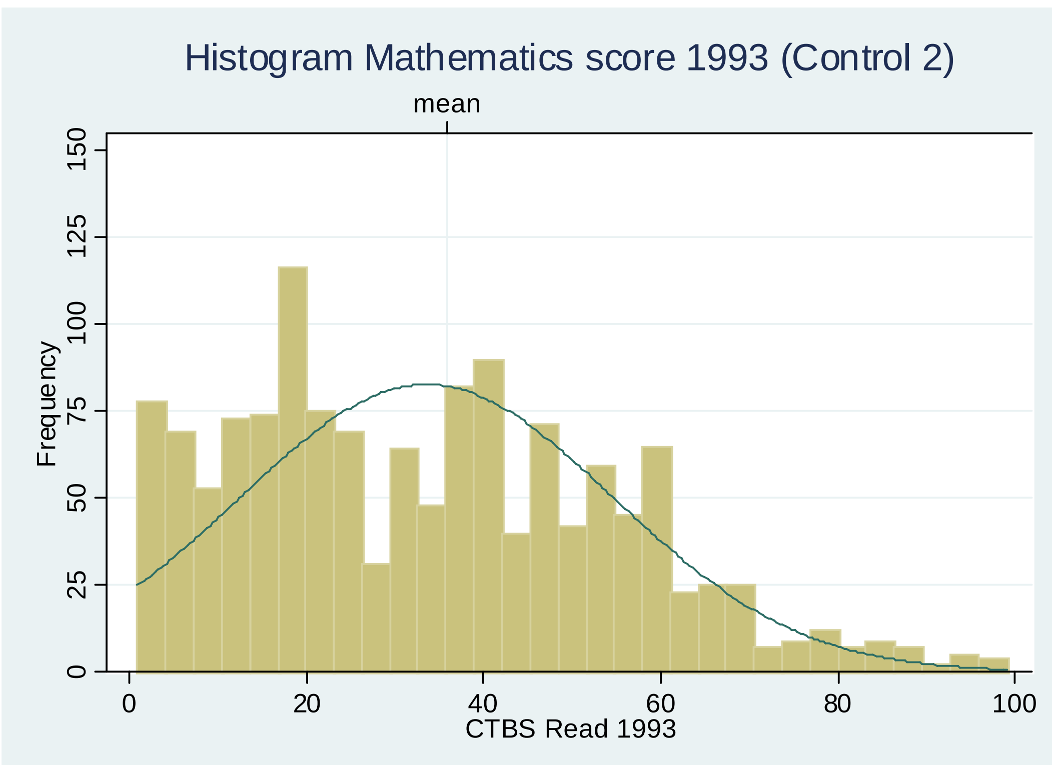
Models	Model 5A "Labest"	Model 5B "durat"	Model 5C "intensity"	Model 5D "Durat & Intensity"
Read Sample 1	217263.4551	217262.6774	217262.6321	217261.1964
Math Sample 1	216535.1454	216536.3498	216534.1207	216533.3825
Read Sample 2	209677.1216	209682.8113	209674.4670	209669.5697
Math Sample 2	209007.0519	209013.3093	209002.3773	208998.5527
Parameters	48	48	48	50

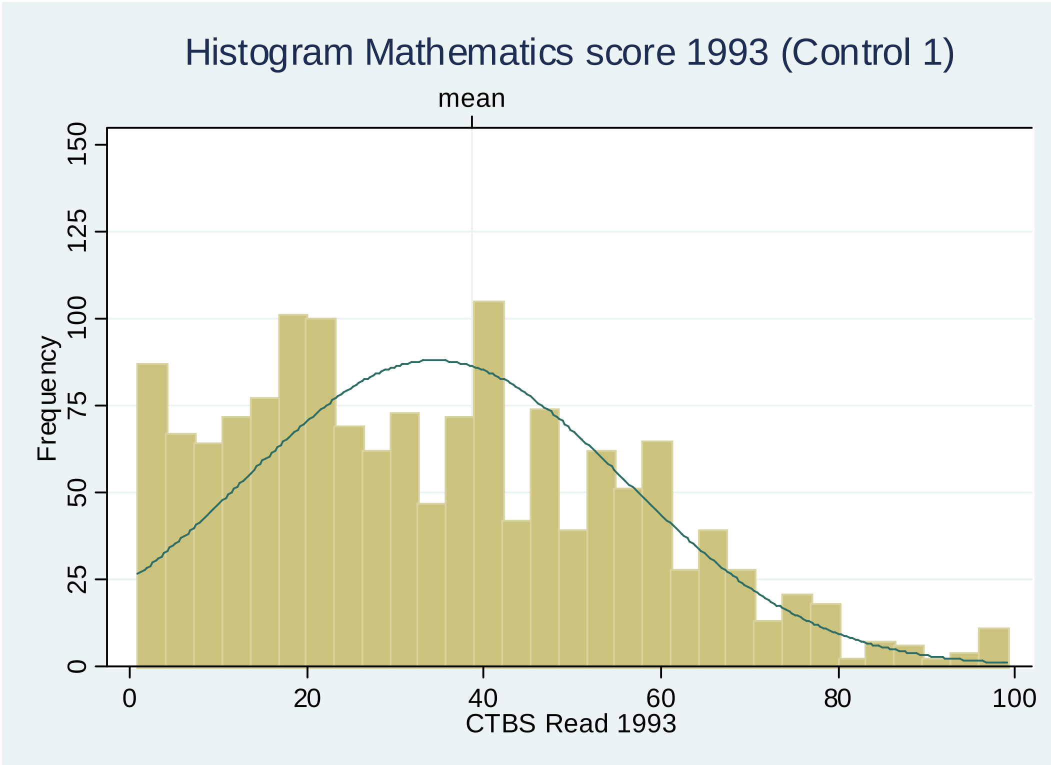
Note. Sample 1 includes only LA's BEST schools. Sample 2 include LA's BEST and non-LA's BEST schools.

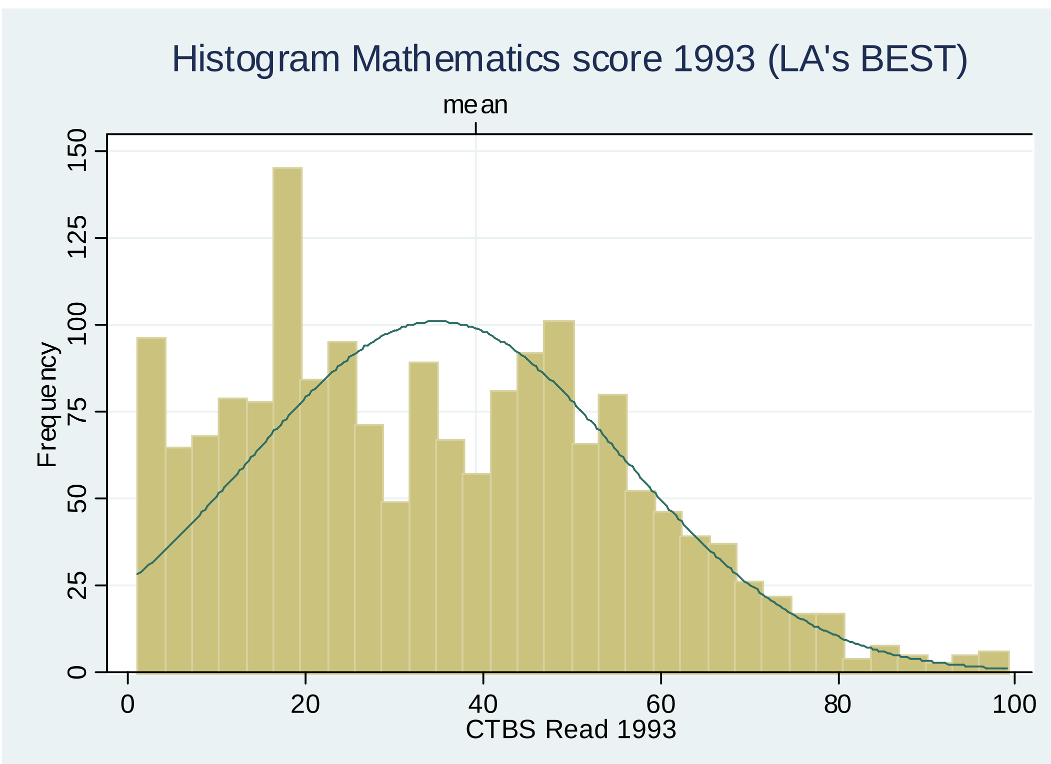

Histogram Reading score 1993 (Control 2)
mean
Frequency
0
25
50
75
100
125
150
0
20
40
60
80
100
CTBS Read 1993


Histogram Reading score 1993 (Control 1)
mean
Frequency
0
25
50
75
100
125
150
0
20
40
60
80
100
CTBS Read 1993


Histogram Reading score 1993 (LA's BEST)
mean
Frequency
0
25
50
75
100
125
150
0
20
40
60
80
100
CTBS Read 1993


Histogram Mathematics score 1993 (Control 2)
mean
Frequency
0
25
50
75
100
125
150
0
20
40
60
80
100
CTBS Read 1993


Histogram Mathematics score 1993 (Control 1)
mean
Frequency
0
25
50
75
100
125
150
0
20
40
60
80
100
CTBS Read 1993


Histogram Mathematics score 1993 (LA's BEST)
mean
Frequency
0
25
50
75
100
125
150
0
20
40
60
80
100
CTBS Read 1993

Appendix E
A Brief History of LA's BEST

Originated with 10 sites in 1988, LA's BEST has grown into 24 sites by 2000 and followed with a period of rapid expansion between the year 2000 to 2005, during this period LA's BEST has expanded the program from 24 sites to 168 sites[55]. A new infrastructure was needed to accommodate this rapid expansion. Realizing this need, Bain & Company has donated half a million dollar worth of pro bono strategic planning for LA's BEST. Part of the strategic plan included the transition of the Board of Directors into the Governing and Advisory Boards. The Governing Board now has the fiduciary responsibilities and policy authority (above and beyond the LAUSD policy authority) and the Advisory Board has the program authority (above and beyond any LAUSD curriculum standards). The management team was also separated into the corporate office and the operations office. The corporate office generates fundraising events and writes proposals securing grants to support operations programming. It is also responsible for generating language for major after school legislation (both state and federal) meanwhile, producing quarterly newsletters, event programs as needed, and an annual report to keep all parties informed. The operations office manages the site staff and coordinates the day-to-day activities that occur on sites. System-wide decision-making is co-managed by corporate leaders in a situational way, primarily involving the corporate and operations offices; for example CEO and COO, or Deputy Administrator and Grant Manager, etc.

Both the corporate and operations offices are lead by the president and CEO, whose major responsibilities are to provide strategic leadership and stewardship of LA's BEST, including reporting to the Board of Directors; providing oversight for the design and management of all programs and initiatives; engaging support, and fiscal and financial resources; promoting community and institutional collaboration; and directing media, community and public relations. These duties are performed with the support of the following corporate staff members: Deputy Administrator, Chief Financial Officer, Director of Fund Development, Director of Communications, Director of Community

[55] LA's BEST is currently operating 178 sites. By the end of 2007, LA's BEST is expected to be operating 180 sites.

Outreach, After School Arts Program (ASAP) Consultant, BEST Friends Coordinator (Harvard fellow for 1 year), and their associates and assistants.

For the purpose of this study, and to illustrate the share of responsibilities at LA's BEST, an abbreviated organization chart is presented in Figure D1.

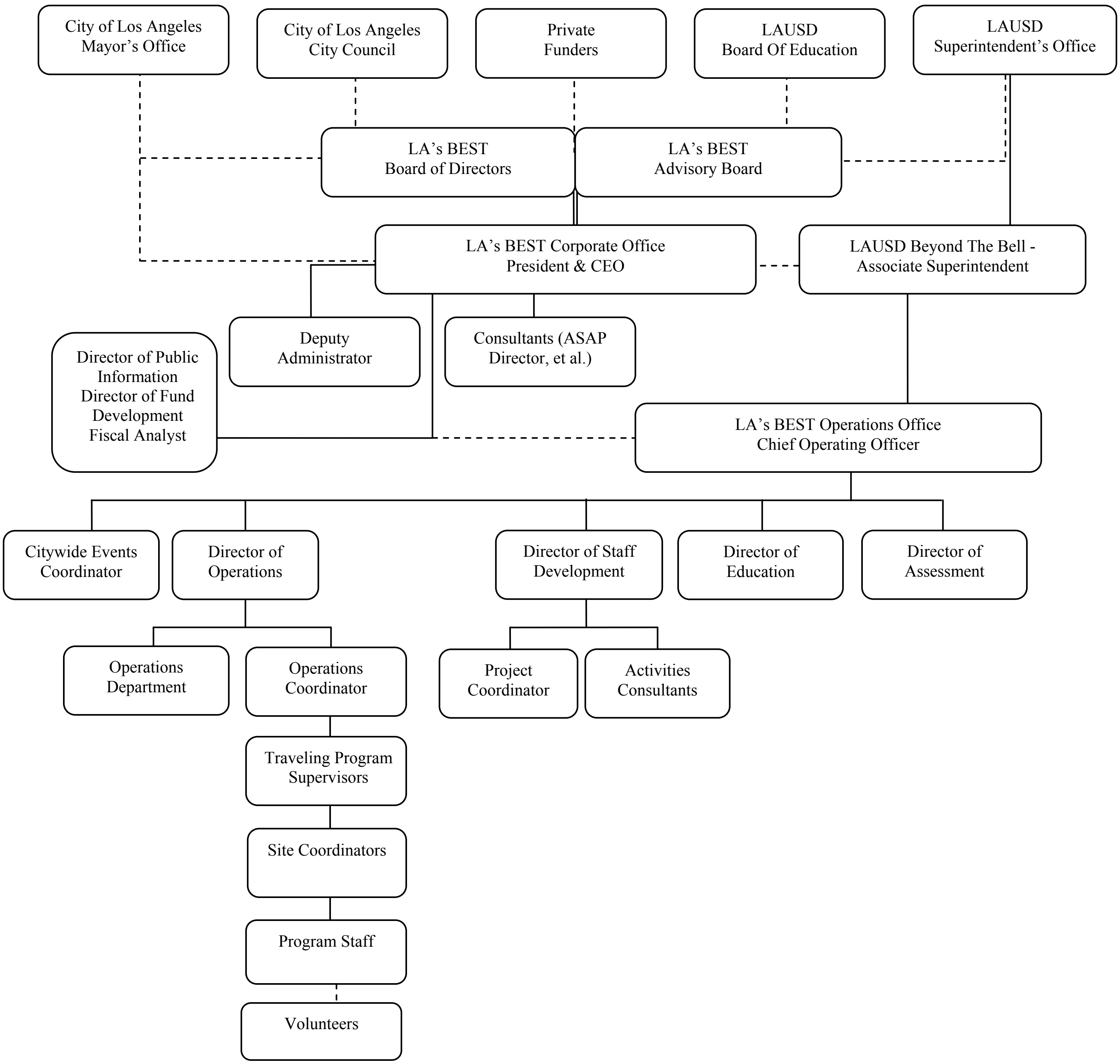


Figure D1. LA's BEST Modified Organization

LA's BEST Corporate Management Staff is located in the Office of the Mayor[56] and closely works with the Operations Management staff located within Los Angeles Unified School District (LAUSD). The field staff includes: 1) a site coordinator[57] for every site, with the ratio of one adult site staff to every 20 students; and 2) traveling program supervisors and activities consultants[58] assigned to clusters of five or six sites. Field staff are employees of the Los Angeles Unified School District, they must be finger printed, obtained a tuberculosis clearance, and pass a criminal background check. They are predominantly Hispanic and African American, with over half living in the local community. This is a deliberate attempt to create site cultures that reflect the daily experiences of the participating students. The field staff also tend to be female, under 25 years of age, and college enrolled. To ensure continuity of the program culture, climate, and mission, supervisory positions such as site coordinators and traveling staff are usually promoted from within. Many of the field staff also work as educational aides or assistants with the District or their home school (see annual report for LA's BEST 1999-2000). This involvement is highly encouraged by LA's BEST both as a relationship building procedures with the school staff and also to provide connections and facilitate communications between school and afterschool staff.

For professional development, LA's BEST offers a staff training day once per year where personnel from all of the sites can take workshops on topic ranging from activity ideas to educational psychology concepts (i.e. motivation, behavior management, and so forth). Mandatory training sessions on specific content areas are provided for staff who teach sports, math, science and other subject matter. In addition, traveling program supervisors and activities consultants serve as mentors and provide an on-going personal training and support to the site staff.

---

[56] The Office of the Mayor has also developed the Mayor's council which is primarily responsible for the structure and implementation of LA's Best at elementary schools in Los Angeles; this has helped to expand the program all throughout the city of Los Angeles.

[57] The field staff members are often predominantly Hispanic and African-American, with over half of them living within the communities they serve. Staff also tend to be female, under the age of 25 years old, and are college-enrolled. Many of the site staff work as educational aides or assistants with LAUSD (1999-2000 annual report for LA's Best). A strict background security check is placed upon them for employment.

[58] The traveling staff are provided with regular trainings in management and curriculum and act as support and supervisors for their clusters. They also served as the liason between the site staff and the operation and corporate offices.

Voluntary services[59] are recruited from the community and society in general. Volunteers aid LA's BEST in three different capacities: 1) academic tutoring during homework sessions; 2) assistance with activities during club or enrichment sessions; and/or 3) assistance in all other aspects for further program enhancement(http://www.lasbest.org/volunteers/opportunities.php).

Volunteers include people from the local community, local colleges/universities, non-profit and civic organizations. They provide an additional source of supervision in areas where the staff may be lacking or in need of assistance. Due to their voluntary status, background checks are not as stringent for volunteer staff.

Figure D2 provides the vision, mission, and values statements for LA's BEST, which plays a critical role in guiding the LA's BEST organization.

Vision
All children need a safe place to be after school with caring, responsible adults and engaging activities that connect each child to his/her school, family and community.

Mission
The mission of LA's BEST is to provide a safe and supervised after school education, enrichment and recreation program for elementary school children, ages 5 to 12, in the City of Los Angeles.

Values
Nothing we do is as important as the effect it has on a child.

Engaging activities develop values, skills, and relationships. Activities are not seen as ends in themselves, but as vehicles for creating values, building skills, and solidifying peer and adult relationships. An engaging activity is one that holds children's attention, awakens their imagination, and inspires them to want to learn more.

All children have equal rights to be accepted, respected, and valued by others. Children are viewed as individuals to be developed, not problems to be solved.

Children should be involved in decision making and program design. If children get to choose how, when, in what, and with whom to be engaged, they are far more likely to enjoy themselves and behave cooperatively.

When we listen for understanding, everyone learns — children and adults alike. We are constantly able to learn from our children as well as each other. Everyone is a learner.

Figure D2. LA's BEST's Vision, Mission, and Values

[59] Voluntary services are not paid.