

CRESST REPORT 749

Denise Huang
Seth Leon
Deborah La Torre
Sima Mostafavi

EXAMINING THE RELATIONSHIP
BETWEEN LA'S BEST
PROGRAM ATTENDANCE AND
ACADEMIC ACHIEVEMENT
OF LA'S BEST STUDENTS

DECEMBER, 2008



National Center for Research on Evaluation, Standards, and Student Testing

Graduate School of Education & Information Studies
UCLA | University of California, Los Angeles

**Examining the Relationship between LA's BEST Program Attendance and
Academic Achievement of LA's BEST Students**

CRESST Report 749

Denise Huang, Seth Leon, Deborah La Torre, & Sima Mostafavi
CRESST/University of California, Los Angeles

December, 2008

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
300 Charles E. Young Drive North
GSE&IS Building, Box 951522
Los Angeles, CA 90095-1522
(310) 206-1532

Copyright © 2008 The Regents of the University of California

The work reported herein was supported by grant number 021891 from LA's BEST with funding to the National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

The findings and opinions expressed in this report are those of the authors and do not necessarily reflect the positions or policies of LA's BEST.

Examining the Relationship between LA's BEST Program Attendance and Academic Achievement of LA's BEST Students

Denise Huang, Seth Leon, Deborah La Torre & Sima Mostafavi
CRESST/University of California, Los Angeles

Abstract

Researchers and policymakers are increasingly interested in the impact of afterschool programs on youth development. Even though numerous studies have investigated the impact of afterschool participation on academic outcomes, there is limited research on the differential impact of afterschool programs based on students' participation rate. This study bridges that research gap and presents results from a study of the effectiveness of the LA's BEST afterschool program based on different levels of student participation. This research tracked 4 years of the academic histories for two cohorts of students participating in LA's BEST. We separated the students in each cohort into four categories based on their intensity of attendance in LA's BEST and then used a propensity based weighting method to remove existing differences in student background characteristics. Hierarchical growth modeling was employed to analyze the academic outcomes. Results indicate that math achievement outcomes of students vary by intensity of program participation. Student participants who attended LA's BEST over 100 days per year demonstrated greater math achievement growth than students with low program attendance. This finding was consistent, and was statistically significant, for both cohorts of students. In contrast, although the trend for English-language arts achievement growth was positive, and followed a developmental pattern similar to math, it did not vary significantly by intensity of program participation. This finding was also consistent for both cohorts of students.

Chapter I: Introduction

In recent years, interest and funding in afterschool programs has increased significantly. For example, California increased its yearly budget for afterschool programs from 120 to 550 million during the 2006–07 fiscal year (California AfterSchool Network, 2007).¹ As a result, funders and policymakers are demanding greater accountability of programs.

Ever since the enactment of No Child Left Behind Act in 2001 (NCLB; 2002), achievement gains resulting from afterschool participation have been of particular interest (Lauer, et al., 2006; Miller 2003). However, findings have been inconsistent (Fashola, 1998;

¹ As mandated by Proposition 49, funding for afterschool programs was increased once the California state budget reached a level making the release of funds feasible (California AfterSchool Network, 2007).

Vanderhaar & Muñoz, 2006). The challenge for researchers is partly due to the wide variation of program goals, difficulty in obtaining valid control groups, the inherent potential of selection bias in the afterschool population, difficulty in obtaining clean records of data, the high transience rates of the students, and in particular, the failure to differentiate among the dosage² students receive (Lauer et al., 2003).

For any intervention project, it is necessary for subjects to receive adequate treatment in order to demonstrate effects. However, afterschool studies rarely examine the importance of dosage. When conducting program evaluations, it is very common for studies to group all participants (regardless of attendance days) as the treatment group and the non-participants (without proper control of pre-existing differences) as the comparison group. Thus, although inconsistent results can stem from many factors, such as those previously mentioned, failing to consider the “dosage” effect is one of the most important. Whereas some students might have regular attendance at an afterschool program, others might “drop-in” as needed. In cases of inconsistent attendance, it is unrealistic to expect significant academic gains.

The goal of this study is to examine the long-term relationship between participation in LA’s BEST and academic achievement. Accordingly, the main research question for this study is as follows:

- Do the achievement outcomes of LA’s BEST students’ vary as a function of their different intensity levels of afterschool participation?

Only recently have researchers started examining the effects of dosage level on the academic outcomes of afterschool programs. These studies have found that students who attend afterschool programs more and experience more exposure, benefit more from the program (Lauer et al., 2003; McComb & Scott-Little, 2003; Frankel & Daley, 2007). The purpose of this study is to further this research by comparing students with different dosage levels using propensity score matching to reduce self-selection bias and to examine the students’ achievement trends over a period of 4 years.

Chapter II: Review of Literature

What academic outcomes are associated with afterschool program participation? When reviewing research on the academic impact of afterschool participation, results are mixed. It is not unusual to come across null (Lauver, 2002; Dynarski, et al., 2003; Vanderhaar &

² Feister, Simpkins and Bouffard (2005), define dosage as a measure of attendance intensity that focuses on the amount of time a participant attends a program within a specified period (e.g., hours per week, days per month, days in a year, etc.).

Muñoz, 2006) and positive (Redd, Cochran, Hair & Moore, 2002; Dynarski, et al., 2003) results.

Studies of afterschool programs have reported on an array of positive academic outcomes. Bergin, Hudson, Chryst & Resetar (1992) found positive associations between afterschool participation and higher achievement scores. Their study followed a group of kindergartners who attended an afterschool program and compared them to a control group. Initially, the standardized test scores of both groups were below national average. However, by the spring of first grade, the treatment group was outperforming the control group and was performing above national norms. Afterschool participation is also associated with higher classroom grades, higher math and reading scores, increased day school attendance, lower dropout rates, higher homework completion rates, and higher graduation rates (Goerge, Cusick, Wasserman & Gladden, 2007; Little & Harris, 2003; Sheley, 1984). While at the same time, others have reported mixed, insignificant or negative outcomes regarding academic performance³, school retention, feelings of safety, and behavior to name a few (Cooper, Charlton, Valentine & Muhlenbruck, 2000; Dynarski, et al., 2003; James, 1997; Vanderhaar & Muñoz, 2006). However, most studies that have evaluated dosage level have found positive effects for students who attend more consistently (McComb & Scott-Little, 2003). The following sections discuss two important considerations in conducting afterschool research and evaluations.

The Dosage Effect

Dosage effect is a critical factor to examine when assessing the effect of an intervention. More specifically, examining dosage helps to determine whether participants are receiving sufficient treatment in order to demonstrate effect. Even though dosage (defined as intensity of participation from here on) is very important in determining program success, recent literature on afterschool programs has only begun to investigate this issue. In general, these studies have found a positive relationship between intensity of participation and positive student outcomes. For instance, Frankel and Daley (2007) found that higher afterschool attendance is associated with higher academic achievement, while Goldschmidt, Huang and Chinen (2007), found that medium (10–14 days per month) and high attendance (15 or more days per month) in an afterschool program is associated with lower juvenile crime rate. In recent years, multiple studies have also found a relationship between afterschool attendance intensity and higher day school attendance (Frankel & Daley, 2007;

³ Some studies looking at both math and reading outcomes have only found effects for math whereas others have only found effects for reading

Huang, Gribbons, Kim, Lee & Baker, 2000; Welsh, Russell, Williams, Reisner & White, 2002; Munoz, 2002).

Specifically, in 2007, Frankel and Daley released a report that found an association between high dosage of afterschool participation and higher math assessment scores, English-language arts assessment scores, and day school attendance. They created four attendance level categories: 1–20 days, 21–50 days, 51–100 days, and more than 100 days per year. They found that, in order to benefit academically, the elementary school students needed to attend the afterschool program for at least 100 days per year and middle school students needed to attend at least 50 days annually.

Similarly, Jenner and Jenner (2007) examined the impact of program participation intensity on academic outcomes. They found a linear and positive relationship between participation level and academic outcomes such as math, reading, language arts, and science scores. Their analyses placed the minimum attendance level necessary for measuring impact at 30 days annually.

Along the same lines, Muñoz (2002) looked at afterschool program participation and student outcomes among inner city students in Louisville, Kentucky. The author established two afterschool program attendance level categories using the mean number of visits by all participants. He found a positive relationship between intensity of afterschool program attendance and day school attendance. In addition, he found non-significant correlations between higher afterschool intensity and lower suspensions as well as greater GPA.

Intensity level of afterschool attendance can also predict social outcomes; for instance, Goldschmidt, Huang, and Chinen (2007) examined the long-term effectiveness of afterschool programs in lowering juvenile crime rates. They found that students who consistently attended LA's BEST demonstrated a substantive significant reduction in juvenile crime as compared to students with inconsistent attendance and no attendance.

Additionally, a meta-analysis (Lauer et al., 2003) examining 35 out-of-school time (OST) programs⁴ for assisting at-risk students in reading and/or math identified the duration of OST as a moderator. They found that for both reading and math, effect sizes were larger for OST programs that were more than 45 hours annually. Unlike the four studies previously mentioned, Lauer and colleagues (2006) looked at program duration instead of students' program attendance. They did this because of incomplete access to attendance data. This

⁴ Out-of-school time refers to activities that children participate in when they are not in school and that are not mandated by school attendance. This may include before school, afterschool, and summer programs (Lauer, et al., 2006).

study defined program duration as the total number of hours that the program was offered to participants rather than the number of days students attended.

Finally, in reviewing research on participation and outcomes in afterschool programs, McComb and Scott-Little (2003) concluded that students who attend afterschool programs more frequently and for longer periods benefit the most. They suggest that afterschool programs should be an integral part of school's academic and developmental programs. They stated that in all cases where data was examined using "intensity" level, results favored students who had participated at higher levels.

Reducing Selection Bias

Another frequent critique of afterschool studies is selection bias (Fashola, 1998; Hollister, 2003; Little & Harris, 2003; Scott-Little, Hamann & Jurs, 2002). Because afterschool program participation is voluntary, students self-select themselves into participation and non-participation groups.⁵ In comparing participating students to non-participating students in the same school, there are inherent biases that researchers need to balance or control in order for the findings to be valid. Furthermore, due to the social context of afterschool programs, reaching the "gold standard" of research is difficult. According to the American Institutes for Research (2002), the "gold standard" is research that meets all of the standards of scientifically based research as called for in the NCLB Act (2002). This includes the use of experimental designs, including randomization and control groups. In reality, it is often difficult, and potentially unethical, for most afterschool programs to randomize their participants unless the programs are grossly oversubscribed. For, unless programs have many more applicants than available spaces, random assignment would mean refusing to accept some students into the program so that they could serve as controls. Students who are refused enrollment may end up unsupervised and without the homework help they desperately need. As a result, many studies lack either true experimental control or a valid comparison group. Thus, most studies in this field are quasi-experimental, with researchers using a comparison group and making use of statistical controls. In these quasi-experimental studies, one needs to be cautious when inferring causality. With this in mind, the present study reduces self-selection bias by removing pre-existing category differences using propensity scores. Propensity scores are estimated in order to account for potential differences in student background characteristics, such as gender and ethnicity. By reducing initial differences across different groups, one can more confidently attribute differences in achievement outcomes to treatment intensity.

⁵ Parents may also choose to enroll or "select" their children for participation.

In summary, although many researchers indicate that afterschool programs are a potentially powerful resource that can help increase student' academic achievement, the reported findings on academic outcomes are mixed. In our brief review of literature, we found that many studies that claim positive outcomes reported academic improvement in students with a higher dosage of afterschool participation, and those that reported null or negative findings more often looked at participants of afterschool as an aggregated group. Recently, researchers have begun to examine the relationship between regular afterschool participation and academic outcomes. Even as the literature states that quality afterschool programs can teach students the academic and social skills that they need to avoid anti-school behaviors and contribute to academic resiliency, sufficient exposure to effective afterschool environments is necessary in order for students to reap the benefits. At the same time, although it is necessary to look at the intensity of participation as a contribution to student outcomes, it is also important to reduce the selection bias that is inherent in the field of afterschool research in order to add validity to the findings of the studies. This study intends to fill the research gap by examining the impact of differential intensity of exposure to afterschool programming, specifically LA's BEST, on student academic achievement and using propensity matching as a technique to reduce self-selection bias. First, we provide a brief description of the LA's BEST program.

The LA's BEST Program

Los Angeles Better Educated Students for Tomorrow (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 write an official 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 5 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 Learners comprise at least half of the student population from most sites. Of this population, the majority's primary language is Spanish; whereas the other percentage of the English Learner population is composed of those whose first language is of Asian or Pacific origin.

Parents often mention homework help and proper supervision as the primary incentives for enrolling their children. Teachers may also recommend students for 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.

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-being so that children can achieve their full potential (Schaps, 2006; Hodgkinson, 2006). At LA's BEST, their 3½ beats focus on the whole-child by emphasizing students' intellectual, social-emotional, and physical development.

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 and social-emotional development such as:

- Sense of safety & security – through providing students with a safe and nurturing environment.

- Healthy lifestyle – through curriculum and activities that promotes drug and gang prevention, healthy eating habits, and plenty of exercise.
- 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; each student can be actively engaged in learning activities that are connected to their school and broader community; And most importantly, each student has access to extra-curricular activities, academic enhancements, and qualified, caring adults.

Because 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; the organization gives each site autonomy to structure their own program, as long as the site coordinator and staff adhere to the foundational principles of LA’s BEST.⁶ As a result, each site has distinct characteristics and program themes (such as arts, self-esteem, conflict resolution, technology, etc). Subsequently, relationships with the day school, levels of school⁷ and community supports also tend to vary with each site (see Huang, et al., 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.

⁶ The snack and homework periods are the common components of all LA’s BEST sites. The education and enrichment sessions are grounded on the principles of being: (a) cognitive/ academic (activities in school subject matter; (b) recreational (physical fitness); and (c) part of the performing arts (i.e. dance, drama, etc.).

⁷ 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 the study of nutrition and healthy habits, exercise programs such as tennis and skating, and the BEST Fit community health fair.
- Community and Culture – 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 fall under four categories:
 - Celebrations, such as Halloween Kidfest, Community Jam, Awards Days.
 - Programs for children, including parent volunteers for daily activities, parent attendance of field trips.
 - Programs for parents, including parent workshops, guest speakers for parental education.
 - Communications/information, including open house events, assemblies, parent-teacher meetings.

The educational and enrichment activities mainly come from three different sources: (a) curricula purchased from education vendors, such as KidzLit⁸ and KidzMath⁹; (b) activities developed by the education and staff development departments at LA’s BEST operations; and, (c) activities designed by the site staff.¹⁰

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 feedback from staff, day school teachers, students and parents; they gauge the short and or long-term effects of 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.

⁸ 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.

⁹ 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.

¹⁰ Site staff members receive support from the program coaches and their site coordinators in developing and/or implementing activities.

Chapter III: Study Design and Methods

Since the formation of LA's BEST in 1988, the National Center for Research on Evaluation, Standards, & Student Testing (CRESST) has been conducting evaluations of the program. As a result, CRESST has established a longitudinal database on these students. The longitudinal database includes student demographics and academic information such as student achievement scores on English-language arts and mathematics standardized tests.

The basis for this study sample is comprised of the LAUSD student database 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 accomplished this task by going back through the historical records and tracking all available information for all students from the 2002–03 school year through the 2006–07 school year.

The following describes the study design and the data analysis strategies for this study.

Study Design

This study employs a quasi-experimental design that consists of a longitudinal sample of both academic and LA's BEST program attendance data. The sample is comprised of roughly 10,000 students from LA's BEST programs. The sample includes two cohorts of students with base years in 2002–03 and 2003–04. We separated students in each cohort into four categories based on their intensity of attendance in the LA's BEST program. We also employed a propensity based weighting method to minimize existing differences in student background characteristics across the four LA's BEST program attendance categories. Once this was completed, we took advantage of this panel structure and applied hierarchical growth modeling to academic outcomes. This method allowed us to examine students' academic growth while controlling for student and school-level background characteristics. Given that we had student background information, we also examined moderating factors such as gender, race/ethnicity, language proficiency, and socioeconomic status.

Data Analysis Methods

We utilized the longitudinal nature of the data and followed academic data 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.¹¹ Second, we separated initial status from growth, thus avoiding spurious negative correlations between where students start and their ensuing growth (Bloomquist, 1977).

Propensity scores were estimated to account for potential differences in student background characteristics. These scores are computed from a large reservoir of potential controls by applying a systematic weighting procedure. In other words, 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. Because we have four comparison groups ordered by attendance intensity, rather than a single treatment and a single control, we created a propensity scalar that corresponds with the ordered likelihood of belonging to one of the four intensity groups. We estimated this propensity scalar using ordinal logistic regression.

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. Because observations are nested within individuals, time intervals need not be constant across individuals as in traditional repeated measures analyses (Raudenbush & Bryk, 2002), and the number of observations per person may vary. Thus, this HLM design allows flexible specification of the covariance structure at every level of the analysis for this study (Snijders & Bosker, 1999).

The HLM analysis is based on a three-level model. Two separate models were conducted for each cohort, one for math and one for English-language arts. In these models, Level 1 represents time nested within students. There are four time points for each achievement model, with achievement at each time point serving as the outcome. Before specifying the growth models, we examined the overall achievement growth patterns to determine whether a quadratic or logarithmic transformation would provide a better fit than a simple linear model. Because neither transformation resulted in an improved fit, we modeled linear growth.

At Level 1 we model achievement to be predicted by time (school year). The Level 1 model has two coefficients for each child including an intercept and a slope. The intercept for this level is initialized at zero for the first time point. Level 2 accounts for student-level effects. At this level, the achievement intercept and the achievement slope over time are

¹¹ such as not requiring balanced data (Raudenbush & Bryk, 2002) and managing missing data due to attrition (Hox, 2002)

modeled as functions of LA's BEST attendance intensity and day school attendance. At Level 3, information regarding mean achievement at the school level is included in the model. Only the intercept and slope are allowed to vary randomly over this level. The slopes, due to the effect of LA's BEST attendance intensity and day school attendance, are assumed constant at this level of the model.

This model is performed on the weighted sample in which differences in background characteristics and the initial achievement outcomes across intensity levels have been removed. The primary relationship of interest is that between attendance intensity and the slope of achievement growth over time. The presence of a significant relationship between attendance intensity and the slope of achievement growth over time, after controlling for day school attendance and other background characteristics, would provide evidence of the LA's BEST intensity of program attendance impact.

Defining the Study Sample

The basis for the sample is comprised of the LAUSD 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 accomplished this task by going back through the historical records and tracking 4 years of background and California Standards Tests (CSTs) achievement data for the students in the two cohorts. The second- and third-grade cohorts were selected because 3 years of complete LA's BEST attendance data and 4 years of complete background data (achievement scores, day school attendance, etc.) was available for students in these cohorts. The following describes how we defined the two cohorts.

Grade 3 cohort (2003–04). Four years of achievement results were available for this cohort spanning the 2002–03 school year through the 2005–06 school year. Only students with valid CST achievement scores and LA's BEST attendance days reported during the study period were analyzed. Subsequently, students who were in third grade in 2003–04 were followed from 2002–03 to their projected fifth-grade year in 2005–06.¹² Because we employed HLM analysis to control for school-level effects, a minimum of 10 LA's BEST students per school was required for admission into the study sample. The resulting samples included 4,031 students in the math sample and 4,060 students in the English-language arts sample from 112 schools.

Grade 2 cohort (2003–04). Four years of achievement results and LA's BEST attendance data were available for this cohort spanning the 2003–04 school year through the

¹² Students may have been retained following the 2003-04 school year.

2006–07 school year. As with the first cohort, we only included students with valid CST achievement scores and LA’s BEST attendance during the study period. We followed students who were in second grade in 2003–04 through 2006–07, their projected year in fifth grade barring retention. Furthermore, because we employed HLM analysis to control for school-level effects, a minimum of 10 LA’s BEST students per school was required for admission into the sample. The resulting sample included 5,995 students in the math sample and 5,991 students in the English-language arts sample from 134 schools.¹³

Defining Attendance Intensity

Examination of student attendance patterns indicates that students participate in LA’s BEST with varying regularity. Therefore, it is necessary to set criterion to measure the intensity of attendance. In order to accomplish this, we computed the average attendance of all students in LA’s BEST over the 3 study years and then categorized attendance into four levels of intensity. In order to expand on the work of Frankel & Daley (2007), we defined the four intensity levels with the same cut points as used in their study. For the Grade 3 cohort, attendance intensity was based on the period from 2003–04 to 2005–06. For the Grade 2 cohort, attendance intensity was based on the period from 2004–05 to 2006–07. As with Frankel & Daley we did not expect students who average less than 20 days of attendance to benefit from the program; therefore, we classified them as Level 1. We classified students attending 21–50 days on average as Level 2, and those attending 51–100 days as Level 3. We defined regular attendance (Level 4) as those students who averaged greater than 100 days of LA’s BEST attendance per year.

Controlling for Existing Population Differences

Because we did not randomly assign students to the four intensity levels, it was necessary to control for existing differences in student background characteristics so that causal interpretations could be explored. In social science, randomized controlled experiments are often difficult to achieve due to study design and or ethical issues; subsequently quasi-experimental designs using propensity scoring methods are gaining widespread use. Typically, these designs employ logistic regression to estimate the probability that a subject is in a treatment group compared to a control group, and then use the propensity outcome to create balance among the student background characteristics. This process can be done using matching, stratum, or weighting techniques. In this study, we created an ordered treatment variable with four levels rather than a simple dichotomous

¹³ The Grade 2 cohort includes larger samples of students and schools due to the inclusion of students from new school sites added to LA’s BEST in the 2006-07 year.

treatment compared to a control. In the literature, adaptations to the basic propensity scoring method have been proposed in the case of an ordinal or dosage based treatment variable. Therefore, we adopt this approach by using ordinal logistic regression within a hierarchical linear modeling (HLM) framework to create a single propensity scalar.

Step 1 – HLM ordinal logistic regression. We employed ordinal logistic regression within an HLM framework to model the relationship between student background characteristics and the likelihood of a student attending the LA’s BEST program at the varying intensity levels. Level 1 (student level) indicators for baseline achievement, day school attendance, parental education, ethnicity (% Hispanic and % Black), Gender (% female), Limited English Proficient (LEP) and Initially Fluent English Proficient (IFEP) status were entered with the four level ordinal intensity variable used as the outcome. Each school represents a Level 2 unit and the average school achievement effect on the student-level intercept was included in the model. We then transformed the model coefficients to create a single propensity scalar, after which we divided the propensity scalar into quintiles. In other words, we gave each student a score of 1–5 based on his or her propensity score. The creation of this propensity quintile is necessary to apply a weighting method intended to remove initial differences in student background characteristics.

Step 2 – Weighting. The purpose behind the creation of the propensity scalar is to control for differences in background characteristics across the attendance intensity categories. To achieve this goal, we inversely weighted cases relative to their propensity outcome so that within each intensity level an equal number of weighted cases resulted in each propensity quintile. We also normalized the weighted cases so that the final weighted sample was the same size as the original un-weighted sample. Once balance existed among student background characteristics across the intensity levels, we could make valid comparisons. When balance was lacking for a specific variable, we added extra terms (i.e., variable squared or interaction terms) to the HLM ordinal logistic regression described in Step 1. We repeated this process until we achieved balance or balance was not achievable. The desired result was a sample with no more differences in background than would be expected from a randomly controlled design. If a significant relationship between a given background variable and attendance intensity was still present after this process, we included that variable as a covariate in the final growth model.

Chapter IV:
Student Cohort Demographic Analysis and HLM Modeling Results

In order to provide more clarity to our analyses, the demographic analysis and modeling results will be presented separately by cohort. The synthesis of the results of the two cohorts will be presented in the Discussion & Conclusion section.

Grade 3 Cohort – Student Population Characteristics

For the Grade 3 cohort, we conducted student achievement and demographic analyses by subject contents: Math and English-language arts.

Student achievement. Tables 1 and 2 present standardized CST achievement means for the Grade 3 cohort in math and English-language arts from the 2002–03 school year through 2005–06 for each intensity category. Student achievement in both math and English-language arts was higher for the students with over 100 days of LA’s BEST attendance in each of the 4 years when compared to students who attended LA’s BEST less often. For example, in 2002–03 the standardized CST achievement mean in math was 0.146 for students who averaged over 100 days of LA’s BEST attendance compared to a standardized CST achievement mean of -0.107 for students who averaged 20 days or less of LA’s BEST attendance. We tested these differences by attendance intensity with four separate one-way ANOVA’s (one for each year). The results were statistically significant for each year in both math and English-language arts ($p < .05$). These findings indicate that there were differences in math and English-language arts CST performance for students with varying levels of LA’s BEST attendance. In addition these differences exist for each year included in this study.

Table 1
Math Achievement by LA’s BEST Attendance Intensity, Grade 3 Cohort

Unweighted standardized math outcome	Average LA’s BEST attendance intensity (2003–04 to 2005–06)				ANOVA results	
	1–20 days (<i>n</i> = 1,131)	21–50 days (<i>n</i> = 784)	51–100 days (<i>n</i> = 744)	Over 100 days (<i>n</i> = 1,372)	<i>F</i> test	Sig.
CST math, 2002–03	-0.107	-0.078	-0.014	0.146	15.752	0.000
CST math, 2003–04	-0.085	-0.103	0.030	0.115	11.727	0.000
CST math, 2004–05	-0.120	-0.115	-0.010	0.175	23.360	0.000
CST math, 2005–06	-0.119	-0.078	-0.003	0.152	17.558	0.000

Table 2

English-Language Arts Achievement by LA's BEST Attendance Intensity, Grade 3 Cohort

Unweighted standardized language arts outcome	Average LA's BEST attendance intensity (2003–04 to 2005–06)				ANOVA results	
	1–20 days (<i>n</i> = 1,144)	21–50 days (<i>n</i> = 785)	51–100 days (<i>n</i> = 749)	Over 100 days (<i>n</i> = 1,382)	<i>F</i> test	Sig.
CST ELA, 2002–03	-0.099	-0.089	0.063	0.158	17.144	0.000
CST ELA, 2003–04	-0.408	-0.415	-0.307	-0.176	14.442	0.000
CST ELA, 2004–05	0.045	0.049	0.161	0.294	19.309	0.000
CST ELA, 2005–06	0.029	0.054	0.149	0.275	16.676	0.000

Student demographics. Tables 3 and 4 present the student background characteristics for the Grade 3 cohort for each intensity category. Not surprisingly, student attendance in day school was associated with the intensity of attendance in LA's BEST. Those students with higher attendance intensity in LA's BEST also attended day school more often. Students who attended LA's BEST more frequently were also more likely to be Black, female, classified as IFEP, and have parents with more than a high school education. Students who attended LA's BEST more frequently were also less likely to be Hispanic, or classified as LEP. All of the background characteristics presented in the math sample had statistically significant differences across the four attendance intensity categories ($p < .05$). In addition, all but one of the background characteristics presented in the English-language arts sample had statistically significant differences across the four attendance intensity categories ($p < .05$). This indicates that the students have different characteristics across the attendance intensity levels. In order to attribute differences in achievement outcomes solely to the level of intensity of participation it is necessary to control for these background differences. A propensity scoring method was used to reduce the differences among the groups in an attempt to create a final sample that would have no significant differences in these characteristics.

Table 3

Background Variables by LA's BEST Attendance Intensity, Math Sample of the Grade 3 Cohort

Unweighted standardized math outcome	Average LA's BEST attendance intensity (2003–04 to 2005–06)				ANOVA results	
	1–20 days (<i>n</i> = 1,131)	21–50 days (<i>n</i> = 784)	51–100 days (<i>n</i> = 744)	Over 100 days (<i>n</i> = 1,372)	<i>F</i> test	Sig.
Day school attendance (2004–05 to 2005–06)	153.850	155.633	156.156	158.325	11.727	0.000
Female	0.478	0.483	0.522	0.582	23.360	0.000
Black	0.050	0.046	0.074	0.073	17.558	0.000
Hispanic	0.901	0.927	0.872	0.845	15.752	0.000
IFEP	0.077	0.079	0.095	0.109	8.646	0.000
LEP	0.812	0.786	0.719	0.686	11.198	0.000
Parent < HS education	0.435	0.418	0.363	0.328	3.600	0.013
Parent is HS grad/No college	0.207	0.207	0.220	0.235	12.637	0.000
Parent had some college	0.112	0.125	0.134	0.169	3.238	0.021

Table 4

Background Variables by LA's BEST Attendance Intensity, English-Language Arts Sample of the Grade 3 Cohort

Unweighted standardized math outcome	Average LA's BEST attendance intensity (2003–04 to 2005–06)				ANOVA results	
	1–20 days (<i>n</i> = 1,131)	21–50 days (<i>n</i> = 784)	51–100 days (<i>n</i> = 744)	Over 100 days (<i>n</i> = 1,372)	<i>F</i> test	Sig.
Day school attendance (2004–05 to 2005–06)	153.740	155.625	156.083	158.138	8.336	0.000
Female	0.478	0.484	0.521	0.582	11.286	0.000
Black	0.050	0.046	0.076	0.072	3.767	0.010
Hispanic	0.902	0.927	0.870	0.847	12.653	0.000
IFEP	0.078	0.079	0.095	0.109	3.126	0.025
LEP	0.809	0.786	0.717	0.687	20.255	0.000
Parent < HS education	0.433	0.419	0.363	0.326	12.202	0.000
Parent is HS grad/No college	0.205	0.206	0.219	0.236	1.477	0.219
Parent had some college	0.112	0.126	0.136	0.171	6.625	0.000

Controlling for existing population differences. Because students were not randomly assigned to the four intensity levels, they displayed different characteristics across the four levels of intensity in attendance. Therefore, it was necessary to control for existing student background characteristics so that we could explore causal interpretations. In Tables 5 and 6, we show the relationship between each background variable and LA’s BEST attendance intensity for the sample after we have made adjustments (weighting based on a propensity scalar).

Table 5

Background Variables by LA’s BEST Attendance Intensity (After Weighting), Math Sample of the Grade 3 Cohort

Background variables	Average LA’s BEST attendance intensity (2003–04 to 2005–06)				ANOVA results	
	1–20 days (<i>n</i> = 1,131)	21–50 days (<i>n</i> = 784)	51–100 days (<i>n</i> = 744)	Over 100 days (<i>n</i> = 1,372)	<i>F</i> test	Sig.
Zscore: CST math, 2002–03	-0.011	0.000	-0.051	-0.001	0.480	0.696
Day school attendance (2004–05 to 2005–06)	154.637	156.126	155.411	156.018	0.968	0.407
Female	0.523	0.515	0.508	0.526	0.244	0.866
Black	0.068	0.058	0.068	0.053	1.107	0.345
Hispanic	0.872	0.910	0.884	0.883	2.329	0.073
IFEP	0.093	0.091	0.095	0.089	0.082	0.970
LEP	0.758	0.745	0.733	0.759	0.720	0.540
Parent < HS education	0.388	0.383	0.375	0.382	0.112	0.953
Parent is HS grad/No college	0.218	0.211	0.218	0.224	0.173	0.915
Parent had some college	0.134	0.148	0.128	0.139	0.496	0.685

The results in Table 5 demonstrate that, through use of the weighting process, we were able to remove nearly all of the bias associated with the relationship between the background variables and attendance intensity for the math sample. For example, in the weighted sample the percentage of female students across the four attendance categories ranges from a low of about 51% (51 to 100 days) to a high of about 53% (over 100 days). The significance test for gender has a *p*-value equal to 0.866, which indicates that these differences are not statistically significant. Before the weighting process was applied the percentage of female students across the four attendance categories ranged from a low of about 48% (1 to 20 days) to a high

of about 58% (over 100 days) and these differences were statistically significant. Other examples show that there was only about a .05 standard deviation difference between the high and low CST math mean and a difference of about a 1.5 attendance days for the day school attendance range. Similar results are seen in the weighted sample for nearly all the background characteristics. Generally these results allow us to conclude that there is balance for the background variables across attendance intensity categories in the weighted sample. Thus, it would be reasonable to expect the results after weighting to be comparable with those from a randomly controlled design. The relationship between being Hispanic and attendance intensity did approach statistical significance ($p > .05$). For this reason, Hispanic status was included as an additional variable in the final growth model for math.

Table 6
Background Variables by LA’s BEST Attendance Intensity (After Weighting), English-Language Arts Sample of the Grade 3 Cohort

Background variables	Average LA’s BEST attendance intensity (2003–04 to 2005–06)				ANOVA results	
	1–20 days (<i>n</i> = 1,144)	21–50 days (<i>n</i> = 785)	51–100 days (<i>n</i> = 739)	Over 100 days (<i>n</i> = 1,382)	<i>F</i> test	Sig.
Zscore: CST ELA, 2002–03	-0.034	-0.023	-0.012	-0.025	0.072	0.975
Day school attendance (2004–05 to 2005–06)	154.396	156.047	155.211	155.780	0.072	0.361
Female	0.517	0.517	0.500	0.524	1.069	0.072
Black	0.067	0.056	0.068	0.054	0.955	0.413
Hispanic	0.873	0.913	0.884	0.882	2.637	0.048
IFEP	0.090	0.088	0.090	0.091	0.018	0.997
LEP	0.761	0.753	0.745	0.755	0.213	0.887
Parent < HS education	0.390	0.385	0.384	0.377	0.141	0.935
Parent is HS grad/No college	0.219	0.211	0.213	0.224	0.210	0.889
Parent had some college	0.130	0.147	0.126	0.143	0.783	0.503

As with the math sample, the results in Table 6 show that through use of the weighting process, we were able to remove most of the bias associated with the relationship between the background variables and attendance intensity for the English-language arts sample. The relationship between being Hispanic and attendance intensity was, however, statistically significant ($p < .05$). This indicates that after weighting there were still some differences

across the attendance intensity categories in the proportion of Hispanic students. In addition, the relationship between being female and attendance intensity approached statistical significance ($p > .05$). For this reason, we included both Hispanic status and gender as additional controlling variables in the final growth model for English-language arts with this cohort.

Three-Level HLM Growth Model Results for Grade 3 Cohort

We employed a three-level hierarchical growth model to examine the impact of afterschool attendance intensity on student achievement. Two separate models were conducted for this cohort, one for math and one for English-language arts.

Math achievement. Table 7 includes the results from the three-level HLM growth model for math. We ran this model on the weighted sample that we had already adjusted to create balance among the background characteristics. The table presents model effects on both the baseline achievement level (intercept) and achievement growth (slope). The P -value indicates the statistical significance level of each effect, whereas the unstandardized B coefficient indicates the magnitude and direction of the effects. We tested the effect of attendance intensity in both LA's BEST and day school against math achievement at baseline (2002–03) and math growth over the course of the study (2002–06). The B coefficient indicates that for every year a student maintains regular LA's BEST attendance (over 100 days) their math achievement will increase by 0.034 standard deviations relative to a student with negligible LA's BEST attendance (0–20 days). This positive achievement growth was statistically significant ($p < .05$). Interestingly, day school attendance is associated with baseline math achievement ($p < .05$) but not with achievement growth ($p > .05$). The school-level math mean effects on the slope show that students in schools with higher mean math achievement at baseline experienced less growth than students in schools with lower baseline performance. To be more precise a student from a school which had a baseline math performance one standard deviation greater than the mean would experience 0.107 standard deviations less growth per year than a student from a school which had average baseline math performance.

Table 7

Three-Level HLM Growth Model for Math, Grade 3 Cohort

	<i>B</i> coefficient	<i>P</i> -value
Effects on the intercept (Math mean at year 2002–03)		
LA's BEST attendance over 100 days	0.010	0.770
LA's BEST attendance 51–100 days	-0.010	0.816
LA's BEST attendance 21–50 days	-0.013	0.785
Hispanic	-0.047	0.493
Day school attendance	0.004	0.001
School-level math mean, 2002–03	0.873	0.000
Effects on the slope (Math growth from 2002–06)		
LA's BEST attendance over 100 days	0.034	0.001
LA's BEST attendance 51–100 days	0.014	0.310
LA's BEST attendance 21–50 days	0.004	0.755
Hispanic	-0.008	0.613
Day school attendance	0.000	0.823
School-level math mean, 2002–03	-0.107	0.013

Figure 1 displays the estimated achievement growth trajectory from baseline for three LA's BEST attendance intensity categories in relation to those students who attended LA's BEST on average 0–20 days. The trajectory for students who attended LA's BEST on average 0–20 days is set at zero to serve as a reference line. Relative to students who attended LA's BEST for 20 days or less, students who attended LA's BEST an average of over 100 days saw their predicted math *Z*-scores grow by just over 0.1 standard deviations over the 3 years from baseline. Although this effect size is not large, the growth rate was significant ($p < .05$) and is an important finding given that this effect occurred after we carefully controlled background characteristics including day school attendance. Relative to students who attended 0–20 days, those who attended 51–100 days appear to experience a small degree of positive growth, although the difference was not large enough to reach statistical significance ($p > .05$).

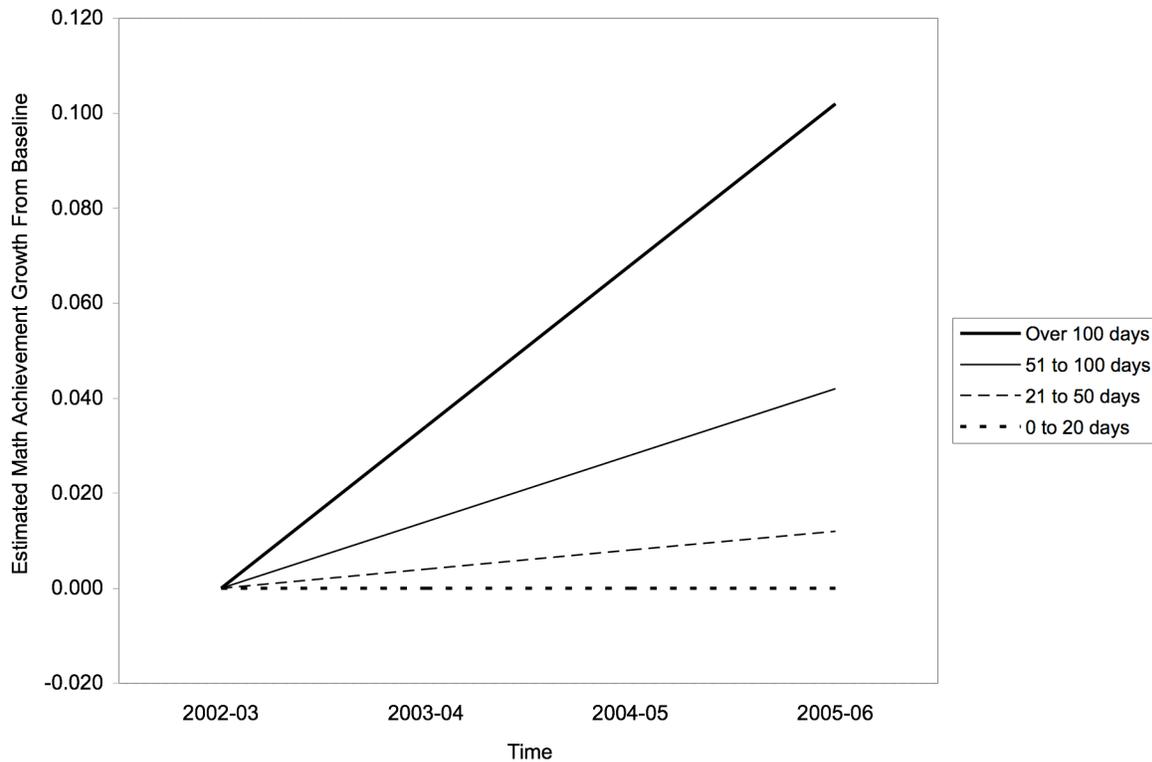


Figure 1. Model estimates, Grade 3 cohort – math achievement over time by LA’s BEST attendance intensity.

English-language arts achievement. Table 8 presents the results from the three-level HLM growth model for English-language arts. As with the math model, we ran this model on the weighted sample that we had already adjusted to create balance among the background characteristics. Again, we tested the effect of attendance intensity in both LA’s BEST and day school against English-language arts achievement at baseline (2003–04) and English-language arts growth over the course of the study (2003–07). Once again, day school attendance is associated with baseline English-language arts achievement ($p < .05$) but not with achievement growth ($p > .05$). Unlike the math sample results however, LA’s BEST attendance intensity was not significantly associated with positive English-language arts achievement growth ($p > .05$).

Table 8. Three-Level HLM Growth Model for English-Language Arts, Grade 3 Cohort

	<i>B</i> coefficient	<i>P</i> -value
Effects on the intercept (ELA mean at year 2002–03)		
LA’s BEST attendance over 100 days	-0.024	0.597
LA’s BEST attendance 51–100 days	0.017	0.688
LA’s BEST attendance 21–50 days	-0.001	0.978
Female	0.101	0.003
Hispanic	-0.141	0.015
Day school attendance	0.004	0.008
School-level ELA mean, 2002–03	1.077	0.000
Effects on the Slope (ELA growth from 2002–06)		
LA’s BEST attendance over 100 days	0.020	0.104
LA’s BEST attendance 51–100 days	-0.001	0.942
LA’s BEST attendance 21–50 days	0.005	0.696
Female	0.022	0.003
Hispanic	0.002	0.874
Day school attendance	0.000	0.694
School-level ELA mean, 2002–03	-0.144	0.000

Figure 2 displays the expected English-language arts achievement growth over time in each LA’s BEST attendance category relative to those students attending less than 20 days. The positive growth trend for students who attended LA’s BEST on average over 100 days was not large enough to reach statistical significance ($p > .05$). The other three lines in Figure 2 are relatively tightly bunched together confirming the finding that there were no significant differences between LA’s BEST attendance intensity and English-language arts achievement.

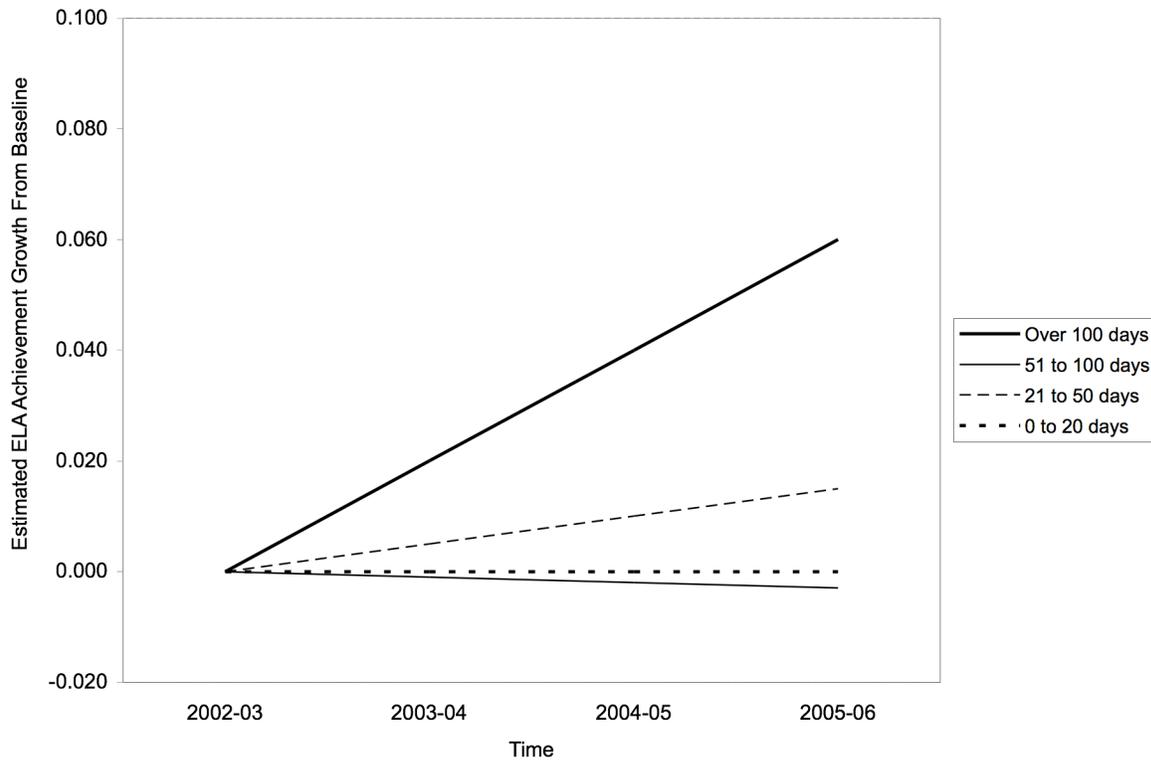


Figure 2. Model estimates, Grade 3 cohort – English-language arts achievement over time by LA’s BEST attendance intensity.

Next, we present the findings on the Grade 2 cohort. First, we provide the achievement analyses, followed with the demographic analyses and the HLM Modeling results.

Grade 2 Cohort - Student Population Characteristics

As with the Grade 3 cohort, we conducted student achievement and demographic analyses by subject contents: Math and English-language arts.

Student achievement. Tables 9 and 10 present standardized CST achievement means for the Grade 2 cohort in math and English-language arts from the 2003–04 school year through the 2006–07 school year for each intensity category. CST achievement scores are not equated across time; therefore, the comparisons of interest in these tables are the differences within each year across intensity levels and not CST achievement score changes over time. Student achievement in both math and English-language arts was highest for the students who attended LA’s BEST over 100 days during each of the 4 years. We tested these differences by attendance intensity with four separate one-way ANOVA’s for each assessment and found them to be significant in each year ($p < .05$).

Table 9. Math Achievement by LA's BEST Attendance Intensity, Grade 2 Cohort

Unweighted standardized math outcome	Average LA's BEST attendance intensity (2004–05 to 2006–07)				ANOVA results	
	1–20 days (<i>n</i> = 1,580)	21–50 days (<i>n</i> = 1,137)	51–100 days (<i>n</i> = 1,213)	Over 100 days (<i>n</i> = 2,065)	<i>F</i> test	Sig.
CST math, 2003–04	-0.090	-0.045	0.001	0.043	5.577	0.001
CST math, 2004–05	-0.056	0.020	0.072	0.117	10.253	0.000
CST math, 2005–06	-0.096	-0.062	0.062	0.115	18.554	0.000
CST math, 2006–07	-0.143	-0.122	-0.005	0.045	11.906	0.000

Table 10. English-Language Arts Achievement by LA's BEST Attendance Intensity, Grade 2 Cohort

Unweighted standardized language arts outcome	Average LA's BEST attendance intensity (2004–05 to 2006–07)				ANOVA results	
	1–20 days (<i>n</i> = 1,576)	21–50 days (<i>n</i> = 1,143)	51–100 days (<i>n</i> = 1,201)	Over 100 days (<i>n</i> = 2,071)	<i>F</i> test	Sig.
CST ELA, 2003–04	-0.179	-0.107	-0.069	0.030	13.276	0.000
CST ELA, 2004–05	-0.425	-0.334	-0.285	-0.202	16.398	0.000
CST ELA, 2005–06	0.099	0.180	0.263	0.325	16.168	0.000
CST ELA, 2006–07	0.056	0.122	0.170	0.223	10.662	0.000

Student Demographics. Tables 11 and 12 present the student background characteristics for the Grade 2 cohort for each intensity category. Similar to the Grade 3 cohort, student attendance in day school was associated with the intensity of attendance in the LA's BEST afterschool program. Those students with higher attendance intensity in LA's BEST also attended day school more often. Students who attended LA's BEST more frequently were also more likely to be female and have parents with more than a high school education. In addition, students who attended LA's BEST more frequently were less likely to be Hispanic, or classified as LEP. Most of the background characteristics presented in Tables 11 and 12 had statistically significant differences across the four attendance intensity categories ($p < .05$). This indicates that there are substantial existing differences in student background characteristics across the four attendance intensity groups. Therefore, it was necessary to control for these differences in order to draw meaningful inferences regarding the effect of LA's BEST attendance intensity.

Table 11. Background Variables by LA's BEST Attendance Intensity, Math Sample of the Grade 2 Cohort

Background variables	Average LA's BEST attendance intensity (2004–05 to 2006–07)				ANOVA results	
	1–20 days (<i>n</i> = 1,580)	21–50 days (<i>n</i> = 1,137)	51–100 days (<i>n</i> = 1,213)	Over 100 days (<i>n</i> = 2,065)	<i>F</i> test	Sig.
Day school attendance (2004–05 to 2006–07)	157.223	157.253	158.004	161.194	19.522	0.000
Female	0.497	0.503	0.516	0.546	3.413	0.017
Black	0.060	0.062	0.080	0.079	2.514	0.057
Hispanic	0.893	0.881	0.872	0.845	6.728	0.000
IFEP	0.056	0.069	0.077	0.076	2.192	0.087
LEP	0.690	0.667	0.655	0.590	14.787	0.000
Parent < HS education	0.358	0.354	0.326	0.285	9.075	0.000
Parent is HS grad/No college	0.189	0.175	0.208	0.193	1.384	0.246
Parent had some college	0.106	0.117	0.124	0.169	11.973	0.000

Table 12. Background Variables by LA's BEST Attendance Intensity, English-Language Arts Sample of the Grade 2 Cohort

Background variables	Average LA's BEST attendance intensity (2004–05 to 2006–07)				ANOVA results	
	1–20 days (<i>n</i> = 1,576)	21–50 days (<i>n</i> = 1,143)	51–100 days (<i>n</i> = 1,201)	Over 100 days (<i>n</i> = 2,071)	<i>F</i> test	Sig.
Day school attendance (2004–05 to 2006–07)	156.875	156.776	157.778	160.515	15.241	0.000
Female	0.500	0.501	0.517	0.553	4.355	0.005
Black	0.062	0.061	0.085	0.080	3.108	0.025
Hispanic	0.891	0.883	0.868	0.844	6.756	0.000
IFEP	0.057	0.068	0.077	0.075	1.934	0.122
LEP	0.690	0.669	0.655	0.591	14.789	0.000
Parent < HS education	0.359	0.355	0.327	0.286	9.194	0.000
Parent is HS grad/No college	0.188	0.178	0.206	0.190	1.034	0.376
Parent had some college	0.108	0.115	0.124	0.169	11.857	0.000

Controlling for existing population differences. As with the Grade 3 cohort, we did not randomly assign students to the four intensity levels. Furthermore, there were substantial existing differences in student background characteristics between the four levels. Therefore, it was necessary to control for existing student background characteristics so that we could explore causal interpretations. In Tables 13 and 14, we show the relationship between each background variable and LA’s BEST attendance intensity for the sample after we have made adjustments (weighting based on a propensity scalar).

Table 13. Background Variables by LA’s BEST Attendance Intensity (After Weighting), Math Sample of the Grade 2 Cohort

Background variables	Average LA’s BEST attendance intensity (2004–05 to 2006–07)				ANOVA results	
	1–20 days (<i>n</i> = 1,580)	21–50 days (<i>n</i> = 1,137)	51–100 days (<i>n</i> = 1,213)	Over 100 days (<i>n</i> = 2,065)	<i>F</i> test	Sig.
Zscore: CST math, 2003–04	-0.057	-0.022	-0.004	-0.030	0.672	0.569
Day school attendance (2004–05 to 2006–07)	157.790	157.616	157.761	159.993	6.806	0.000
Female	0.518	0.518	0.516	0.511	0.077	0.973
Black	0.070	0.068	0.081	0.063	1.299	0.273
Hispanic	0.879	0.873	0.871	0.871	0.205	0.893
IFEP	0.060	0.072	0.078	0.068	1.241	0.293
LEP	0.654	0.643	0.648	0.646	0.138	0.937
Parent < HS education	0.336	0.340	0.324	0.319	0.695	0.555
Parent is HS grad/No college	0.190	0.176	0.207	0.195	1.296	0.274
Parent had some college	0.121	0.124	0.124	0.134	0.554	0.646

Table 14. Background Variables by LA’s BEST Attendance Intensity (After Weighting), English-Language Arts Sample of the Grade 2 Cohort

Background variables	Average attendance intensity (2004–05 to 2006–07)				ANOVA results	
	1–20 days (<i>n</i> = 1,427)	21–50 days (<i>n</i> = 881)	51–100 days (<i>n</i> = 1,034)	Over 100 days (<i>n</i> = 1,577)	<i>F</i> test	Sig.
Zscore: CST ELA, 2003–04	-0.113	-0.063	-0.067	-0.092	0.724	0.538
Day school attendance (2004–05 to 2006–07)	157.560	157.144	157.478	159.072	3.441	0.016
Female	0.517	0.513	0.516	0.523	0.125	0.945
Black	0.071	0.064	0.087	0.063	2.516	0.056
Hispanic	0.878	0.876	0.866	0.873	0.317	0.813
IFEP	0.063	0.072	0.079	0.068	0.932	0.424
LEP	0.653	0.647	0.649	0.646	0.072	0.975
Parent < HS education	0.332	0.337	0.329	0.326	0.156	0.926
Parent is HS grad/No college	0.188	0.180	0.203	0.192	0.719	0.541
Parent had some college	0.122	0.123	0.125	0.134	0.501	0.682

The results in Tables 13 and 14 demonstrate that, through use of the weighting process, we were able to remove most of the bias associated with the relationship between the background variables and attendance intensity. For example, in the weighted math sample the percentage of female students across the four attendance categories ranges from a low of about 51% for those who attended over 100 days to a high of about 52% in the other 3 categories. The significance test for gender has a *p*-value equal to 0.973, which indicates that these differences are not statistically significant. Before the weighting process was applied the percentage of female students across the four attendance categories ranged from a low of about 50% (1–20 days) to a high of about 55% (over 100 days) and these differences were statistically significant. Other examples show that there was only about a 0.03 standard deviation difference between the high and low CST math mean, and the percentage of LEP students ranged from 64% to 65% across the attendance categories. Similar results are seen in the weighted sample for most of the background characteristics. There were, however, still significant differences in day school attendance intensity across the four LA’s BEST attendance categories ($p < .05$). This indicates that after weighting there were still some differences across the attendance intensity categories in the average number of days attending day school. In addition, the relationship between being Black and attendance intensity

approached statistical significance in the English-language arts sample ($p > .05$). For this reason, we included both Black status and day school attendance as additional controlling variables in the final growth model for English-language arts with this cohort.

Three-Level HLM Growth Model Results for Grade 2 Cohort

As with the Grade 3 cohort, we employed a hierarchical growth model to examine the impact of afterschool attendance intensity on student achievement. Two separate models were conducted for this cohort, one for math and one for English-language arts. This model is performed on the weighted sample in which differences in background characteristics and the initial achievement outcome across intensity levels have been removed. After weighting the sample, there were still differences in the average number of days attending day school between the LA's BEST attendance categories. In order to control for these remaining differences, we included day school attendance as a covariate in the growth models for the Grade 2 cohort.

Math achievement. Table 15 presents the results from the three-level HLM growth model for math. As previously mentioned, we ran this model on the weighted sample that we had already adjusted to create balance among the background characteristics. We tested the effect of attendance intensity in both LA's BEST and day school against math achievement at baseline (2003–04) and math growth over the course of the study (2003–07).

Regular LA's BEST attendance (over 100 days) was again significantly associated with positive achievement growth relative to students with LA's BEST attendance intensity of over 0–20 days per year ($p < .05$). The interpretation of the *B* coefficient indicates that for every year a student maintains regular LA's BEST attendance (over 100 days) their math achievement will increase by 0.029 standard deviations relative to a student with negligible LA's BEST attendance (0–20 days). Once again, day school attendance is associated with baseline math achievement ($p < .05$) but not with achievement growth ($p > .05$).

School-level achievement means from Grade 2 cohort students is positively associated with baseline math achievement ($p < .05$) but negatively associated with achievement growth ($p < .05$). As would be expected, this indicates that students in schools with higher mean math achievement at baseline experienced less growth than those in schools with lower mean math achievement at baseline (see Fraenkel & Wallen, 1993).

Table 15. Three-Level HLM Growth Model for Math, Grade 2 Cohort

	<i>B</i> coefficient	<i>P</i> -value
Effects on the intercept (Math mean at year 2003–04)		
LA’s BEST attendance over 100 days	0.037	0.231
LA’s BEST attendance 51–100 days	0.065	0.048
LA’s BEST attendance 21–50 days	0.063	0.092
Day school attendance	0.006	0.000
School-level math mean, 2002–03	0.934	0.000
Effects on the slope (Math growth from 2003–07)		
LA’s BEST attendance over 100 days	0.029	0.001
LA’s BEST attendance 51–100 days	0.019	0.064
LA’s BEST attendance 21–50 days	-0.012	0.273
Day school attendance	0.000	0.538
School-level math mean, 2003–04	-0.087	0.015

Figure 3 displays the positive association between LA’s BEST attendance intensity and math achievement. Although CST achievement scores are not equated across time, the growth model allows for relative comparisons between the four intensity categories. For this reason we display the expected achievement growth over time in each LA’s BEST attendance category in relation to those students who attended LA’s BEST on average 0–20 days. Relative to students who attended LA’s BEST 20 days or less, students who attended LA’s BEST an average of over 100 days saw their predicted math Z-scores grow by about 0.09 standard deviations over the 3 years from baseline ($p < .05$). The relative growth for students who attended LA’s BEST on average 51–100 days was not significant ($p > .05$). The negative growth trend for students who attended LA’s BEST on average 21–50 days was also not large enough to reach statistical significance ($p > .05$).

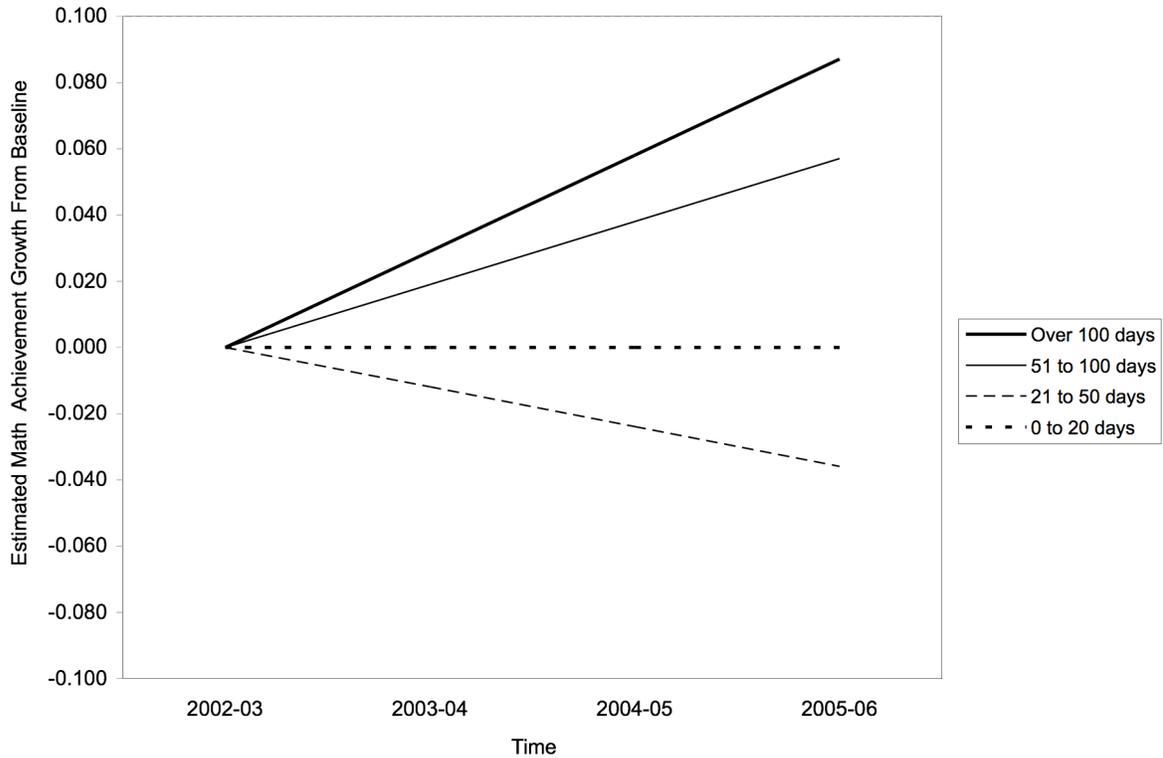


Figure 3. Model estimates, Grade 2 cohort – math achievement over time by LA’s BEST attendance intensity.

English-language arts achievement. Table 16 presents the results from the three-level HLM growth model for English-language arts. As with the other models, we ran this model on the weighted sample that we had already adjusted to create balance among the background characteristics. Again, we tested the effect of attendance intensity in both LA’s BEST and day school against English-language arts achievement at baseline (2003–04) and English-language arts growth over the course of the study (2003–07).

Similar to the Grade 3 cohort, LA’s BEST attendance intensity was not significantly associated with positive English-language arts achievement growth ($p > .05$). Furthermore, as with the Grade 3 cohort, day school attendance is associated with baseline English-language arts achievement ($p < .05$) but not with achievement growth ($p > .05$). In other words, students in schools with higher mean English-language arts achievement at baseline experienced less growth than students in schools with lower baseline performance.

Table 16. Three-level HLM Growth Model for English-Language Arts, Grade 2 Cohort

	<i>B</i> coefficient	<i>P</i> -value
Effects on the intercept (ELA mean at year 2003–04)		
LA’s BEST attendance over 100 days	0.027	0.424
LA’s BEST attendance 51–100 days	0.061	0.088
LA’s BEST attendance 21–50 days	0.068	0.119
Day school attendance	0.006	0.000
Black	0.061	0.191
School-level ELA mean, 2003–04	0.975	0.000
Effects on the slope (ELA growth from 2003–07)		
LA’s BEST attendance over 100 days	0.008	0.380
LA’s BEST attendance 51–100 days	0.009	0.315
LA’s BEST attendance 21–50 days	-0.006	0.496
Day school attendance	0.000	0.111
Black	-0.023	0.093
School-level ELA mean, 2003–04	-0.107	0.000

Figure 4 displays the expected English-language arts growth over time in each LA’s BEST attendance category also relative to those students who attended LA’s BEST an average of 0–20 days. This figure shows that the achievement trajectories for the four intensity levels are bunched tightly together. For example, the expected English-language arts growth over the study period for students with average LA’s BEST attendance of over 100 days was just 0.024 standard deviations above the expected growth for those who attended LA’s BEST an average of 0–20 days. Furthermore, the figure represents the finding that there were no significant differences in achievement growth due to LA’s BEST attendance intensity ($p > .05$).

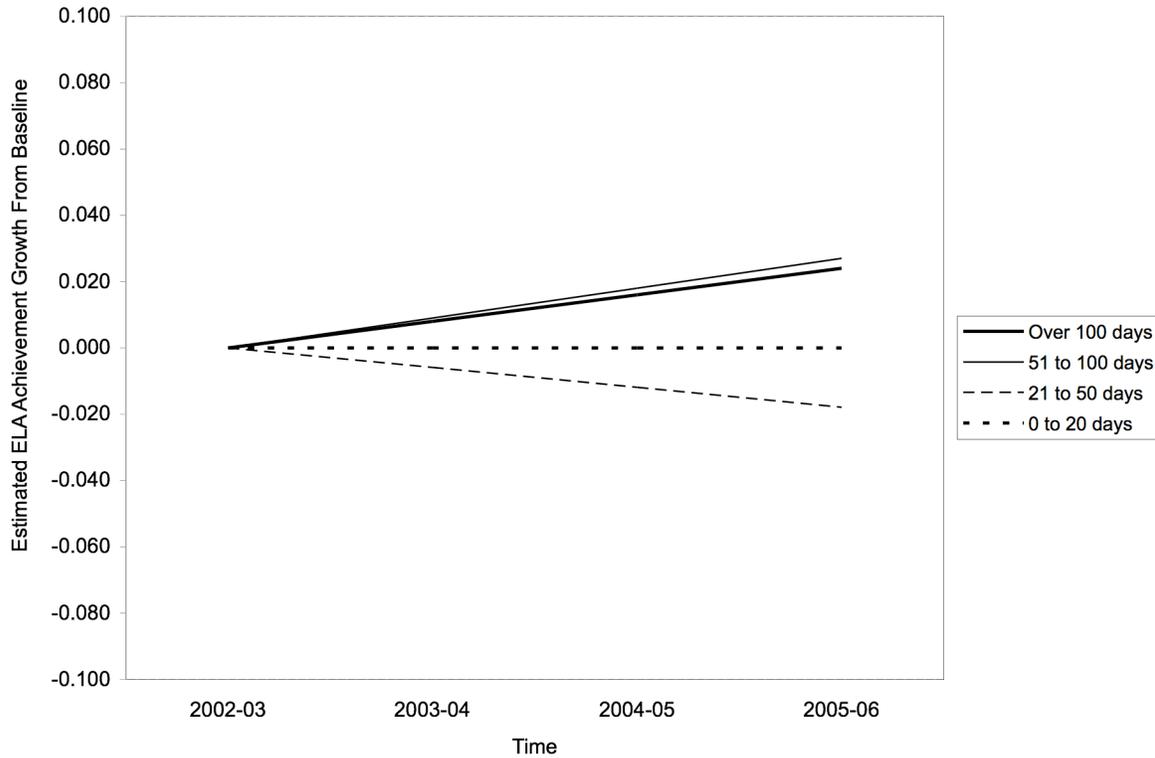


Figure 4. Model estimates, Grade 2 cohort – English-language arts achievement over time by LA’s BEST attendance intensity.

In Summary

Results of the analysis suggest that regular attendance in the LA’s BEST program (over 100 days per year) leads to positive math achievement growth when compared to students with low attendance in the program. This finding was consistent in two separate cohorts of students who we followed over a 4-year period and was statistically significant. The finding of positive impact for regular LA’s BEST attendance on math achievement growth was present after carefully accounting for existing differences in student background characteristics, in addition to important indicators such as students’ initial performance levels and their day school attendance over the study period. In contrast, we found that students’ achievement growth in English-language arts was not significantly related to the students’ intensity of attendance in the LA’s BEST program. This finding was also consistent for both cohorts of students represented in this study.

Chapter V: Discussion and Conclusion

This study set out to fill a research gap by using rigorous methodology to study the effects of “dosage” (intensity of afterschool attendance) on students’ academic outcomes. It extends the current literature on the impact of afterschool programs in two key ways: First, the analyses explicitly modeled achievement longitudinally for 4 years. Second, the study used a large sample of over 10,000 students and took extensive care to apply an advanced multilevel propensity matching technique to establish a valid study sample from which we could generate valid inferences.

Implications for Methodology

Outcome studies of afterschool programs typically are designed to compare participants with non-participants based on any program attendance. Consequently, participants may attend one day in an afterschool program and still be included in the treatment group. Furthermore, non-participants may have been enrolled in other afterschool activities and still be included in a control group. As stated in a report by Frankel and Daley (2007), two very important issues are ignored by most studies: First, “How did the non-participants spend their time afterschool?” and second, “How intensive was the participants’ program attendance?” (p. 12). Expanding on Frankel and Daley’s strategy and addressing their concerns, this study used statistical strategies to reduce selection bias and confirm their findings on the importance of “dosage” for afterschool participants. Similar to their study, this study grouped afterschool students by their intensity of attendance into four groups (i.e., 0–20, 21–50, 51–100, and over 100 days), and compared the three higher intensity levels against the low intensity group, thus addressing their second question concerning intensity of participation. Additionally, by comparing the low dosage students to high dosage students, this study reduces their fore-mentioned concern on how non-participants spend their out-of-school time.¹⁴

Although rare in afterschool studies, it is common in intervention and medical studies to examine the effect of different levels of treatment or dosage received and compare groups receiving low dosage to groups receiving higher dosage (Imbens, 2000; Leon, Mueller, Solomon & Keller, 2001). In this study, we considered it logical to compare the low attendance students to those with regular program attendance. The rationale is that because

¹⁴ LA’s BEST requires 5 days of attendance per week. Based on this rationale, this study considered low dosage students unlikely to be simultaneously enrolled in another afterschool program. Despite this, propensity scoring and covariance methods were used to remove most of the observable characteristic differences anticipated between the low dosage and other intensity groups.

these students have demonstrated the intent to receive treatment (through enrollment in LA's BEST), they can be considered to have very similar background characteristics to those students with regular LA's BEST attendance, which makes them a superior control group than those students who have never attended the program. Furthermore, the selection bias issues that apply to control students who have never attended the program¹⁵ are likely to be greater than for those who have demonstrated some level of need for the program through their enrollment. Thus, by confining our analyses to students who had some contact with the LA's BEST program we removed a potential source for self-selection bias.

Despite this, we realized that many self-selection differences would still exist among the students who participated in LA's BEST at the various intensity levels. Therefore, we used propensity scores to balance the samples and covariates to eliminate any pre-existing differences. By employing a study design that compares low attendance students to high attendance students, and uses propensity scores to weight the existing differences among the four intensity groups, we were able to address most of the selection bias issues.

Implication of Results

Results of the analysis provide evidence that regular attendance in the LA's BEST program (over 100 days per year) leads to positive math achievement growth when compared to students with low attendance in the program. This finding was consistent in two separate cohorts of students whom we followed over a 4-year period and was statistically significant. Furthermore, it is important to note that this result is obtained after carefully accounting for existing differences in students' background characteristics, so that the most plausible explanation of this statistical difference is in the intensity of attendance.

In contrast, although the trend of English-language arts achievement growth is positive and follows a developmental pattern similar to math, it is not significantly related to the students' intensity of attendance. This finding was also consistent for both cohorts of students represented in this study.

Multilevel longitudinal models are used to model student academic achievement over time. The multilevel modeling is statistically necessary to account for the nested structure of the data, but also provides a tool with which we can examine important between-school variation in program implementation. Results from this modeling imply that math achievement growth is higher for school sites that initially scored lower at the baseline

¹⁵ such as having a role model at home who attends to their needs, enrolling in other afterschool activities, being tutored by a tutoring agency and so forth

period, therefore attendance in LA's BEST may have the best potential to benefit the students enrolled in those schools.

It is also interesting to learn that day school attendance is associated with baseline math and English-language arts performance but not with achievement growth. In contrast, regular LA's BEST attendance (over 100 days) is significantly related to achievement growth in math. This finding suggests that regular attendance at LA's BEST can have a positive growth effect on student achievement beyond the effect of day school attendance.

Furthermore, implications from this study highlight that simple indicators of program participation are inadequate to capture program effects fully. For a program to have impact on students' achievement, the students need to receive sufficient exposure. Participation level would be a better indicator of program effects until the field can find methodologies that control the self-selection biases that are inherent and hidden in the non-participants. Supporting Frankel and Daley's finding (2007), this study also found that regular afterschool program attendance of at least 100 days per year is necessary to reap the program benefits.

As shown in Figures 1 through 4, students in the Level 1 and 2 intensity groups (0–20 days and 21–50 days, respectively) show flat or slightly negative growth trends, whereas students in the Level 3 and 4 intensity groups (51–100 and over 100 days, respectively) display positive achievement growth. The figures also illustrate that as afterschool attendance intensity increases, achievement growth increases as well, with the Level 2 group revealing a steeper slope than the Level 1 group. The exception is the English-language arts sample of the Grade 3 cohort, where the Level 1 and 2 intensity groups bunch close together. These results indicate that LA's BEST is capable of making a difference in math achievement growth, but students need to have regular attendance to reap the benefits of the program.

Concerning program implementation, this study found that Hispanics, English Language Learners, male students, and students from families with lower parent education levels are less likely to have regular attendance (over 100 days). Therefore, LA's BEST can increase the benefits of the program to these students by examining the needs of these students and families closely and by offering incentives and program activities that will entice their regular attendance.

Conclusion

This study sets out to fill a research gap by using rigorous methodology to study the effects of “dosage” (intensity of afterschool attendance) on students' academic outcomes. The research tracked approximately 10,000 students for 4 years. We found that students who attended LA's BEST for over 100 days per year showed statistically significant achievement

growth in math as compared to students who participated 20 days or less per year. This achievement growth is more evident in school sites that scored lower in math at the baseline level suggesting that students from schools that are lower performing gain most from the program. In other words, LA's BEST is serving their targeted population (low performing students) as intended. LA's BEST can improve their effectiveness by encouraging all students to participate at a minimum of 101 days per year.

References

- American Institutes for Research. (2002). *Evidence of effects on student achievement*. Unpublished manuscript. Washington DC: U. S. Department of Education.
- Bergin, D. A., Hudson, L. M., Chryst, C. F., & Resetar, M. (1992). An afterschool intervention program for educationally disadvantaged young children. *The Urban review*, 24(3), 203–217.
- Bloomquist, N. (1977). On the relation between change and initial value. *Journal of the American Statistical Association*, 72, 746–749.
- California AfterSchool Network (2007). *California's funding landscape*. Retrieved December 17, 2007, from http://www.afterschoolnetwork.org/as_landscape
- Cooper, H., Charlton, K., Valentine, J. C., & Muhlenbruck, L. (2000). Making the most of summer school: A meta-analytic and narrative review. *Monographs of the Society for Research in Child Development*, 65(1, Serial No. 260).
- Dynarski, M., Moore, M., Mullens, J., Gleason, P., James-Burdumy, S., Rosenberg, L., et al. (2003). *When schools stay open late: The national evaluation of the 21st Century Community Learning Centers Program, first-year findings*. Princeton, NJ: Mathematica Policy Research.
- 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).
- Feister, L. M., Simpkins, S. D., & Bouffard, S. M. (2005, Spring). Present and accounted for: Measuring attendance in out-of-school time programs. *New Directions for Youth Development*, 105, 91–107.
- Fraenkel, J. R., & Wallen, N. E. (1993). *How to design and evaluate research in education* (2nd ed.). New York: McGraw-Hill.
- Frankel, S., & Daley, G. (2007). *An evaluation of after school programs provided by Beyond the Bell's partner agencies*. Los Angeles: Beyond the Bell Branch, Los Angeles Unified School District.
- Goerge, R. M., Cusick, G. R., Wasserman, M., & Gladden, R. M. (2007). *After-school programs and academic impact: A study of Chicago's After School Matters*. Chicago: Chapin Hall Center for Children.
- Goldschmidt, P., Huang, D., & Chinen, M. (2007). *The long-term effects of after-school programming on educational adjustment and juvenile crime: A study of the LA's BEST after-school program*. Los Angeles: University of California, National Center for Research, on Evaluation, Standards, & Student Testing (CRESST).
- Hodgkinson, H. (2006). *The whole child in a fractured world*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Hollister, R. (2003). *The growth in after-school programs and their impact*. Washington, DC: Brookings Institution.

- Hox, J. J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Huang, D., Gribbons, B., Kim, K. S., Lee, C., & Baker, E. L. (2000). *A decade of results: The impact of the LA's BEST after school enrichment program on subsequent student achievement and performance*. Unpublished manuscript, University of California, Los Angeles: National Center for Research, on Evaluation, Standards, & Student Testing (CRESST).
- Huang, D., Miyoshi, J., La Torre, D., Marshall, A., Pérez, P., & Peterson, C. (2007). *Exploring the intellectual, social and organizational capitals at LA's BEST* (CRESST Tech. Rep. No. 714). Los Angeles: University of California, National Center for Research, on Evaluation, Standards, & Student Testing (CRESST).
- Imbens, G. W. (2000). The role of propensity score in estimating dose-response functions. *Biometrika*, 87(3), 706–710.
- James, D. W. (1997). Some things do make a difference for youth: A compendium of evaluations of youth programs and practices. Washington, DC: American Youth Policy Forum. (ERIC Document Reproduction Service No. ED 409 462).
- Jenner, E., & Jenner, L. W. (2007). Results from a first-year evaluation of academic impacts of an after-school program for at-risk students. *Journal of Education for Students Placed At Risk*, 12(2), 213–237.
- Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. A., Snow, D., and Martin-Glenn, M. (2003). *The effectiveness of out-of-school-time strategies in assisting low-achieving students in reading and mathematics*. Aurora, CO: Mid-continent Research for Education and Learning.
- Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., & Martin-Glenn, M. L. (2006). Out-of-School-Time Programs: A Meta-Analysis of Effects for At-Risk Students. *Review of educational research*, 76(2), 275–313.
- Lauver, S. C. (2002). Assessing the benefits of an after-school program for urban youth: An impact and process evaluation. (Doctoral dissertation, University of Pennsylvania, 2002). *Digital Abstracts International*, 63, 02.
- Leon, A. C., Mueller, T. I., Solomon, D. A., & Keller, M. B. (2001). A dynamic adaptation of the propensity score adjustment for effectiveness analyses of ordinal doses of treatment. *Statistics in Medicine*, 20, 1487–1498.
- Little, P. M. D., & Harris, E. (2003, July). A review of out-of-school time program quasi-experimental and experimental evaluation results. *Out-of-School Time Evaluation Snapshot*, 1.
- McComb, E. M., & Scott-Little, C. (2003). *After-school programs: Evaluations and outcomes*. Greensboro, NC: SERVE.
- Miller, B. M. (2003). *Critical hours: Afterschool programs and educational success*. Quincy, MA: Nellie Mae Education Foundation. (ERIC Document Reproduction Service No. ED 482 794).

- Muñoz, M. A. (2002). *Outcome-based community—schools partnerships: The impact of the after-school programs on non-academic and academic indicators*. (ERIC Document Reproduction Service No. ED 468 973).
- No Child Left Behind Act of 2001, Pub. L No. 107-110, § 115 Stat.1425 (2002). Retrieved December 17, 2007, from <http://www.ed.gov/legislation/ESEA02/>
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Redd, Z., Cochran, S., Hair, E., & Moore, K. (2002). *Academic achievement programs and youth development: A synthesis*. Washington, DC: Child Trends, Inc.
- Rogosa, D., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin*, 92(3), 726–748.
- Schaps, E. (2006). *Educating the whole child*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Scott-Little, C., Hamann, M. S., & Jurs, S. G. (2002). Evaluations of after-school programs: A meta-evaluation of methodologies and narrative summary findings. *American Journal of Evaluation*, 23(4), 387–419.
- Sheley, J. F. (1984). Evaluation of the centralized, structured, after-school tutorial. *Journal of Educational Research*, 77(4), 213–218.
- Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.
- Vanderhaar, J., & Muñoz, M. A. (2006). *Educating at-risk African American males: Formative and summative evaluation of the Street Academy Program*. (ERIC Document Reproduction Service No. ED 495 958).
- Welsh, M. E., Russell, C. A., Williams, I., Reisner, E. R., & White, R. N. (2002). *Promoting learning and school attendance through after-school programs: Student-level changes in educational performance across TASC's first three years*. Washington, DC: Policy Studies Associates, Inc.

TECHNICAL APPENDIX

Step 1 – HLM Ordinal Logistic Regression

We employed ordinal logistic regression within a HLM framework to model the relationship between student background characteristics and the likelihood of a student attending the LA's BEST program at the varying intensity levels. Level 1 (student level) indicators include continuous measures for baseline achievement, and day school attendance and dummy variables for parental education (less than high school, High school graduate no college, some college), ethnicity (Hispanic and Black), gender (female), LEP, and IFEP status. The four-level ordinal attendance intensity variable used as the outcome. Each school represents a Level 2 unit and the average school achievement effect on the student-level intercept is included in the model. An example of the model equation syntax is shown below:

Level-1 Model

$$\begin{aligned}\text{Prob}[R = 1|B] &= P'(1) = P(1) \\ \text{Prob}[R \leq 2|B] &= P'(2) = P(1) + P(2) \\ \text{Prob}[R \leq 3|B] &= P'(3) = P(1) + P(2) + P(3) \\ \text{Prob}[R \leq 4|B] &= 1.0\end{aligned}$$

where

$$\begin{aligned}P(1) &= \text{Prob}[Y(1) = 1|B] = \text{Probability of being in LA's BEST (0–20 days)} \\ P(2) &= \text{Prob}[Y(2) = 1|B] = \text{Probability of being in LA's BEST (21–50 days)} \\ P(3) &= \text{Prob}[Y(3) = 1|B] = \text{Probability of being in LA's BEST (51–100 days)}\end{aligned}$$

$$\log\left[\frac{P'(1)}{1 - P'(1)}\right] = B_0 + B_1*(FEMALE) + B_2*(BLACK) + B_3*(HISPANIC) + B_4*(R_ATTENDANCE) + B_5*(LTHS) + B_6*(HSGRAD) + B_7*(SOME_COLL) + B_8*(LEP) + B_9*(IFEP) + B_{10}*(STUDENT_CST) + B_{11}*(R_ATTENDANCE2)$$

$$\log\left[\frac{P'(2)}{1 - P'(2)}\right] = B_0 + B_1*(FEMALE) + B_2*(BLACK) + B_3*(HISPANIC) + B_4*(R_ATTENDANCE) + B_5*(LTHS) + B_6*(HSGRAD) + B_7*(SOME_COLL) + B_8*(LEP) + B_9*(IFEP) + B_{10}*(STUDENT_CST) + B_{11}*(R_ATTENDANCE2) + d(2)$$

$$\log\left[\frac{P'(3)}{1 - P'(3)}\right] = B_0 + B_1*(FEMALE) + B_2*(BLACK) + B_3*(HISPANIC) + B_4*(R_ATTENDANCE) + B_5*(LTHS) + B_6*(HSGRAD) + B_7*(SOME_COLL) + B_8*(LEP) + B_9*(IFEP) + B_{10}*(STUDENT_CST) + B_{11}*(R_ATTENDANCE2) + d(3)$$

Level-2 Model

(School CST achievement modeled against student intercept)

$$\begin{aligned}B_0 &= G_{00} + G_{01}*(SCHOOL_CST) + U_0 \\ B_1 &= G_{10}, B_2 = G_{20}, B_3 = G_{30}, B_4 = G_{40}, B_5 = G_{50}, B_6 = G_{60} \\ B_7 &= G_{70}, B_8 = G_{80}, B_9 = G_{90}, B_{10} = G_{100}, B_{11} = G_{110}\end{aligned}$$

Resulting model coefficients were then transformed so that a single propensity scalar was created, after which the propensity scalar was divided into quintiles.

Step 2 – Weighting

The purpose behind the creation of the propensity scalar was to control for differences in background characteristics across the attendance intensity categories. To achieve this goal, we inversely weighted the cases relative to their propensity outcome so that within each of the intensity levels an equal number of weighted cases resulted in each propensity quintile. We also normalized the weighted cases so that the final weighted sample was the same size as the original un-weighted sample.

The following SPSS code is used to accomplish this task:

```
** first compute aggregate propensity scalar mean by ‘propensity scalar quintile’ & and the LA’s Best Intensity variable *****.
```

```
AGGREGATE  
  /OUTFILE = *  
  MODE = ADDVARIABLES  
  /BREAK = att_intensity pr_quintile  
  /scalar_mean = MEAN(scalar).
```

```
*** compute temporary weight based on ratio **.  
compute wt1 = scalar_mean/scalar.
```

```
** Compute aggregate sum of cases in each intensity category **.
```

```
AGGREGATE  
  /OUTFILE = *  
  MODE = ADDVARIABLES  
  /BREAK = att_intensity  
  /attend_sum = n.
```

```
weight by wt1.
```

```
** Compute aggregate weighted sum of cases in each intensity by quintile category **.
```

```
AGGREGATE  
  /OUTFILE = *  
  MODE = ADDVARIABLES  
  /BREAK = att_intensity pr_quintile  
  /weight_sum = n.
```

```
** Compute final normalized weight ***.  
compute fweight = wt1*((attend_sum/5)/weight_sum).  
weight by fweight.
```

```
** Check crosstab to be sure that within each intensity category each propensity quintile is  
equally represented in the weighted sample. Also check that each intensity category weighted  
sample size is unchanged from the raw sample **.
```

```
CROSSTABS
```

```
/TABLES = att_intensity BY pr_quintile
```

```
/FORMAT = AVALUE TABLES
```

```
/CELLS = COUNT
```

```
/COUNT ROUND CELL .
```

Once balance exists among student background characteristics across intensity levels valid comparisons can be made. When balance was lacking for a specific variable, we added extra terms (variable squared or interaction terms) to the HLM ordinal logistic regression described in Step 1. We repeated this process until we achieved balance or balance was not possibly achievable. The desired result was a sample where there would be no more differences in background than would be expected from a randomly controlled design. If a significant relationship between a given background variable and attendance intensity was still present after this process we included that variable as a covariate in the final growth model.

Modeling Achievement Growth – Three-level HLM Growth Model

We employed a three-level hierarchical growth model to examine the impact of afterschool attendance intensity on student achievement. In this model, Level 1 represents time nested within students. For the Grade 3 cohort there are four time points (2002–03 to 2005–06), with achievement at each time point serving as the outcome. Similarly, for the Grade 2 cohort there are also four time points (2003–04 to 2006–07), with achievement at each time point serving as the outcome. The Level 1 intercept is initialized at the first time point (2002–03 for the Grade 3 cohort). Level 2 accounts for student-level effects. At this level, LA’s BEST attendance intensity is modeled against the Level 1 achievement intercept and the achievement slope over time. We also included day school attendance in the model as a student-level covariate. Like LA’s BEST attendance, we modeled day school attendance against the Level 1 achievement intercept and the achievement slope over time. Level 3 accounts for school-level variation. School-level baseline achievement in the assessment being examined is also modeled against the Level 1 achievement intercept and the achievement slope over time.

This model is performed on the weighted sample in which differences in background characteristics and the initial achievement outcome across intensity levels have been removed. Therefore, we did not expect any effect of LA's BEST attendance intensity on the achievement intercept. The primary relationship of interest is that between attendance intensity and the slope of achievement growth over time. The presence of a significant relationship between attendance intensity and the slope of achievement growth over time, after controlling for day school attendance and other background characteristics, would provide evidence of the LA's BEST program attendance impact.

An example of the model for the Grade 2 cohort is shown below:

Level-1 Model (CST achievement modeled across time)

$$Y = P0 + P1*(TIME) + E$$

Level-2 Model

(Student LA's BEST and day school attendance modeled against CST achievement intercept and slope)

$$P0 = B00 + B01*(R_ATTENDANCE) + B02*(HI_Intensity) + B03*(MED_Intensity) + B04*(LOW_Intensity) + R0$$

$$P1 = B10 + B11*(R_ATTENDANCE) + B12*(HI_Intensity) + B13*(MED_Intensity) + B14*(LOW_Intensity) + R1$$

Level-3 Model

(School CST achievement modeled against student intercept and slope)

$$B00 = G000 + G001(SCHOOL_CST) + U00$$

$$B01 = G010$$

$$B02 = G020$$

$$B03 = G030$$

$$B04 = G040$$

$$B10 = G100 + G101(SCHOOL_CST) + U10$$

$$B11 = G110$$

$$B12 = G120$$

$$B13 = G130$$

$$B14 = G140$$