CRESST REPORT 757

Denise Huang
Seth Leon
Aletha M. Harven
Deborah La Torre
Sima Mostafavi

EXPLORING THE RELATIONSHIP BETWEEN LA'S BEST PROGRAM ATTENDANCE AND COGNITIVE GAINS OF LA'S BEST STUDENTS

APRIL, 2009

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

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

Exploring the Relationships between LA's BEST Program Attendance and Cognitive Gains of LA's BEST Students

CRESST Report 757

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

April, 2009

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

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).

EXPLORING THE RELATIONSHIPS BETWEEN LA'S BEST PROGRAM ATTENDANCE AND COGNITIVE GAINS OF LA'S BEST STUDENTS

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

Abstract

The purpose of this report is to examine the extent to which participation in the Los Angeles Better Educated Students for Tomorrow (LA's BEST) afterschool program leads to positive achievement outcomes in math and English-language arts. A quasi-experimental design is utilized, and hierarchical linear modeling (HLM) is employed to examine the relations between intensity of program participation and achievement outcomes across 4 years of data using two cohorts of students. Results reveal that regular attendance (over 100 days per year) in the LA's BEST afterschool program leads to higher achievement in California Standards Test (CST) math performance but not in CST English-language arts performance. Therefore, LA's BEST can improve their program outcomes by setting program structures, activities, and policies to encourage all students to attend regularly.

Introduction

This study serves as an addendum to the 2008 report *Examining the Relationship between LA's BEST Program Attendance and Academic Achievement of LA's BEST Students* (Huang, Leon, La Torre & Mostafavi, 2008: CRESST Tech. Rep. No. 749). Results of the study 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 (1–20 days per year). This finding was consistent in two separate cohorts of students whom we followed over a 4-year period and was statistically significant. The result is obtained after carefully accounting for existing differences in students' background characteristics with propensity scoring methods so that the most plausible explanation of this statistic difference is in the regularity of attendance. Because the 2008 study compared LA's BEST students with different intensity levels of participation, this report intends to provide additional support to the study by adding a group of students who participated in LA's BEST during the baseline years, but did not participate in the program during the study period. Our hypothesis is that the regular attendance group (over 100 days per year) will duplicate the results of the 2008 study and the no attendance group will mirror the results of the low attendance group, thus further validating the use of low and no attendance students in LA's BEST as comparisons.

The research questions for this study are as follows:

- Does attendance in the LA's BEST afterschool program lead to positive achievement outcomes in math and English-language arts?
- Do the achievement outcomes of LA's BEST students' vary as a function of their different intensity levels of LA's BEST participation?

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 of this study sample is comprised of the Los Angeles Unified School District (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 accomplish 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, sample, and 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 student sample is comprised of roughly 6,000 students from LA's BEST programs. The sample includes two cohorts of students with base years in 2002–03 and 2003–04, respectively. We separated students in each cohort into four groups based on their intensity of attendance in the LA's BEST program. Once this was completed, we took advantage of this panel structure and applied hierarchical linear modeling (HLM) to academic outcomes. This method allowed us to examine students' academic improvement based on their different levels of intensity, while controlling for student and school-level background characteristics.

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 accomplish 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. In order to apply appropriate statistical techniques, students were required to attend the same school for all four years that their cohort was in the study for admission to the sample. Given that we are examining the relations between program attendance and achievement outcomes across multiple time points, the study periods were defined as:

- Baseline years (2002–03 and 2003–04, depending on cohort)
- The follow-up period (2003–06 and 2004–07)

Students were then categorized based on their program attendance dates:

- Comparison group – Those who attended LA's BEST during the baseline, but not during the follow-up period (2003–06 and 2004–07).
- Program participant groups – Those students who did not attend LA's BEST at baseline, but attended during the follow-up period (with varying levels of intensity).

It should be noted that we intentionally eliminated students who never attended the LA's BEST program in an effort to reduce the potential effect of self-selection bias. For example, when families and students elect not to participate in afterschool programs, they do so for a variety of reasons. These reasons may include, but are not limited to: a lack of interest in the program's activities; students may have an adult or mentor at home to supervise and provide help with homework; and students may participate in other afterschool activities. These variations make interpretations for study findings somewhat complex. In other words, these students may differ from the students who attended the program in meaningful ways, but these differences may or may not be captured and controlled by existing background data (e.g., parent educational level, race/ethnicity, etc.).

As mentioned in the 2008 report (Huang et al.), it is necessary for any intervention project to expose students to sufficient treatment in order to demonstrate effects. However, afterschool studies rarely examine dosage.[1] In this study, we considered it logical to further the research on dosage by comparing students who only participated in the baseline years to those students who only participated in the follow-up periods. The rationale is that students who received treatment at baseline (through enrollment in LA's BEST) can be considered more likely to have similar background characteristics to the LA's BEST students who participated in the follow-up periods than would a group of students who were never enrolled in the program. Thus, these students make a superior comparison group than those students who have never attended the program. Moreover, we would expect the low attendance program participant group to share many of the characteristics of the comparison group. The

[1] Defined as intensity of participation in this study.

findings on these two groups should be quite similar and may serve as additional validation of dosage effects. As we hypothesized, if these two groups demonstrate similar negative or neutral effects, while the moderate and regular attendance groups demonstrate escalating positive effects, the treatment (LA's BEST experience) demonstrates promise, making dosage a critical factor in determining program success. The following describes the characteristics of the two cohorts.

Grade 2 cohort (2003–04). Four years of student demographic and academic information were available for this cohort spanning the 2003–04 school year through the 2006–07 school year. This cohort included 3,734 students in the math sample and 3,716 students in the English-language arts sample. Students in this cohort are predominantly Latino (88.7%), slightly more than half are female (51.3%), about two-thirds (66.4%) are classified as Limited English Proficient (LEP), and about one-third (32.7%) have parents who either graduated from high school or attended some level of college.

Grade 3 cohort (2003–04). Four years of student demographic and academic information were available for this cohort. Subsequently, students who were in Grade 3 in 2003–04 were followed from 2002–03 through 2005–06, their projected year in Grade 5 barring retention. This cohort included 2,332 students in both the math and English-language arts samples. Participants in this cohort are primarily Latino (90.5%), half are female (50.0%), over two-thirds are classified as LEP (79.6%), and about one-third (34.4%) have parents who either graduated from high school or attended some level of college.

Defining Attendance Intensity

As was previously mentioned, we separated students in each cohort into four groups based on their intensity of attendance in the LA's BEST program. Within each cohort, we classified students who attended LA's BEST over 20 days at baseline, but did not attend LA's BEST during the follow-up period as Group 1. We classified students attending 1–20 days on average during the follow-up period as Group 2, and those attending 21–99 days during follow-up as Group 3. We defined regular attendance (Group 4) as those students who averaged 100 days or greater of LA's BEST attendance per year during the follow-up period. Table 1 displays the criteria used to define the four groups.

Table 1

Group Definitions for the Grade 2 and Grade 3 Cohorts

Group	Baseline	Follow-up period
Group 1	Attended LA's BEST (Over 20 days)	Did not attend LA's BEST
Group 2	Did not attend LA's BEST	Attended LA's BEST (1–20 days per year)
Group 3	Did not attend LA's BEST	Attended LA's BEST (21–99 days per year)
Group 4	Did not attend LA's BEST	Attended LA's BEST (100 or more days per year)

Note. Group 1 constitutes the comparison group whereas Groups 2–4 constitute the program participant groups.

Data Analysis Methods

In order to examine the effects of LA's BEST on student achievement, we employed an HLM design that has the advantages of directly modeling trajectories and being more flexible than traditional analyses. This design allows us to move beyond traditional pre/post analysis, which is limited by data requirements and explanatory possibilities (Rogosa, Brandt & Zimowski, 1982; Raudenbush & Byrk, 2002). With an HLM approach we account not only for student level variation, but also for variation across schools that may influence student achievement outcomes. Two separate HLM approaches were used. First, we employed a "residual gain" approach to compare group outcomes in each cohort. This process is often described as a "value added" analysis because the performance at the end of the follow-up period is compared to the "best prediction" for each student based on their baseline characteristics. Because each student's LA's BEST program group membership is not considered in the HLM model that produces the best prediction, any difference between actual performance and the best prediction that varies substantially from zero for any given group would represent the value added by the group membership.[2] Second, we employed a mixed model (HLM) approach to the pooled data using a factor to represent the groups.

There were two necessary steps in the residual gain approach. In Step 1, the achievement outcome was predicted using baseline variables such as the initial achievement score, regular school attendance, gender, parent education level, language proficiency status and race/ethnicity. Note that in Step 1, the different levels of intensity of participation were not included in the model. This was done so that the prediction of achievement at the follow-up year could be obtained without this knowledge. In Step 2, the difference between actual achievement at the end of the follow-up period and the model-based predictions (obtained at

[2] In this study, the intensity of participation represents the value added by the group membership.

Step 1 without the knowledge of intensity of attendance) were compared among each group descriptively.

The second approach was applied to pooled data across the two cohorts. This mixed model approach also employed HLM but this time a factor was included to represent the four groups. As with the first HLM approach achievement outcomes at the end of the follow-up period were predicted and baseline variables such as the initial achievement score, regular school attendance, gender, parent education level, language proficiency status and race/ethnicity were included as covariates. Estimated means for each group were computed controlling for the baseline variables and 95% confidence intervals for the differences between the comparisons are displayed.

Results

In order to provide more clarity to our analyses, the modeling results will be presented by content area: Math and English-language arts. The synthesis of the results will be presented in the Discussion and Conclusion section.

Math Results

For math achievement, we conducted residual gain analyses for each cohort and then employed mixed modeling to the pooled data (across the two cohorts) using a factor to represent the four groups.

Residual gain analyses. Table 2 displays the difference between predicted and actual math achievement outcomes for the grade 2 cohort. Predicted outcomes were obtained from the Step 1 model that accounted for baseline achievement and background characteristics but ignored group membership (intensity of attendance). Means of each student's actual minus predicted scores shows that students in Group 4 scored about 8 scale points higher than the model predicted. In contrast, students in Groups 1 and 2 on average scored lower than the model predicted. As can be seen in Table 3, residual gains for the grade 3 cohort follow a similar pattern. Once again, students in Group 4 scored higher than the model predicted whereas, students in Groups 1 and 2 on average scored lower than the model predicted.

Table 2

Residual Gains by Group, Math Achievement of the Grade 2 Cohort

		Actual scale score		Predicted scale score		Residual gains (actual-predicted)	
	N	Group means	*SD*	Group means	*SD*	Group means	*SD*
Group 1	430	335.94	76.81	338.96	50.04	-3.02	55.21
Group 2	1054	332.71	74.67	336.26	50.09	-3.55	53.85
Group 3	1329	337.58	74.35	337.60	50.62	-0.01	53.86
Group 4	648	345.57	76.03	337.76	48.23	7.81	56.22

Table 3

Residual Gains by Group, Math Achievement of the Grade 3 Cohort

		Actual scale score		Predicted scale score		Residual gains (actual-predicted)	
	N	Group means	*SD*	Group means	*SD*	Group means	*SD*
Group 1	257	329.08	76.62	331.25	54.27	-2.18	58.31
Group 2	809	323.60	75.94	326.08	48.95	-2.49	55.81
Group 3	883	327.78	74.00	327.79	49.25	-0.01	55.07
Group 4	383	345.72	84.36	338.98	55.52	6.74	58.90

Analyses of pooled data. To test the significance of the differences found in the residual gain analyses, a mixed model was performed controlling for baseline variables with the LA's BEST grouping variable included in the model. Estimated means were produced for each group and differences between the regular attendance group (Group 4) and the three other groups were tested for significance. In Table 4 descriptive statistics for math achievement are presented at the baseline and follow-up time points for the pooled data. The mixed model results in Table 5 reveal that students in Group 4 (with regular attendance) performed significantly better than students in the other three groups ($p < .05$). The significance values shown in Table 3 apply a Bonferroni correction to account for multiple comparisons.

This suggests that students who attended LA's BEST regularly during the follow-up period performed better than expected when compared to students who did not attend the program or who attended the program with less regularity. The estimated mean at the end of

the follow-up period was about 12 scale points higher for the regular attendance than for both the group with no LA's BEST attendance and the group with low LA's BEST attendance in the follow-up period. It should be noted, however, that the effect size of these differences was small when compared to the standard deviations of the actual group means at follow-up (see Table 4).

Table 4

Baseline and Follow-up CST Scale Score Means and Standard Deviations, Math Achievement of the Pooled Cohorts

		Baseline		Follow-up	
	N	Mean	*SD*	Mean	*SD*
Group 1	687	337.93	67.64	333.37	76.76
Group 2	1863	333.46	66.39	328.75	75.34
Group 3	2212	336.38	67.97	333.67	74.35
Group 4	1031	339.34	68.95	345.62	79.18

Note. CST = California Standards Test.

Table 5

Multiple Comparisons Estimated Mean Differences (to Group 4), Math Achievement of the Pooled Cohorts

	Mean estimated scale score difference	Std. error	Significance value	95% Confidence interval	
				Lower bound	Upper bound
Group 4 - Group 1 (342.58 - 330.74)	-11.84	2.88	0.000	-18.743	-4.930
Group 4 - Group 2 (342.58 - 330.81)	-11.77	2.29	0.000	-17.261	-6.278
Group 4 - Group 3 (342.58 - 330.74)	-8.23	2.20	0.001	-13.492	-2.972

English-Language Arts Results

As with math achievement, we conducted residual gain analyses for each cohort and then employed mixed modeling to the pooled data using a factor to represent the four groups.

Residual gain analyses. Tables 6 and 7 present the difference between predicted and actual English-language arts outcomes for both Grade 2 and Grade 3 cohorts. The means of each student's actual minus predicted scores shows that average differences for students in

each group were within approximately 1 scale score point. This suggests that the prediction model produced fairly accurate predictions of actual scores for each group and attendance intensity did not have an effect on the predicted score.

Table 6

Residual Gains by Group, English-Language Arts Achievement of the Grade 2 Cohort

		Actual scale score		Predicted scale score		Residual gains (actual-predicted)	
	N	Group means	*SD*	Group means	*SD*	Group means	*SD*
Group 1	423	321.78	46.07	321.16	31.23	0.62	31.22
Group 2	1044	318.11	42.70	319.19	39.57	-1.08	30.32
Group 3	1327	322.56	43.21	322.40	30.98	0.16	30.09
Group 4	648	326.08	44.17	325.07	30.82	1.01	31.11

Table 7

Residual Gains by Group, English-Language Arts Achievement of the Grade 3 Cohort

		Actual scale score		Predicted scale score		Residual gains (actual-predicted)	
	N	Group means	*SD*	Group means	*SD*	Group means	*SD*
Group 1	256	317.53	44.68	316.71	33.38	0.818	31.54
Group 2	808	315.90	44.91	316.43	31.26	-0.53	31.97
Group 3	883	318.75	43.28	319.01	30.55	-0.26	28.86
Group 4	383	327.77	46.47	326.59	33.52	1.18	29.68

Analyses of pooled data. As with the math achievement data, findings from the residual gain analyses for English-language arts were confirmed using a mixed model that controlled for baseline variables with the LA's BEST grouping variable included in the model. Estimated means were produced for each group and differences between the regular attendance group (Group 4) and the three other groups were tested for significance. In Table 8, descriptive statistics for English-language arts achievement are presented at the baseline and follow-up time points for the pooled data. Results in Table 9 indicate that there was no significant difference among the groups ($p > .05$), confirming that program attendance and intensity of attendance did not influence English-language arts performance.

Table 8.

Baseline and Follow-up CST Scale Sore Means and Standard deviations, English-Language Arts Achievement of the Pooled Cohorts

		Baseline		Follow-up	
	N	Mean	*SD*	Mean	*SD*
Group 1	679	308.85	48.90	320.18	45.57
Group 2	1852	307.51	46.33	317.15	43.68
Group 3	2210	311.36	47.36	321.03	43.27
Group 4	1031	317.36	48.10	326.71	45.03

Note. CST = California Standards Test.

Table 9.

Multiple Comparisons Estimated Mean Differences (to Group 4), English-Language Arts Achievement of the Pooled Cohorts

	Mean estimated scale score difference	Std. error	Significance value	95% confidence interval	
				Lower bound	Upper bound
Group 4 - Group 1 (321.76 - 320.68)	-1.08	1.59	1.000	-4.884	2.721
Group 4 - Group 2 (321.76 - 319.29)	-2.47	1.26	0.149	-5.484	0.544
Group 4 - Group3 (321.76 - 320.37)	-1.39	1.20	0.743	-4.277	1.492

Discussion and Conclusion

As with the 2008 report (Huang et al.), 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. 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 following two cohorts of students longitudinally for 4 years and by employing statistical strategies to reduce self selection bias, thus improving the comparability of students with different levels of attendance intensity.

In the current study, it was hypothesized that the regular attendance group would duplicate the results of the 2008 study and the no attendance group would mirror the results

of the low attendance group. The current findings reaffirm this hypothesis. Results of the analysis show that regular attendance (of 100 days or more per year) in the LA's BEST program is significantly associated with positive CST math achievement growth when compared to students with low or no attendance in the program during the follow-up period. The statistical tests were performed on pooled data across two cohorts of students. The differences in achievement growth for the comparison groups as measured by residual gains were consistent across the two separate cohorts of students whom we followed over a period of 4 years. Furthermore, it is important to note that these results were obtained after carefully accounting for existing differences in students' background characteristics, so that the most plausible explanation of these statistical differences is in the regularity of LA's BEST attendance.

In terms of student's English-language arts achievement, the results reveal that students with regular attendance (100 days or more per year) during the follow-up period did not have significantly higher CST English-language arts residual gains when compared to students who did not attend the program or had low (1–20 days) or moderate (21–99 days) attendance intensity during the follow-up period. This finding is also consistent across both cohorts, suggesting that additional focus in this content area will be needed in order to demonstrate significant results. At the same time, it is also important to keep in mind that two-thirds or more of LA's BEST students are English Language Learners and speak a second language. Additional factors such as language spoken at home, opportunities to read and communicate in English, student motivation/engagement, parental/peer support, community environment and so forth may all influence students' language development and should be examined further in future studies.

Conclusion

In conclusion, this report provides additional evidence to the 2008 report (Huang et al.) that LA's BEST is a program that shows promise in improving students' CST math performance. However, as with any intervention project, students need to attend regularly in order to reap the program benefits. The current study suggests that 100 or more days of annual attendance is necessary. Implications from this study also 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.

References

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.

Huang, D., Leon, S., La Torre, D., & Mostafavi, S. (2008). *Examining the relationship between LA's BEST program attendance and academic achievement of LA's BEST students* (CRESST Tech. Rep. No. 749). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. A., Snow, D., & 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.

McComb, E. M., & Scott-Little, C. (2003). *After-school programs: Evaluations and outcomes*. Greensboro, NC: University of North Carolina, Southeastern Regional Vision for Education (SERVE).

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

Rogosa, D., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin*, *92*(3)*,* 726–748.