

**An Evaluation of After School Programs
Provided by Beyond the Bell's Partner Agencies**

**Executive Summary
&
Synopsis of Methodology and Findings**

**Submitted to:
Richard Roberts,
Beyond the Bell Branch, LAUSD**

**By
Steven Frankel, President
Research Support Services (RSS)
Marina del Rey, CA**

**Glenn Daley, Director
Program Evaluation and Research Branch
Los Angeles Unified School District**

March, 2007

Table of Contents

EXECUTIVE SUMMARY	3
Summary of Findings.....	3
Policy Implications	5
Assistance from PERB.....	6
Further Information.....	7
SYNOPSIS OF METHODOLOGY AND FINDINGS.....	8
Section One – Overview	8
1.1 Purpose of Evaluation.....	8
1.2 Agencies Whose Students Were Included	8
1.3 Demographic Characteristics of Children in the Annual Evaluation	10
1.4 Data Sources & Outcome Measures	11
Section Two – Why Use a Value-Added Model?.....	12
Section Three – English Language Arts	15
3.1 Elementary School English Language Arts CST Scores.....	15
3.2 Middle School English Language Arts CST Scores	19
Section Four – Mathematics	22
4.1 Elementary School Mathematics CST Scores.....	22
4.2 Middle School Mathematics CST Scores	24
Section Five – School Attendance.....	27
5.1 Elementary School Attendance.....	27
5.2 Middle School Attendance.....	30
Section Six – Teacher-Assessed Behavior	33
6.1 Elementary School Teacher-Assessed Behavior.....	33
6.2 Middle School Teacher-Assessed Behavior Ratings	35

Executive Summary

Summary of Findings

An Evaluation of After School Programs Provided by Beyond The Bell's Partner Agencies explores the hypothesis that participation in after school programs results in academic, attendance and behavior improvements that only occur when students attend a minimum number of daily sessions.

The study is based upon 2004-2006 data describing more than 50,000 Los Angeles Unified School District (LAUSD) after school program participants, of which 97% are students of color; 92% qualify for free- or reduced-price lunches; and only 28% are English-only or native English speakers. As demonstrated by these figures, it is clear that Beyond The Bell (BTB) and its partner agencies serve a huge, largely at-risk population that would provide daunting challenges for any school district in the nation.

Both providers and supporters of after school programs agree that their primary purpose is to keep children safe in a stimulating and caring environment. However many advocates and practitioners strongly feel that these programs also improve the academic performance, attendance and behavior of many of the participants.

Most after school program evaluations have not supported these contentions. This has caused many educational policy makers to assume that due to the limited time and scope of the academic components of after school programs, and the fact that most of the staff working with the children have had very limited training in teaching, the expectation that after school programs should contribute to academic gains is unrealistic. To a lesser degree, the same lack of expectations applies to improvements in school attendance and behavior; with the argument being that most after school programs probably don't exert a sufficiently strong influence on children to expect those kinds of benefits to carry over to the "regular" school day.

In this study we hypothesized that if students attend at least a minimum number of times a year (achieving what is sometimes referred to as a *minimum dosage*), such benefits may be realized. We tested this hypothesis using a *value-added design* in which evaluators build their studies around the question of what a given program contributes to student's learning or behavior; above and beyond what the students would have achieved anyway had they not participated in the program.

We learned in this study that the minimum annual attendance levels (*dosages*) at which improvements in California Standards Test (CST) scores and improvements in school daily attendance occur are 100 days per year for elementary students and 50 days per year for middle school students. Improved teacher-assessed behavior ratings for middle school students also occur at the 50-day level.

These findings were determined by two sets of analyses:

- The mean CST scores, daily attendance rates and behavior ratings of the *high dosage* students were compared to the performance of the lower dosage students, and were found to be greater by amounts that were both educationally and statistically significant.
- Students who participated in the after school program (ASP) in 2005-06 were grouped based on the number of days they attended; and the performance of each group was compared to the performance of the same group of students a year earlier. To make these comparisons more valid, variables that would have biased the findings were statistically filtered out so that all that was being measured was the *value added* by the programs. In the high dosage cohorts, improvements were greater than those predicted by the statistical *value-added model* whereas this did not occur in any of the lower dosage cohorts.

The variables that were filtered out by the statistical value-added model were students' previous CST scores and attendance levels from the 2004-05 school year, ethnicity, gender, English language proficiency and school lunch eligibility; as well as their schools' average test scores, attendance levels, school lunch participation and size. This permitted the researchers to document the effect of after school attendance, without the other variables influencing the results.

In addition to demonstrating that high dosage groups did demonstrably better than their low dosage counterparts, the study also found that the BTB after school programs are unlikely to improve the CST scores, daily attendance and behavior ratings of students who do not reach the minimum dosage levels. This has policy implications which are discussed below.

Some caution must be exercised in applying the findings of this study. It will be noted that, when examining the value-added graphic displays in the companion report, the *Synopsis of Methodology and Findings* the value-added model always predicted that the high dosage groups would improve their CST performance from the year before, even though factors including past test performance and school test performance were corrected for by the model. And, in all cases, these groups improved their test performance even more than the model predicted. Similarly, groups of low dosages students were predicted to perform worse than they did the year before; and in every case they did perform as badly – or worse – than the model predicted.

There are two possible reasons for this: 1) the high dosage groups include many students who have been high dosage after school participants for several years and the positive effects of these experiences are compounding their gains; or, 2) Many of students in the high dosage group are *self-selected* and benefit from *innate advantages* that can't be corrected by any statistical model such as higher parental aspirations for their children or the availability of homework help, tutoring and after school supervision by parents or older siblings.

It is likely that this issue of *causality*, which often arises when studying issues such as whether students perform better in parochial schools because of the programs or because of the children who are enrolled, will never be resolved entirely. However it is safe to say that – regardless of the reasons – the value-added model predicted that the high dosage groups would improve their CST performance and they always did so; and the value-added model predicted that the low dosage groups' academic performance would decline and this always occurred.

We plan to learn more about the issue of whether gains or losses were caused by the dosage levels, or by innate factors for which the value-added model couldn't account, through this year's qualitative study. A major goal of that study will be to determine, through student and parent interviews and questionnaires, whether the students in the high dosage groups enjoy innate advantages over their peers in the low dosage groups, or whether most of the differences in performance can be attributed to the after school experience.

Policy Implications

Most after school programs provided by outside agencies have attendance-based funding. In other words, they receive a fixed amount of money per day for each day their participants attended. This *earned reimbursement* strategy, and the agencies' commitments to serving their communities, results in them devoting considerable effort to keeping their programs full. While Proposition 49 makes the relationship between attendance and funding less direct, it still maintains an earned reimbursement approach by insisting that agencies either meet their attendance projections or face financial sanctions.

The grant language of the federal 21st Century Learning Grants and the California After School Education & Safety (ASES) programs includes expectations that agencies will strive for minimum dosage levels of 5 days a week for elementary students and 3 days a week for middle school students. This study suggests that these requirements require higher dosage than is necessary to achieve significant achievement results.

Nevertheless, the results of this study emphasize the importance of dosage and support the notion of BTB working in partnership with its agencies to at least meet the minimum dosages found to be effective in this evaluation (100 days a year for elementary students and 50 days a year for middle school students) so that these become minimums for most students. While these levels are appreciably below the 5- and 3-day a week dosages called for in the legislation (those work out to 180 days for elementary students and 108 days for middle school students), they are a good starting point. Also – based on this study – they will lead to improvements in elementary and middle school students' CST scores and attendance; and to improvements in middle school students' teacher-based behavior ratings as well.

Improving dosage levels will likely engage BTB and its partner agencies in several new initiatives:

- Learning more about the agencies, program sites, program offerings, climate and incentives that are causing students to attend for at least the minimum dosage levels.
- Monitoring student dosage on an ongoing basis; and providing sites with management reports showing monthly, quarterly and annual dosage trends for program sites and individual students.
- Considering program changes and incentives that will drive dosage rates even higher. At least one of the largest agencies already requires school sites with waiting lists to remove from their programs students are not meeting minimum dosage levels. Others are

offering attendance awards and incentives. The concept of exploring other kinds of incentives for agencies and schools exceeding minimum dosage levels (e.g., 80% of the students attending for 80% of the days the program is offered) is also worthy of consideration.

One way of learning more about how agencies improve their dosage rates will be to revise the qualitative studies of after school programs to ensure that adequate numbers of both high and low dosage program sites are included in the sample; and that much more emphasis is placed on documenting the features and strategies employed by programs that contribute to high dosage rates. Also, surveys aimed at determining which kinds of program offerings and features are likely to lead to greater participation should be directed at high dosage, low dosage and nonparticipating students and parents. A related strategy will be to revise quantitative studies (such as this one) so that they can delve into the retention effects of students participating in specific kinds of activities. They also need to incorporate more effective ways of documenting improvements in student behavior (such as analyzing suspensions by reason codes); and developing statistical profiles of those kinds of students who are most likely to participate in these programs at high dosage levels.

Studies of high school after school programs are also needed. These should be aimed at determining the program offerings, schedules, incentives and minimum dosage levels that will help students to pass the high school exit examinations, stay in school and graduate. These studies are critical if fewer students are to be “left behind.”

But above all, it should be emphasized that the benefits of the after school programs documented in this report are in addition to the other – very substantial – benefits that after school programs provide. These include safety, warmth, resiliency and individual attention for students. In the long run these benefits are probably even more important than those documented in this study. For these reasons, it is critical that as the BTB programs increase their average dosage levels, the needs of parents and students who only require “drop-in” programs not be forgotten.

However, it is likely that as parents and students learn about the benefits of high dosage after school programs, fewer will opt for the “drop-in” alternative since students who attend for at least the minimum dosage levels are likely to experience the “best of both worlds”: the traditional benefits provided by most after school programs and the academic, attendance and behavior improvements that high dosage after school programs provide.

Assistance from PERB

Grateful acknowledgement is made to Glenn Daley, Director of LAUSD’s Program Evaluation Research Branch (PERB), who developed and guided the use of the value-added analytic techniques upon which these analyses are based and is the co-author of this report.

Further Information

This *Executive Summary* is intended for policy makers, practitioners and other interested parties who need a brief, non-technical summary of the study's findings and policy implications. The companion report, *Synopsis of Methodology and Findings*, that follows includes a relatively non-technical description of the study's methodology and findings; and the tables and statistical charts that support the findings and policy recommendations contained in this *Executive Summary*. There are also 2 versions of the full report which includes a third volume: *Work Papers*. This is a highly technical 200 page report which contains all of the statistical commentary, tables and charts developed by the project staff; reproduced almost exactly like it was originally written. A public version is complete except for the agency comparisons; and a restricted version includes detailed findings related to the performance of the individual private agencies.

Synopsis of Methodology and Findings

Section One – Overview

1.1 Purpose of Evaluation

An Evaluation of After School Programs Provided by Beyond The Bell's Partner Agencies explores the hypothesis that participation in after school programs results in academic, attendance and behavior improvements that only occur when students attend a minimum number of daily sessions.

We know from previous studies that the overall relationship between after school attendance and in-school achievement variables such as test scores, attendance rates and behavior ratings is usually not that high; but in this study we hypothesized that if students attend at least a minimum number of times a year (achieving what is sometimes referred to as a *minimum dosage*), such benefits may be realized. We tested this hypothesis using a *value-added design* in which evaluators build their studies around the question of what a given program contributes to student's learning or behavior; above and beyond what the students would have achieved anyway had they not participated in the program.

This study uses California Standards Test (CST) *scale scores* that range from 150-600 to measure improvements in academic performance. The scale scores have proven to be far more sensitive than the *categorical scores* (Far Below Basic, Below Basic, Basic, Proficient & Advanced) that are often used by school districts and the California Department of Education (CDE) to report CST results. Unlike the categorical scores, the scale scores (which are also provided by CDE) report test score improvements made within a category (e.g., a group of students moving from a mean scale score of 305 in the 2005 Math CST to a mean scale score of 345 in the 2006 Math CST). This increase represents a huge gain of 40 scale score points even though both scores are still located within the Basic category. If categorical scores were used to report the same data, the students would be listed as not having shown any progress; since they hadn't move from Basic (scale scores of 300-349) to Proficient (scale scores of 350-399).

1.2 Agencies Whose Students Were Included

Exhibit 1.2.1 lists the community-based agencies whose students were included in the study.

Exhibit 1.2.1: Agencies Providing Services

Programs At Elementary Schools	Art Share Los Angeles A World Fit For Kids! Boys and Girls Clubs P. F. Bresee Foundation Building Up Los Angeles Carney Educational Services City of San Fernando Children, Youth and Family Collaborative Gang Alternatives Program Great Beginnings for Black Babies Kids Protectors of the Environment Learning for Life Para Los Niños Center for Counseling S.T.A.R. Inc. Woodcraft Rangers
Programs at Middle Schools	After School All-Stars A World Fit For Kids! Boys and Girls Clubs P. F. Bresee Foundation Building Up Los Angeles Carney Educational Services Literacy, Arts, Culture, Education, Recreation Martin Luther King Legacy Foundation Teen Prime Time Woodcraft Rangers YWCA of Greater Los Angeles

Records from more than 50,000 participants were analyzed for this evaluation; making it one of the largest studies of its type conducted in a single school district. While data from the three latest school years were analyzed, this *Synopsis of Methodology & Findings* focuses on the latest two school years with the 2004-05 data used to establish baselines and the 2005-06 data used to assess gains. These data, analyses and findings have been labeled in the full report the *Annual Evaluation*.

A similar set of analyses were conducted for three years (2005-06, 2004-05 and 2003-04), with the 2003-04 data used as a baseline. The findings from these analyses were very similar and never contradictory; but they didn't include data for the 2005-06 3rd graders since it isn't until the 4th grade that students have accumulated 3 years of CST scores. Therefore the decision was made to have the *Synopsis of Methodology & Findings* focus on the Annual Evaluation which does include 3rd grade CST scores for the 2005-06 cohort. The 3-year data, analyses and findings have been labeled the *Biannual Evaluation* and are included in the full report.

1.3 Demographic Characteristics of Children in the Annual Evaluation

Exhibit 1-2 describes the participants included in the Annual Evaluation of which 97% are students of color; 92% qualify for free- or reduced-price lunches; and only 28% are English-only or native English speakers. As demonstrated by these figures, it is clear that BTB and its partner agencies serve a huge, largely at-risk population that would provide daunting challenges for any school district in the nation. The totals in this exhibit reflect the fact that demographic data were missing for some participants.

Exhibit 1.4.1: Demographic Characteristics of 2005-06 Students

Gender

	<i>Frequency</i>	<i>Percent</i>
Female	24,896	48.5
Male	26,400	51.5

Ethnicity

	<i>Frequency</i>	<i>Percent</i>
African-American	7,151	13.9
Hispanic	40,589	79.1
White	1,643	3.2
Other	1,913	3.7

Meal Program Eligibility

	<i>Frequency</i>	<i>Percent</i>
No	4,018	7.8
Yes	47,278	92.2

Language Classification

	<i>Frequency</i>	<i>Percent</i>
EO: English only/Native Speaker	14,127	27.5
IFEP: Initial Fluent English Speaker	4,752	9.3
LEP/ELL: English Language Learner	19,907	38.8
RFEP: Reclassified Fluent English Speaker	12,510	24.4

Generally speaking, this is a population whose CST scale scores tend to be highest when they're in grades 2-4 and then decline steadily for the remainder of their school careers. Only about half typically graduate. Thus intervention programs, such as after school programs that can either

slow the academic decline or reverse its direction, for substantial numbers of students are probably worthy of strong support.

1.4 Data Sources & Outcome Measures

Most of the data for this evaluation came from the School Information Branch of LAUSD which provided the demographic, test score, daily attendance and behavior rating data. The individual after school agencies provided the after school attendance data.

The outcome measures used in this evaluation are based on those required by the state and federal agencies funding the BTB programs: academic performance, behavior, and regular school attendance. Specifically:

- Academic performance is assessed using CST English Language Arts scale scores and the CST Mathematics scale scores.
- Regular school attendance is assessed as the percentage of days a student was present out of the total number of days during the 2005-06 school year for which the student was enrolled (typically 180 days).
- Behavior is assessed using teacher-assigned ratings for behavior that are added annually to each student's permanent record. These ratings are reported differently for elementary and middle school students. Elementary school teachers assign ratings for 14 different items involving behavior, cooperation and work habits. Only 4 ratings are used in middle school. The Behavior ratings used in these analyses were the sum of the 14 or 4 ratings, and this score was only calculated for students that were not missing any of the individual ratings.
- After School Attendance is recorded as being the number of days for which the individual children, who had attended a BTB after school program for at least one day, were signed into their BTB programs. Because of bookkeeping problems at the sites and problems with missing or multiple student ID numbers, these counts are sometimes lower than the actual number of days some students attended. Improvements in attendance recording procedures that have already implemented will likely result in improved attendance counts starting in December 2006.

Section Two – Why Use a Value-Added Model?

Both providers and supporters of after school programs agree that the primary purpose of BTB programs is to keep children safe in a stimulating and caring environment after the regular school day ends. However many practitioners strongly feel that these programs also improve the academic performance, attendance and behavior of many of its participants.

Most after school program evaluations have not supported these contentions. This has caused many educational policy makers to assume that due to the limited time and scope of the academic components of after school programs, and the fact that most of the staff working with the children has no training in teaching language or math, the expectation that after school programs can contribute to academic gains is unrealistic. To a lesser degree, the same lack of expectations has often spilled over to improvements in school attendance and behavior, with the argument being that after school programs don't exert a sufficiently strong influence on children to expect their benefits to carry over to the "regular" school day.

Many of the studies that have failed to find linkages have relied on *quasi-experimental matched-pair designs* in which the children who have participated in the after school program were compared to children from the same school who haven't participated in that particular after school program. However, the lack of positive findings may have been due to how these designs have typically been implemented.

In most cases, the matches have been made primarily on whether or not a child has attended as little as a single session of the after school program, and then on race, sex, ethnicity, free lunch status, English language proficiency rating and previous achievement test scores. While this methodology is inexpensive to implement because all the data are available from district records, and appears to be very comprehensive, it in fact ignores two key considerations:

- *How did the "non-participating" students spend their after school time?* Were they receiving *as much or more* academic support and/or enrichment than the after school program provided (e.g., by being supervised and tutored at-home by parents or older siblings, or participating in a competing after school program, intensive academic program, team sports, performance art or church programs)? Or, were they indeed getting little enrichment or academic support as most matched-pair designs implicitly assume?
- *How intensive was the after school program participation of students included in the treatment group?* Would being enrolled in an after school program and attending for one or two days during the entire school year qualify them for the "treatment group;" or would they have to meet minimal guidelines such as having attended for at least 20 days or 50 days a year?

Unless both of these variables (the after school experiences of the comparison group and the minimum participation levels of the "treatment groups" are given careful consideration in quasi-experimental designs of after school program evaluations, the results are likely to be inconclusive, or hopelessly confounded.

For this study, BTB managers sought permission from PERB to have the contractor, Research Support Services (RSS), perform analyses based on a value-added model. PERB not only granted permission, but they used their strong expertise in value-added models to play a leading role in revising the study's design.

The value-added analyses are based on the concept of *dosage*, which has been included in previous BTB evaluations conducted by RSS. For the purposes of this report, *dosage* is defined as the number of days each student participated in a Beyond the Bell program during the 2005-06 school year. The concept of *dosage* is different from that of *attendance*. Consider two hypothetical after school programs which both accrued a total of 18,000 attendance days during the 2005-06 school year:

- Program A earned the 18,000 attendance days by having the same 100 students attend for all 180 days that the program was offered. This is an extreme example of a *high-dosage* program in which all the students had 100% attendance.
- Program B earned the 18,000 attendance days by having 500 students attend over 180 days. In this program the average student attended only 36 of the 180 days, giving them an average attendance rate of 20%. This is an extreme example of a *low-dosage* program.

The primary hypotheses explored by the value-added analyses in this study was that under these circumstances the 100 students in Program A would realize significant gains in academics, attendance and behavior; whereas the 500 students in Program B would not benefit in these ways.

In the value-added analyses, students who did not attend any program sessions were eliminated from consideration since there could be no "value-added" for them. Thus the value-added analyses examined 4 cohorts of students consisting of those attending BTB programs in 2005-06 for:

- 1-20 days
- 21-50 days
- 51-100 days
- More than 100 days

The analyses focused on the degree to which each of these "dosages" was related to improvements in the outcome variables. In the value-added analyses, pre-existing differences between the students in these groups were controlled statistically using a two-stage process. In the first stage a "statistical filter" was constructed which removed the effects of the following variables from the outcome measures. The following *student demographic characteristics* were filtered out:

- Level of performance on the outcome being evaluated (CST English/Language Art, CST Math, Daily Attendance, and Behavior Ratings) the year before (2004-05).
- Ethnicity (Black, Hispanic, White & Other)
- Gender (Male & Female)
- Language classification (EO, IFEP, LEP & RFEP)

- Meal program participant (Yes & No)

The following *school characteristics* were also filtered out:

- Mean baseline value on the achievement, attendance or behavior outcome being evaluated
- Percentage of students in the school's meal program
- Number of students attending the school

This filter was applied to each grade separately creating *standardized residuals* for that grade. The standardized residuals predicted how students for that grade *should* score on each outcome measure *if the after school program dosage had no effect*. Then these theoretical *predicted scores* were compared to the actual scores to determine the *value added* by the after school programs.

A more detailed description of the regression analyses that were used to create the standardized residuals is included in the full report. Here it is sufficient to say that the grade level residuals enabled the researchers to:

- Graphically depict the degree to which students in each after-school attendance cohort exceeded, or failed to reach, the performance predictions emerging from these *value-added* analyses (hence their name) that incorporated the statistical filter to eliminate demographic and school-level effects.
- Graphically depict the degree to which the students attending programs offered by each individual agency exceeded, or failed to reach, the performance predictions emerging from the value-added analyses.

These value-added analyses were complemented by statistically testing whether differences in performance for the students who met minimum dosage levels (100 days for elementary students and 50 days for middle school students), and those that did not meet those levels, were both *practically* and *statistically* significant. These differences were tested *without* the statistical filters that are used in the value-added analyses being applied.

The results of all of these analyses comprise the remainder of the *Synopsis of Methodology & Findings*. More detailed analyses (e.g., those for each grade level rather than for all elementary or middle school students) are included in the full report.

Section Three – English Language Arts

3.1 Elementary School English Language Arts CST Scores

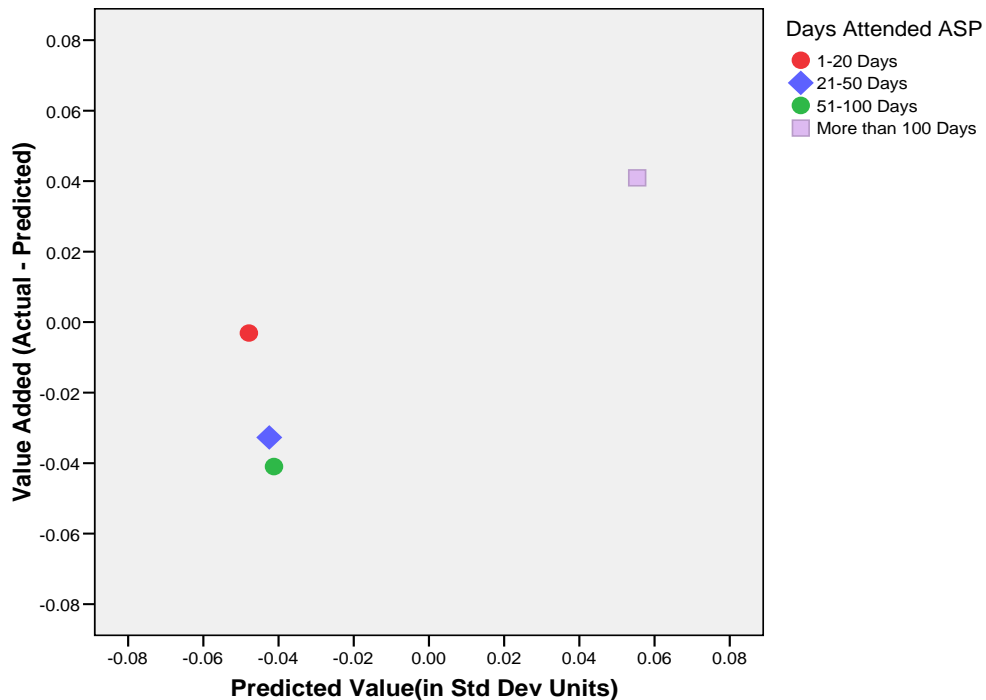
The study found statistically significant differences between the average English Language Arts (ELA) CST scores for students in the high dosage group (“More than 100 Days”) compared to students in the low dosage group (“1-100 days”).

Exhibit 3.1.1: Descriptive Statistics of English Language Arts CST Scale Scores for Elementary School Students Attending 1-100 vs. More than 100 Days

DOSAGE	<i>n</i>	<i>M</i>	<i>SD</i>
Low: 1-100 Days	10,089	314.94	50.62
High: > 100 Days	6,150	319.97	51.85

This exhibit shows that students who attended more than 100 days on average scored 5 scale score points higher than students who attended less than 100 days. The difference between these two groups of students was statistically significant, $t(12747,16) = -6.06, p < .01$.

Exhibit 3.1.2: Elementary School English Language Arts CST Value-Added Analysis



Residuals computed separately for each grade

The key to interpreting this kind of graph is to understand what the x-axis (Predicted Value) and the y-axis (Value Added) represent.

- **Students are sorted along the x-axis according to previous academic achievement.** In many cases, in a simplified model, previous test scores alone predict more than 50% of the variation in current test scores, with other factors such as ethnicity, sex, English language proficiency, school, etc., accounting for additional variation. Thus the right or positive end of the x-axis represents generally high-achieving students, and the left or negative end of the x-axis represents generally low-achieving students.

In this case, the model predicted that only the students in the “More Than 100 Days” group would increase their scores. The amount of the increase or decrease is in *Z-scores*; which measure the distance of a point from the mean score of a distribution in terms of *standard deviations*. A *Z-score* of +1.00 means that the point is one standard deviation above the mean; and in this case the model predicted that the “More Than 100 Days” group would increase their scores by about 6% of a standard deviation (+0.06 on the x-axis); whereas the model predicted that the students in all the other groups would see their scores drop by about 4% of a standard deviation (-0.04). Given that the standard deviations shown in Exhibit 3.1.1 were about 50 scale score points, an increase of 6% translates to a predicted increase of 3 scale score points; and a decrease of 4% translates to a predicted decrease of 2 scale score points.

- **The y-axis shows the residual: the academic gain or loss experienced by each student in the current year relative to the average for SIMILAR students.** Here we see that the students in the “More Than 100 Days” group actually scored about 4% better (equal to about 2 scale score points for this distribution) than the model predicted [look upwards from the 0.00 point on the y-axis], whereas the students in the other 3 groups scored about 4% worse (equal to about 2 scale score points) than the statistical model predicted.

Among students with high predicted achievement, those who participated in after-school programs were BOTH more likely to have high dosage rates than low dosage rates AND more likely to have positive value added in achievement than the average of their peers who did not participate. Among students with low predicted achievement, those who participated in after-school programs were BOTH more likely to have low dosage rates than high dosage rates AND more likely to have negative value added in achievement than the average of their peers who did not participate. This suggests that there are systematic differences between high-achieving high-dosage participants and low-achieving low-dosage participants. Therefore, it is impossible to generalize from this evidence to a broad claim that high after-school dosage alone will increase academic achievement among low-achieving students. However, it is possible to make the following generalization: Low-dosage participation among low-achieving student populations does not help their achievement and may detract from it. Thus, program design should certainly be directed toward increasing dosage for these students. Whether such increased dosage would in turn contribute to positive academic gains for these students remains to be seen. The evidence here suggests the possibility, but does not answer the question definitively.

In this instance, as in all the others in this report, the correlation between the results from the statistical filter and the number of days each student attended the after school program was positive but weak, $r(11210) = .02, p < .05$. The correlation is based on the data from individual

students, rather than from the grouped data shown in the previous exhibit. The correlation accounted for less than 1% of the variance in the scores. While the statistical analysis software ruled that this correlation was significant, this resulted from the large number of students included in the analyses; rather than from relationships between the variables. This weak correlation helps confirm findings that the relationship between dosage and CST scores only became meaningful when students attended for at least the minimum number of dosage days.

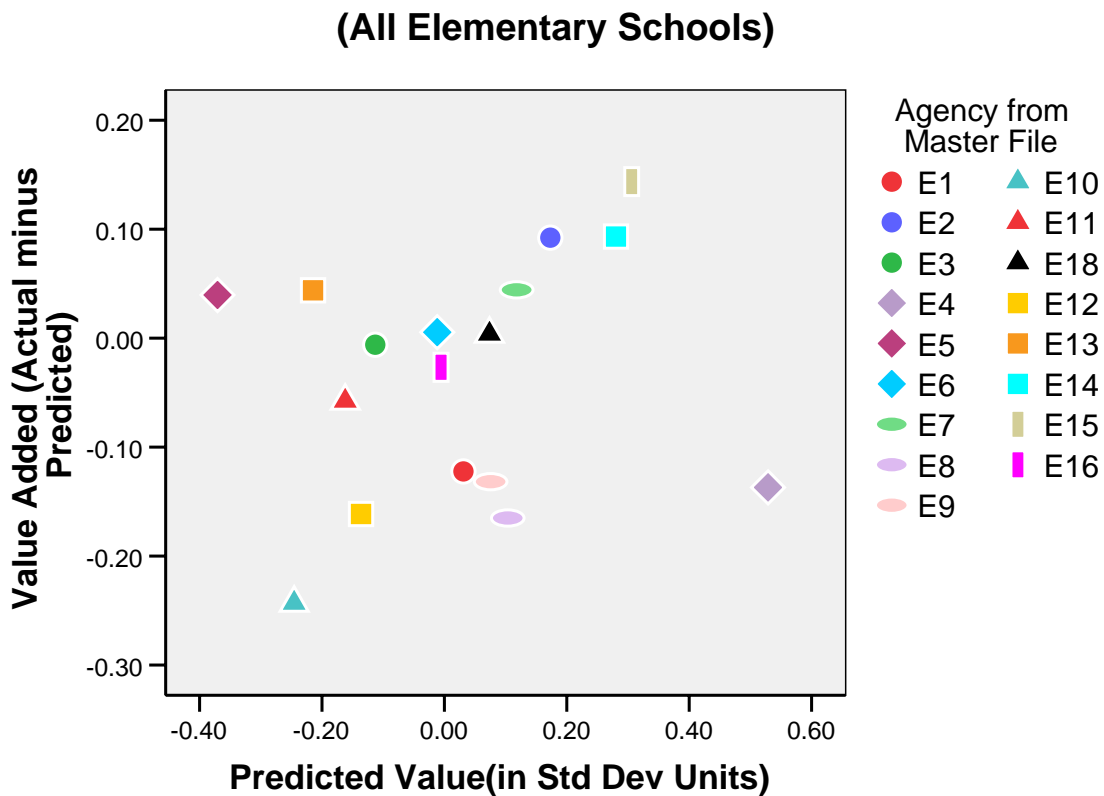
It also should be noted that in this graph, as well as in many of the similar graphs in this synopsis and in the full report, the data points generally show a *positive slope*; climbing from the lower left to the upper right of the graph as days of participation increases. While there are exceptions to this phenomena (note the relatively high performance of the 0-20 day group in this exhibit and the reverse slopes of the attendance graphs), they raise the question among sophisticated readers as to whether there are systematic differences among these groups that are in addition to their dosage levels? We've addressed this issue in the *Executive Summary* in non-technical terms by saying that it's not always possible to state whether increased dosage has led to improved performance in these groups, or whether the members of the high dosage are somewhat different in intangible ways such as parental attitudes or availability of homework help at home; and this has led to their high performance. While we have acknowledged that to some degree both theories may be true, there is still ample evidence that high dosage participation is linked to improved CST performance and that low dosage participation is often associated with declines in academic performance.

Here, getting a bit more technical, we will stipulate that in developing the value-added model, particular care was taken to create a "flat prediction surface" by assuring that the regressions the correlations between group membership and performance factors such as previous test performance, ethnicity, English proficiency, school meal program participation, etc., were close to zero before the results of after school program participation were analyzed. Detailed charts showing the degree to which this was accomplished are included in the full report.

This in turn leads us to believe that the positive slope of the model was NOT an artifact of the model, but was largely caused partly by factors such as receiving high dosages of after school care for several years that represent additional benefits of the after school high dosage experience and partly by factors such as the high dosage students being – to at least some degree – a self-selected group.

We will be able to learn more about this issue through this year's qualitative study. A major goal of that study will be to determine, through student and parent interviews and questionnaires, whether the students in the high dosage groups enjoy innate advantages over their peers in the low dosage groups, or whether most of the differences in performance resulted from the after school experience.

Exhibit 3.1.3: Elementary School English Language Arts CST Agency Value-Added Analysis, $F(16, 10885) = 2.87, p < .001$



Residuals computed separately for each grade

Exhibit 3.1.3 shows the distribution of data points for each of the agencies that served elementary school children in 2005-06. Since this is not an evaluation of the individual agencies, the names that go with the symbols have been masked. This type of graph is interpreted much like the previous type:

- Agencies whose data points are located to the left of the 0.00 point on the x-axis are serving students whose 2005-06 performance was expected to decline; whereas those located to the right of the 0.00 served students whose performance was expected to increase.
- Agencies whose data point are located above the 0.00 point on the y-axis are serving students who scored higher than the model predicted; whereas those located under the 0.00 served students who scored lower than the model predicted.

From a policy perspective, the most important finding resulting from this graph, and other ones like it, is that there is considerable “spread” among the data points representing the various agencies. As noted in the heading, these results are significant at the .001 level. The distribution of agencies on the y-axis suggests that agencies falling at roughly the same points on the x-axis vary considerably in the scores achieved by their students on the 2005-06 tests. This provides opportunities to perform in-depth studies of the agencies falling at different points on this graph to identify highly effective strategies and practices for improving dosages and test scores; as well as perform similar analyses for agency’s individual school sites.

3.2 Middle School English Language Arts CST Scores

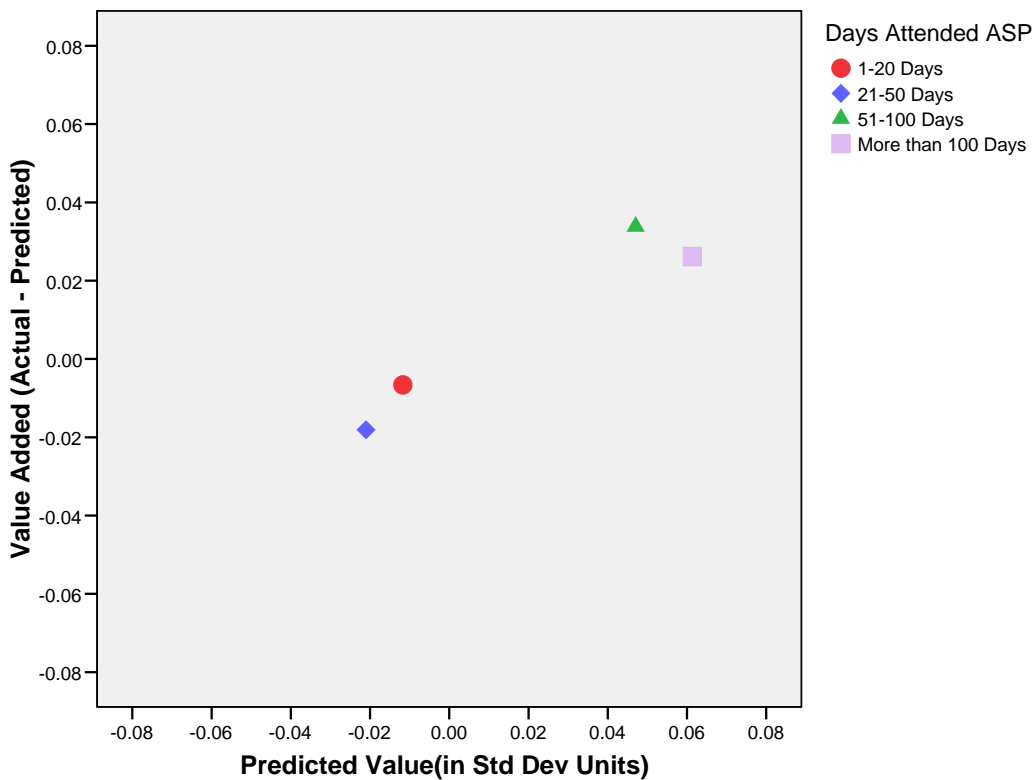
The study found statistically significant differences between the average English Language Arts CST scores for students in the high dosage group (“More than 50 days”) compared to students in the low dosage group (“1-50 days”).

Exhibit 3.2.1: Descriptive statistics of English Language Arts CST Scale Scores for Middle School Students Attending 1-50 vs. More than 50 Days

DOSAGE	<i>n</i>	<i>M</i>	<i>SD</i>
Low: 1-50 Days	22,064	310.65	48.22
High: > 50 Days	7,879	314.14	49.44

This exhibit shows that students who attended more than 50 days scored an average of 3.5 scale score points higher than students who attended less than 50 days. The difference between these two groups of students was statistically significant, $t(13579.68) = -5.55, p < .01$.

Exhibit 3.2.2: Middle School English Language Arts CST Value-Added Analysis

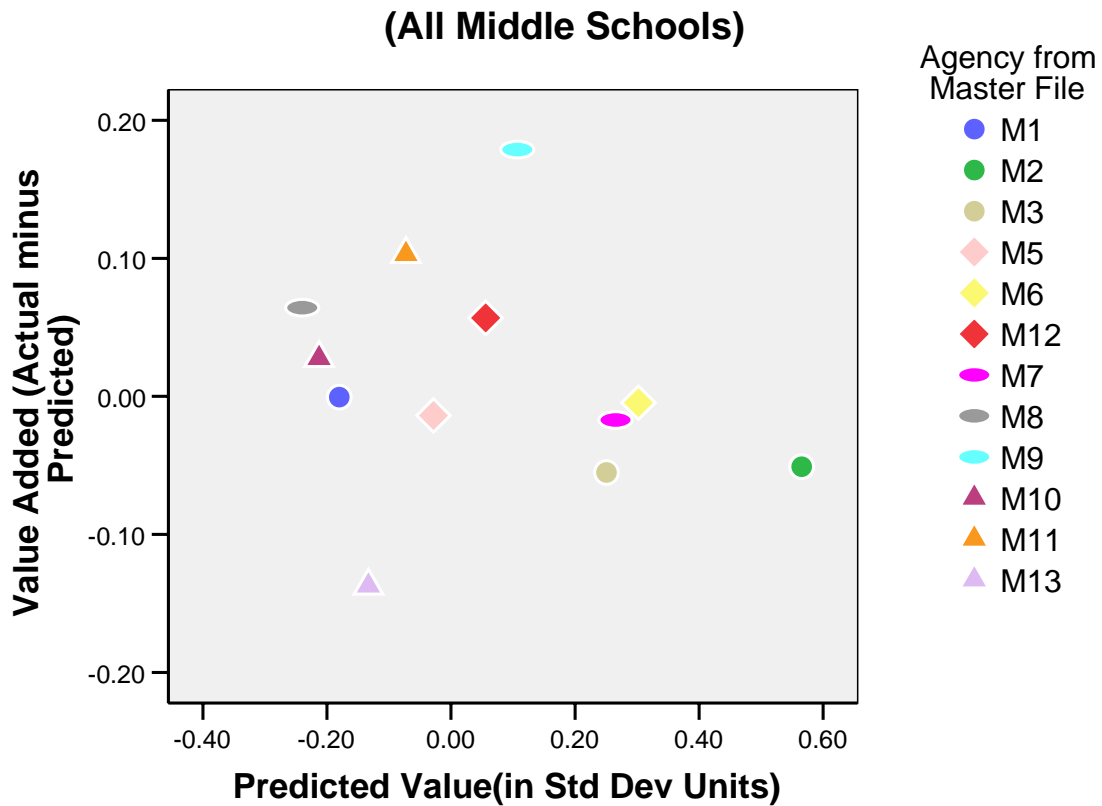


Residuals computed separately for each grade

The high dosage cohorts for middle schools are the “51-100 Days” and “More than 100 Days” groups. These were the only ones that scored above the 0.00 point on the y-axis. All the others scored at or below their predicted levels. This provides evidence those middle school students who attended BTB programs for more than 50 days last year did appreciably better than students who attended for fewer days.

The correlation between the results from the statistical filter and the number of days each student attended the after school program was positive but weak, $r(28538) = .001, p < .05$. The correlation is based on the data from individual students, rather than from the grouped data shown in the previous exhibit. This correlation is very weak, accounting for less than 1% of the variance in the scores. While the statistical analysis software ruled that this correlation was significant, this resulted from the huge number of students included in the analyses; rather than from relationships between the variables. This weak correlation helps confirm findings that the relationship between dosage and CST scores only became meaningful when students attended for at least the minimum number of dosage days.

Exhibit 3.2.3: Middle School English Language Arts CST Agency Value-Added Analysis, $F(11, 28246) = 6.72, p < .001$



Residuals computed separately for each grade

This exhibit shows the distribution of data points for each of the agencies that served middle school children in 2005-06. Since this is not an evaluation of the individual agencies, *per se*, the names that go with each symbol have been masked. The distribution of agencies on the y-axis suggests that agencies falling at roughly the same points on the x-axis vary considerably in the scores achieved by their students on the 2005-06 tests. This provides opportunities to perform in-depth studies of the agencies falling at different points on this graph to identify highly effective strategies and practices for improving dosages and test scores; as well as perform similar analyses for agency’s individual school sites.

Section Four – Mathematics

4.1 Elementary School Mathematics CST Scores

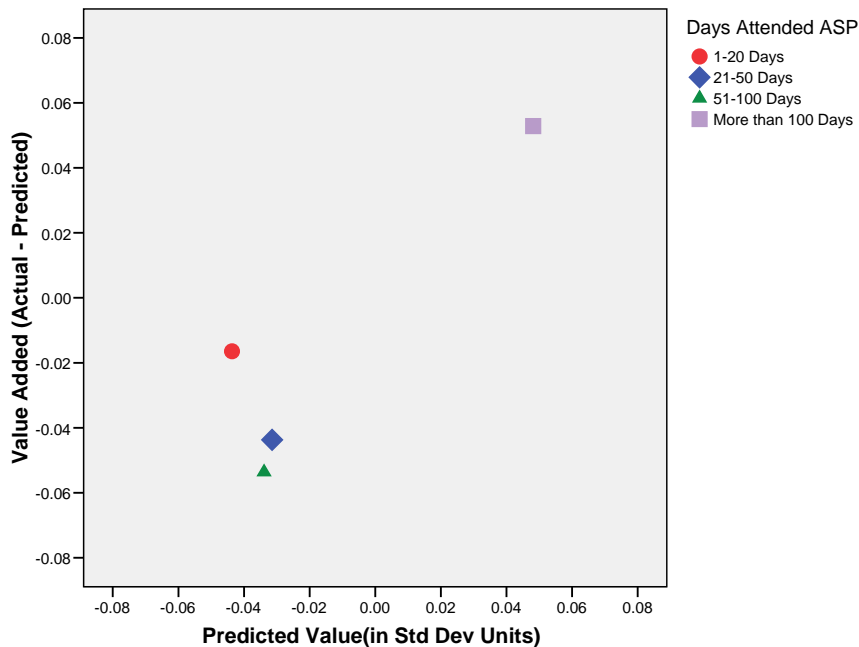
The study found statistically significant differences between the average mathematics CST scores for students in the high dosage group (“More than 100 days) compared to students in the low dosage group (“1-100 days”).

Exhibit 4.1.1: Descriptive Statistics of Mathematics CST Scale Scores for Elementary School Students Attending 1-100 vs. More than 100 Days

DOSAGE	<i>n</i>	<i>M</i>	<i>SD</i>
Low: 1-100 Days	10,083	335.53	75.11
High: > 100 Days	6,147	344.04	76.06

Students who attended more than 100 days scored an average of 8.5 scale score points higher than students who attended fewer than 100 days, $t(16228) = -6.94, p < .001$. This difference between groups places the high dosage group within 6 points of reaching the Proficient category, and wasn’t affected by the value-added analysis results presented below.

Exhibit 4.1.2: Elementary School Mathematics CST Value-Added Analysis

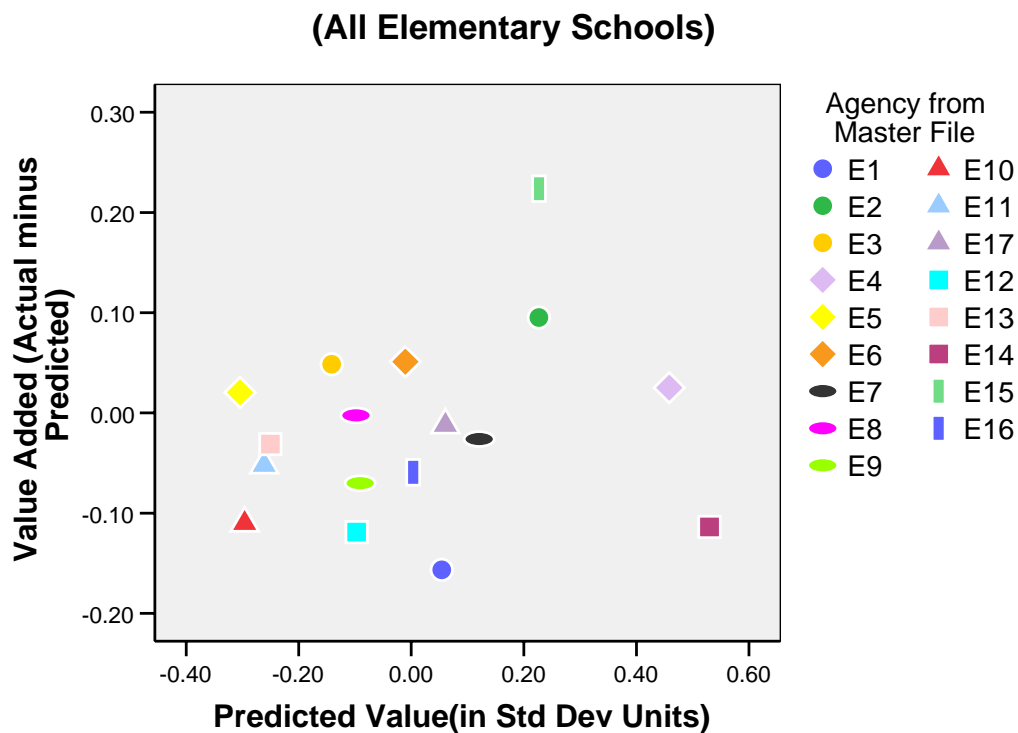


Residuals computed separately for each grade

The high dosage group for elementary school math is “More than 100 Days”. This group of students scored more than 5% above the 0.00 point on the y-axis. All the other groups scored at or below their predicted levels. This provides evidence those elementary school students who attended BTB programs for more than 100 days last year did appreciably better than students who attended for fewer days.

The correlation between the results from the statistical filter and the number of days each student attended the after school program was positive and weak, $r(11194) = .04, p < .01$. The correlation is based on the data from individual students, rather than from the grouped data shown in the previous exhibit. This correlation is very weak, accounting for less than 1% of the variance in the scores. While the statistical analysis software ruled that this correlation was significant, this resulted from the huge number of students included in the analyses; rather than from relationships between the variables. This weak correlation helps to confirm the findings that the relationship between dosage and CST scores only became meaningful when students attended for at least the minimum number of dosage days.

Exhibit 4.1.3: Elementary School Mathematics CST Agency Value-Added Analysis, $F(16, 10869) = 3.57, p < .001$



Residuals computed separately for each grade

This exhibit shows the distribution of data points for each of the agencies that served middle school children in 2005-06. Since this is not an evaluation of the individual agencies, *per se*, the names that go with each symbol have been masked. The distribution of agencies on the y-axis suggests that agencies falling at roughly the same points on the x-axis vary considerably in the scores achieved by their students on the 2005-06 tests. This provides opportunities to perform in-depth studies of the agencies falling at different points on this graph to identify highly effective strategies and practices for improving dosages and test scores; as well as perform similar analyses for agency’s individual school sites.

4.2 Middle School Mathematics CST Scores

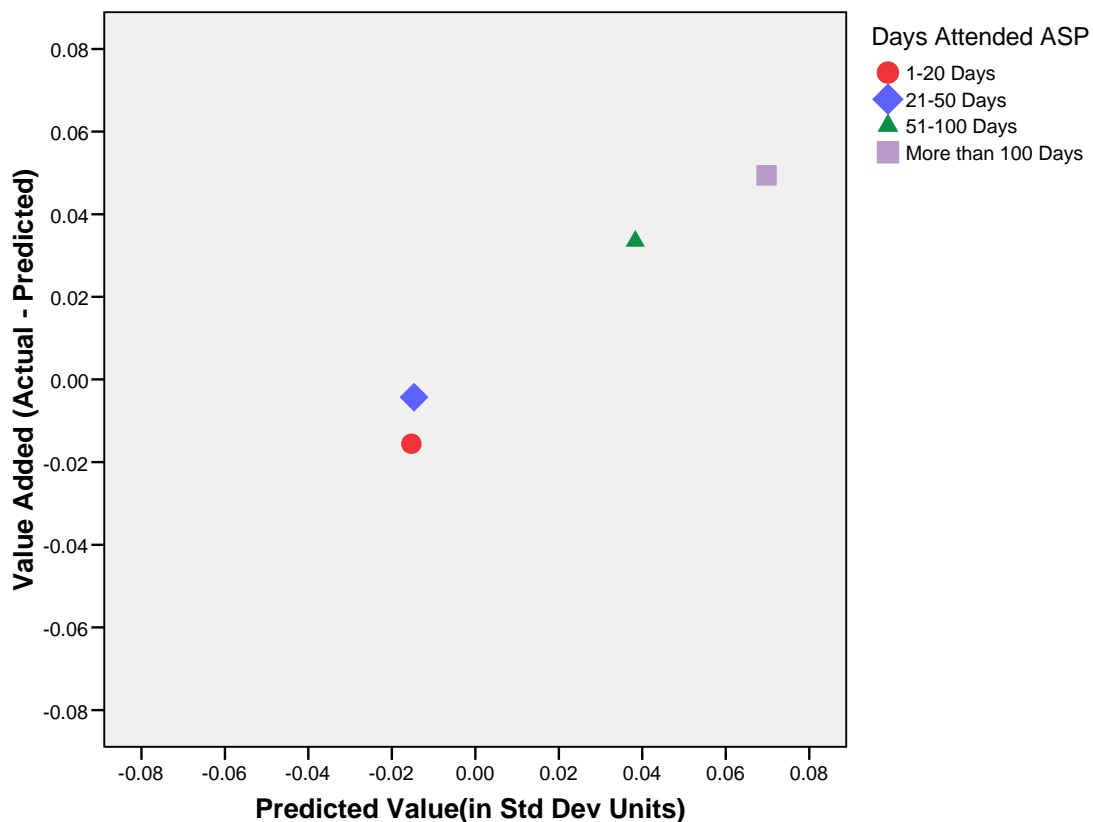
The study found statistically significant differences between the average mathematics CST scores for students in the high dosage group (“More than 100 days”) compared to students in the low dosage group (“1-100 days”).

Exhibit 4.2.1: Descriptive statistics of Mathematics CST Scale Scores for Middle School Students Attending 1-50 vs. More than 50 Days

DOSAGE	<i>n</i>	<i>M</i>	<i>SD</i>
Low: 1-50 Days	21,971	301.49	54.38
High: > 50 Days	7,847	306.27	57.27

Students who attended more than 100 days scored an average of 6 scale score points higher than students who attended fewer than 100 days ($306.27 - 301.49 = 4.78$), $p < .01$. The low dosage group is only 1.5 point away from dropping into the Below Basic category.

Exhibit 4.2.2: Middle School Mathematics CST Value-Added Analysis

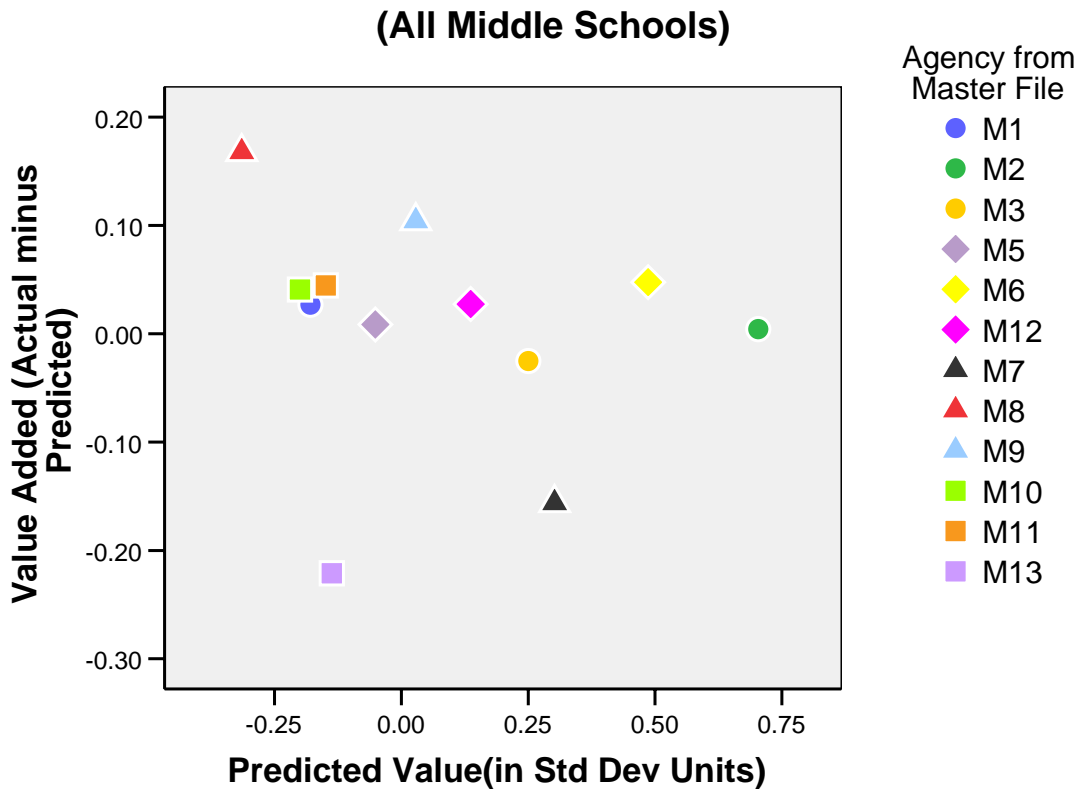


Residuals computed separately for each grade

The high dosage cohorts for middle schools are the “51-100 Days” and “More than 100 Days” groups. These were the only ones that scored above the 0.00 point on the y-axis. All the others scored at or below their predicted levels. This provides evidence those middle school students who attended BTB programs for more than 50 days last year did appreciably better on the Math CST than the model predicted.

The correlation between the results from the statistical filter and the number of days each student attended the after school program was positive and weak, $r(28229) = .02, p < .01$. The correlation is based on the data from individual students, rather than from the grouped data shown in the previous exhibit. This correlation is very weak, accounting for less than 1% of the variance in the scores. While the statistical analysis software ruled that this correlation was significant, this resulted from the huge number of students included in the analyses; rather than from relationships between the variables. This weak correlation helps to confirm the findings that the relationship between dosage and CST scores only became meaningful when students attended for at least the minimum number of dosage days.

Exhibit 4.2.3: Middle School Mathematics CST Agency Value-Added Analysis, $F(11, 28116) = 10.00, p < .001$



Residuals computed separately for each grade

This exhibit shows the distribution of data points for each of the agencies that served middle school children in 2005-06. Since this is not an evaluation of the individual agencies, *per se*, the names that go with each symbol have been masked. The distribution of agencies on the y-axis suggests that agencies falling at roughly the same points on the x-axis vary considerably in the scores achieved by their students on the 2005-06 tests. This provides opportunities to perform in-depth studies of the agencies falling at different points on this graph to identify highly effective strategies and practices for improving dosages and test scores; as well as perform similar analyses for agency's individual school sites.

Section Five – School Attendance

5.1 Elementary School Attendance

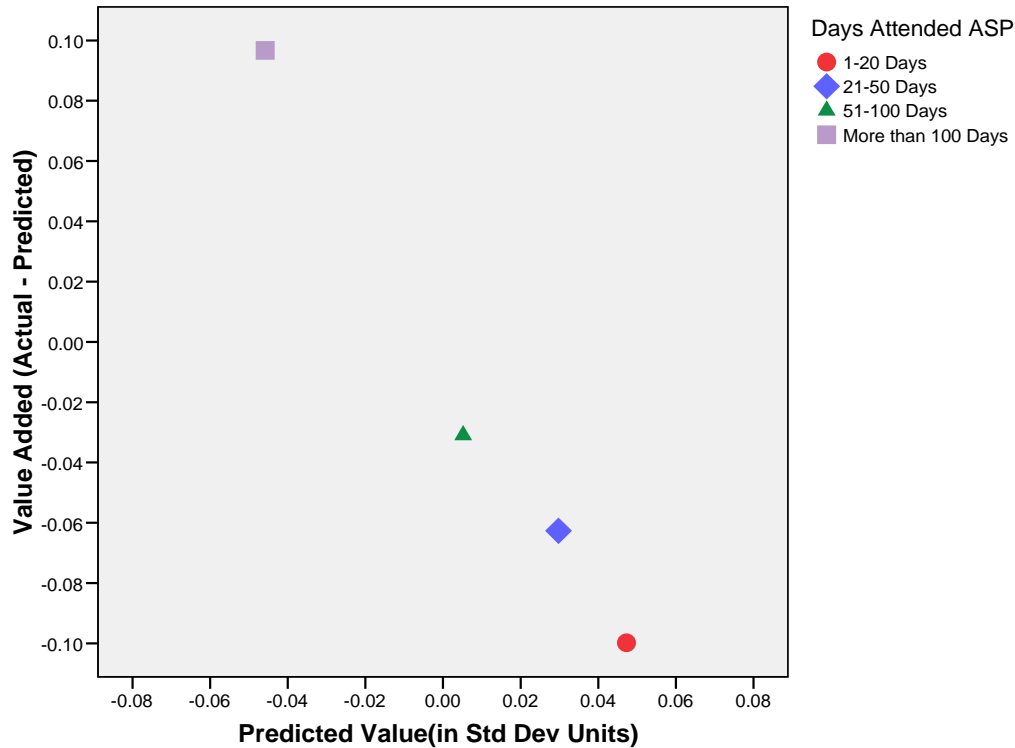
The study found statistically significant differences between the school attendance levels for students in the high dosage group (“More than 100 days”) compared to students in the low dosage group (“1-100 days”).

Exhibit 5.1.1: Descriptive Statistics of School Attendance Percentages for Elementary Students Attending 1-100 vs. More than 100 Days

DOSAGE	<i>n</i>	<i>M</i>	<i>SD</i>
Low: 1-100 Days	12,172	95.51	5.98
High: > 100 Days	7,434	96.46	4.74

Students who attended more than 100 days attended school at a rate of 96.5 percent vs. 95.51 percent for the low dosage students, $t(18331.10) = -12.28, p < .01$. On average, this is equivalent to attending about 2 additional days of school for the more than 100 days group of students (high dosage) compared to the less than 100 days group of students (low dosage).

Exhibit 5.1.2 Elementary School Attendance Value-Added Analysis

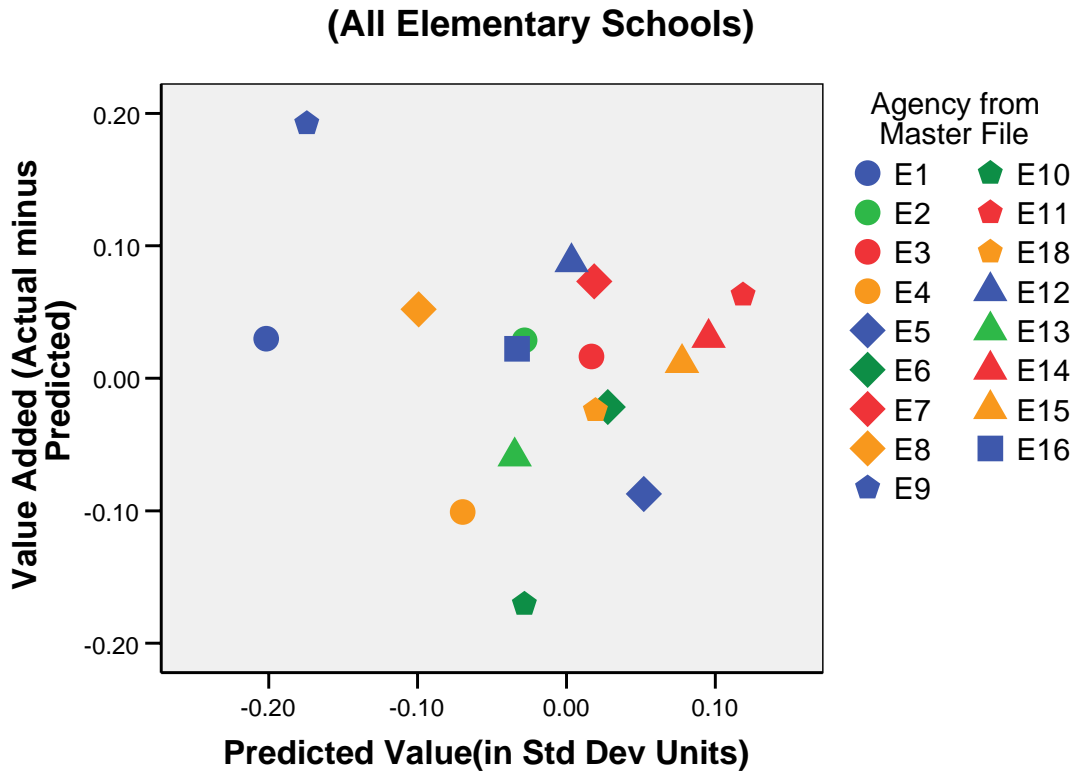


Residuals computed separately for each grade

The high dosage cohort for elementary school attendance is the “More than 100 Days” group which scored nearly 10% above the 0.00 point on the y-axis. All the other cohorts scored at or below their predicted levels. This provides evidence those elementary school students who attended BTB programs for more than 100 days last year did appreciably better than students who attended for fewer days. Further research is needed to determine why these data points are inverted compared to those of the other graphs of this type.

The correlation between the results from the statistical filter and the number of days each student attended the after school program was positive and weak, $r(11145) = .09, p < .01$. The correlation is based on the data from individual students, rather than from the grouped data shown in the previous exhibit. This correlation is very weak, accounting for less than 1% of the variance in the scores. While the statistical analysis software ruled that this correlation was significant, this resulted from the huge number of students included in the analyses; rather than from relationships between the variables. This weak correlation helps to confirm the findings that the relationship between dosage and CST scores only became meaningful when students attended for at least the minimum number of dosage days.

Exhibit 5.1.4: Elementary School Attendance Agency Value-Added Analysis, $F(16, 10792) = 1.64, p = .05$



Residuals computed separately for each grade

This exhibit shows the distribution of data points for each of the agencies that served elementary school children in 2005-06. Since this is not an evaluation of the individual agencies, *per se*, the names that go with each symbol have been masked. The distribution of agencies on the y-axis suggests that agencies falling at roughly the same points on the x-axis vary considerably in the scores achieved by their students on the 2005-06 tests. This provides opportunities to perform in-depth studies of the agencies falling at different points on this graph to identify highly effective strategies and practices for improving dosages and test scores; as well as perform similar analyses for agency's individual school sites.

5.2 Middle School Attendance

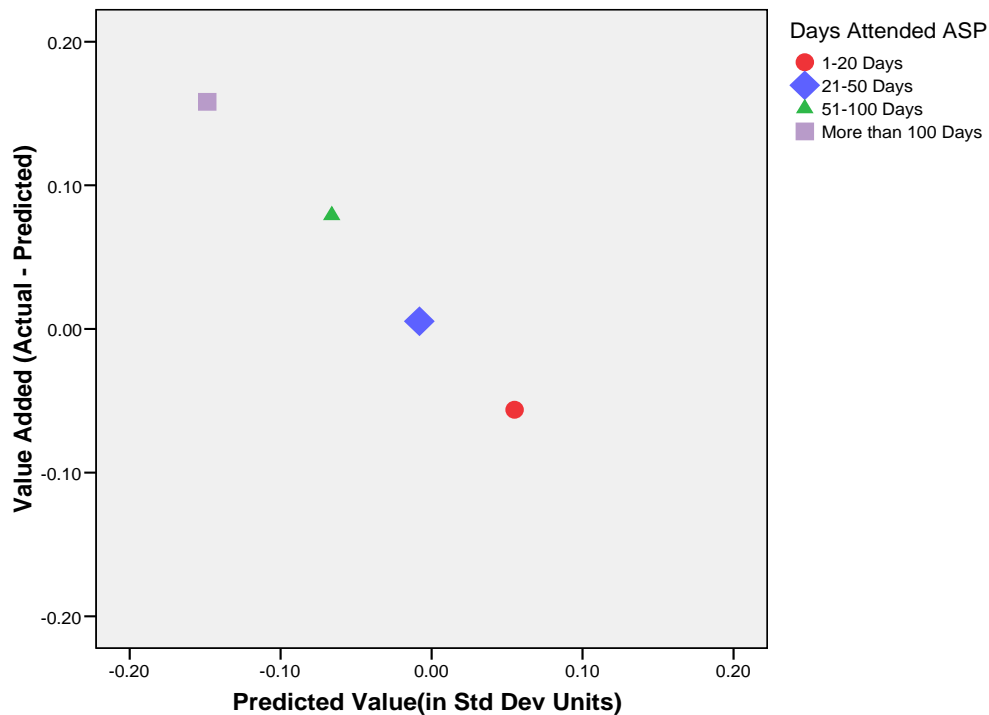
The study found educationally and statistically significant differences between the school attendance levels for students whose dosages in 2005-06 met the minimum dosage levels vs. those whose dosages were lower.

Exhibit 5.2.1: Descriptive Statistics of School Attendance Percentages for Middle School Students Attending 1-50 vs. More than 50 Days

DOSAGE	<i>n</i>	<i>M</i>	<i>SD</i>
Low: 1-50 Days	23,286	94.07	6.89
High: > 50 Days	8,209	95.24	5.80

Students who attended the after school program more than 100 days attended the regular school day at a higher rate (95%) compared to students who attended the after school program less than 100 days (94%), $t(162927.29) = -14.97, p < .01$.

Exhibit 5.2.2: Middle School Attendance Value-Added Analysis



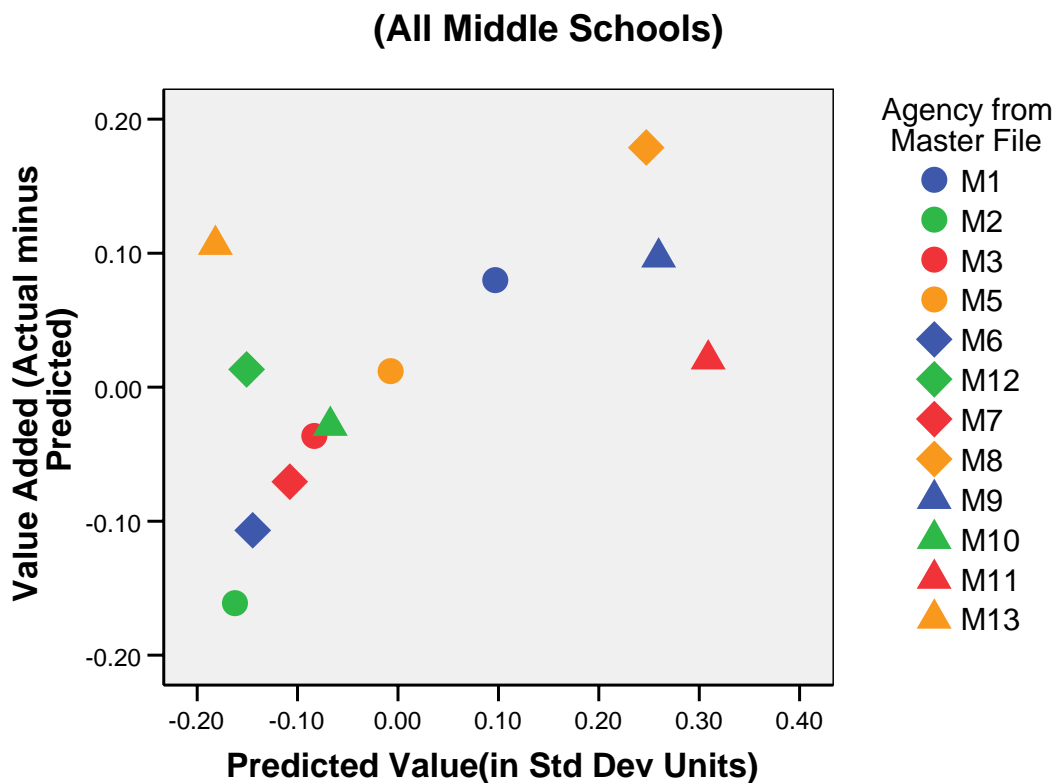
Residuals computed separately for each grade

The high dosage cohorts for middle schools are the “51-100 Days” and “More than 100 Days” groups. These were the only ones that scored above the 0.00 point on the y-axis with the “More than 100 Days” group exceeding the Predicted Value by more 15% of a standard deviation. All the other groups scored at or below their predicted levels. This provides evidence those middle

school students who attended BTB programs for more than 50 days last year did appreciably better on daily attendance than the model predicted. It will take further study to determine why these data points are inverted compared to those of the other graphs of this type.

The correlation between the results from the statistical filter and the number of days each student attended the after school program was positive and weak, $r(30012) = .07, p < .01$. The correlation is based on the data from individual students, rather than from the grouped data shown in the previous exhibit. This correlation is very weak, accounting for less than 1% of the variance in the scores. While the statistical analysis software ruled that this correlation was significant, this resulted from the huge number of students included in the analyses; rather than from relationships between the variables. This weak correlation helps to confirm the findings that the relationship between dosage and CST scores only became meaningful when students attended for at least the minimum number of dosage days.

Exhibit 5.2.3: Middle School Attendance Agency Value-Added Analysis,
 $F(11, 29901) = 8.14, p < .001$



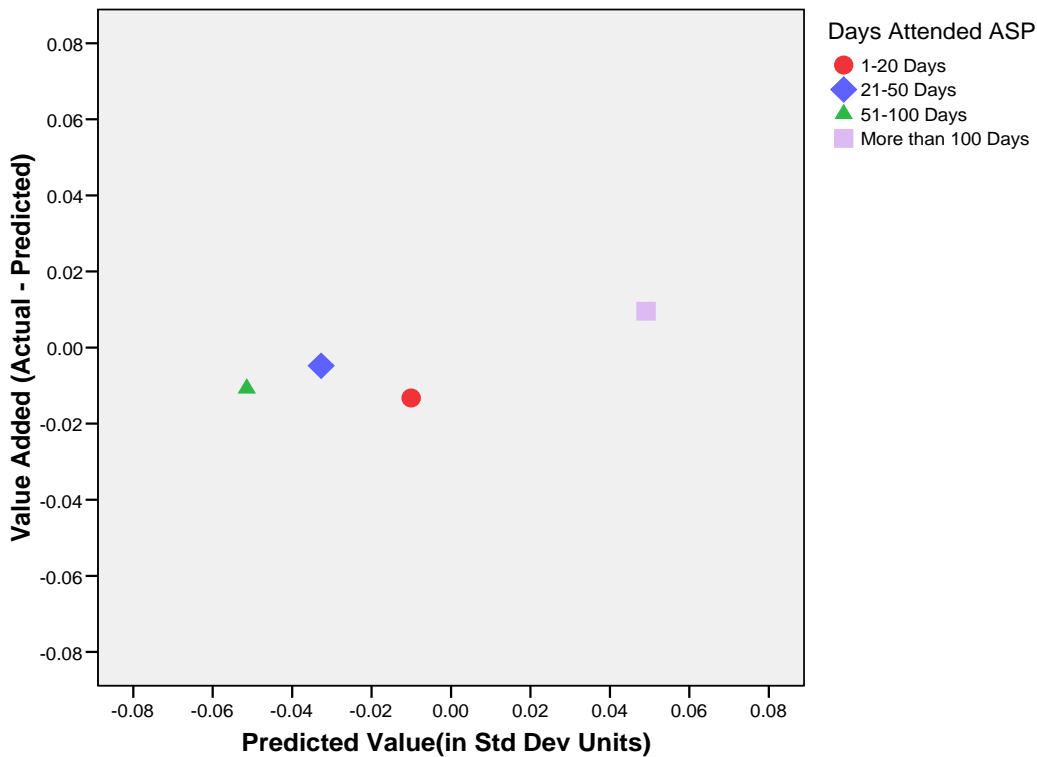
Residuals computed separately for each grade

This exhibit shows the distribution of data points for each of the agencies that served middle school children in 2005-06. Since this is not an evaluation of the individual agencies, *per se*, the names that go with each symbol have been masked. The distribution of agencies on the y-axis suggests that agencies falling at roughly the same points on the x-axis vary considerably in the scores achieved by their students on the 2005-06 tests. This provides opportunities to perform in-depth studies of the agencies falling at different points on this graph to identify highly effective strategies and practices for improving dosages and test scores; as well as perform similar analyses for agency's individual school sites.

Section Six – Teacher-Assessed Behavior

6.1 Elementary School Teacher-Assessed Behavior

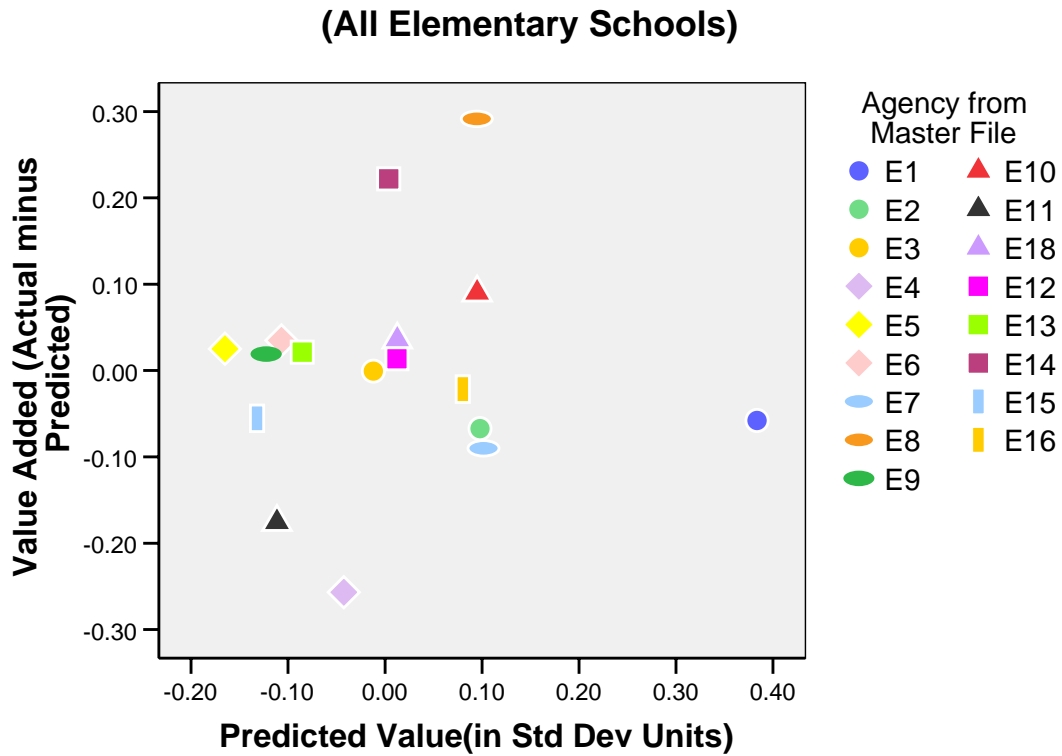
Exhibit 6.1.1: Elementary School Teacher Assessed Behavior Value-Added Analysis



Residuals computed separately for each grade

Exhibit 6.1.1 shows that none of the 4 elementary after school attendance groups scored appreciably above the Predicted Level on Teacher-Assessed Behavior. This was the only data set for which this occurred.

Exhibit 6.1.2: Elementary School Teacher-Assessed Behavior Agency Value-Added Analysis, $F(16, 10687) = 2.90, p < .001$

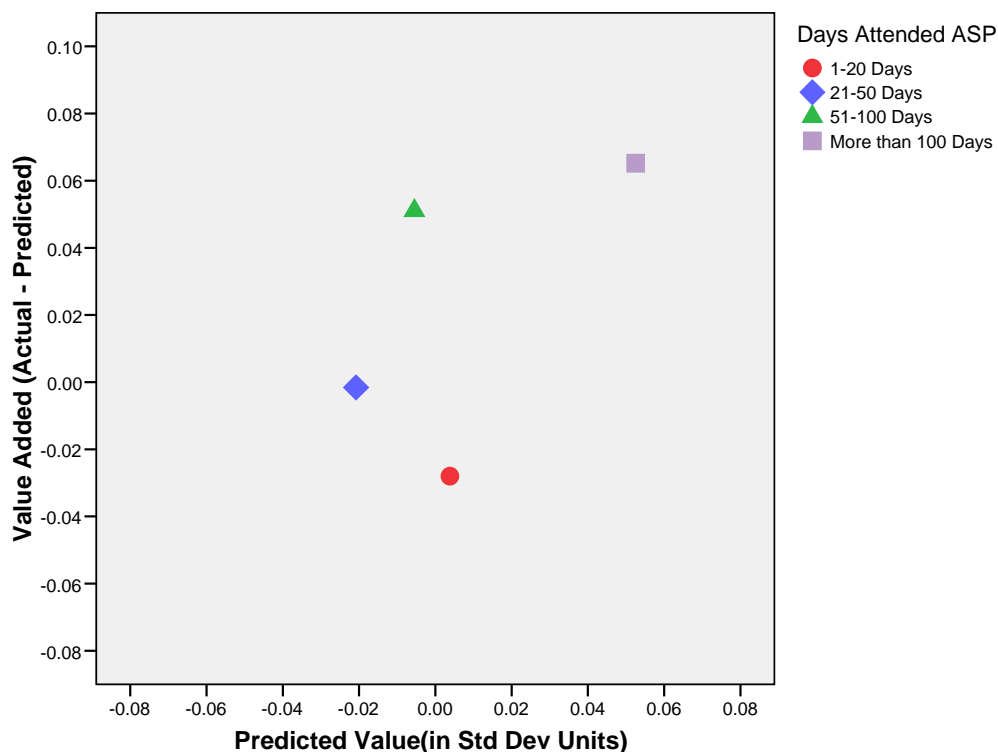


Residuals computed separately for each grade

This exhibit shows the distribution of data points for each of the agencies that served elementary school children in 2005-06. Due to the nature of the findings on Exhibits 6.1.1, we suggest that this exhibit be interpreted with some caution; even though there is a good “spread” on the vertical axis. We suggest that these differences may be due more to behavior rating practices in specific elementary schools rather than to the influence of the agencies.

6.2 Middle School Teacher-Assessed Behavior Ratings

Exhibit 6.2.1: Middle School Teacher-Assessed Behavior Value-Added Analysis

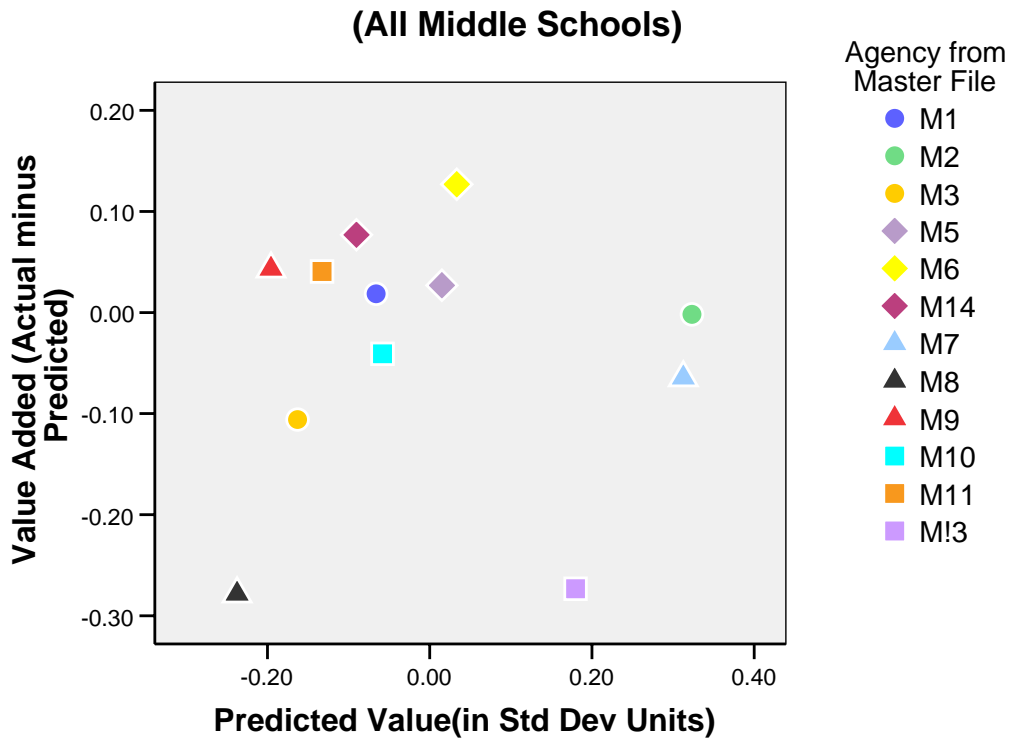


Residuals computed separately for each grade

The high dosage cohorts for middle schools are the “51-100 Days” and “More than 100 Days” groups. These were the only ones that scored above the 0.00 point on the y-axis. All the other groups scored at or below their predicted levels. This provides evidence those middle school students who attended BTB programs for more than 50 days last year did appreciably better on Teacher-Assessed Behavior than the model predicted

The correlation between the results from the statistical filter and the number of days each student attended the after school program was positive and weak, $r(20495) = .04, p < .01$. The correlation is based on the data from individual students, rather than from the grouped data shown in the previous exhibit. This correlation is very weak, accounting for less than 1% of the variance in the scores. While the statistical analysis software ruled that this correlation was significant, this resulted from the huge number of students included in the analyses; rather than from relationships between the variables. This weak correlation helps to confirm the findings that the relationship between dosage and daily attendance levels only became meaningful when students attended for at least the minimum number of dosage days.

**Exhibit 6.2.2: Middle School Teacher-Assessed Behavior Agency Value-Added Analysis,
 $F(11, 20382) = 9.93, p < .001$**



Residuals computed separately for each grade

This exhibit shows the distribution of data points for each of the agencies that served middle school children in 2005-06. Since this is not an evaluation of the individual agencies, *per se*, the names that go with each symbol have been masked. The distribution of agencies on the y-axis suggests that agencies falling at roughly the same points on the x-axis vary considerably in the scores achieved by their students on the 2005-06 tests. This provides opportunities to perform in-depth studies of the agencies falling at different points on this graph to identify highly effective strategies and practices for improving dosages and test scores; as well as perform similar analyses for agency’s individual school sites.