A Study of the Effectiveness of the

the Learning Tree™ Program

In LAUSD'S Beyond the Bell Middle Schools (2004-2005)

JULY 2005

Action Learning Systems, Inc.™

174 W. Foothill Blvd. #306 Monrovia, CA 91016 T: 626-357-8041 F: 626-357-5031
1510 J. Street, Suite 150 Sacramento, CA 95814 T: 916-447-0762 F: 916-447-0766

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

EXECUTIVE SUMMARY

The purpose of this evaluation was to study the effectiveness of Action Learning System, Inc.'s Learning Tree Program in the middle school academies of LAUSD's Beyond the Bell Branch. The integration of all evidence collected through the systematic study of program implementation, student outcomes, and teacher perceptions have led to very positive and constructive conclusions.

The key research questions posed to guide this evaluation were as follows:

1. To what extent was the Learning Tree Program fully implemented in the Beyond the Bell Extended Learning Academies? What were the effects of implementation on middle school student outcomes?

2. To what extent was there a program effect on student outcomes for students participating in a Learning Tree curricular area (i.e. Mathematics) as compared to those who did not participate?

3. Did program effects vary by student sub-group?

4. What was the relationship between attendance and student outcomes? Did attendance vary across the two curricular areas (English/LA and Mathematics)?

5. What were the perceptions of the Learning Tree Program?

To respond to these questions, the evaluation design employed both qualitative and quantitative methods. A sample of four schools was randomly selected for study. The two curricular areas of Learning Tree (Mathematics and English/LA) were studied consecutively over two sessions. A quasi-experimental design using statistically comparable groups was used to study program effect. Data collected through direct observations of instruction and teacher surveys were analyzed in response to the questions of program implementation and perceptions of effectiveness.

Presented below is a summary of the key findings:

- The level of implementation of the Learning Tree curriculum in the classroom was positively linked to student performance. The majority of observed classrooms received ratings corresponding to "attempting to faithfully replicate" the program.

- Statistical evidence of significant program effect was observed between English/Language Arts students and a "statistically comparable" group of students. Program effect was not observed between Mathematics students and their comparison counterparts. Findings did support the likelihood of reinforcement of weekday instruction, particularly at the 6th grade level.
- Students in LAUSD's calendar Track A consistently outperformed their comparison group counterparts in both the Mathematics and English/Language Arts programs.
- Student attendance was positively linked to student performance. As students increased their participation, test scores increased.
- Learning Tree teachers were very positive in their perceptions of the program and its effectiveness. The coaching component was perceived as very effective in providing both program and instructional support.

Despite limitations of design and methodology that were presented throughout the evaluation process, the findings and conclusions were both positive and informative. It was clear that program implementation and attendance were two areas that necessitate improvement. Future study using random assignment would lead to more definitive evidence of program effect. Also, focused questions of the connection between Learning Tree and traditional weekday instruction may be of interest. Overall, the evidence presented has supported the Learning Tree Program as an effective out-of-school-time intervention program for the Beyond the Bell middle school population.

INTRODUCTION

Action Learning Systems, Inc. (ALS) is a California-based educational consulting company, dedicated to a results-driven approach to education. ALS provides whole school reform and school and district-level intervention to low-performing schools throughout the state. ALS is also a state-approved provider of Supplemental Educational Services (SES), AB75, and has eight approved School Assistance Improvement Team (SAIT) leads. ALS has provided supplemental, out-of school time instructional programs to middle school students through the Los Angeles Unified School District's (LAUSD) Beyond the Bell Branch (BTB) since 2002.

The Learning Tree Intervention Program

The Learning Tree (LT) Intervention Program for Mathematics and English/Language Arts uses research-based curriculum and pedagogy based on the findings of the National Reading Panel report *Teaching Children to Read* (National Reading Panel, 2000), the National Reading Council's report *Preventing Reading Difficulties in Young Children* (Snow, Burns & Griffin, 1998), and the National Institute for Literacy's *Put Reading First* (2003). These studies provided the conceptual basis of the Learning Tree Intervention Program. Research on reading instruction and pedagogical strategies have served as the framework for the Learning Tree curriculum. The use of reading strategies, self-monitoring, text-handling and oral interaction between teachers and peers in combination with phonics and tutoring have been linked to the improvement of student achievement in reading (Camilli, Vargas, & Yurecko, 2003). The research-based instructional strategy Reciprocal Teaching is interwoven throughout the LT curriculum. Reciprocal Teaching is a strategy through which students are taught a series of reading comprehension skills and comprehension monitoring techniques (Brown, Palincsar, & Armbruster, 1984). These comprehension skills (questioning, clarifying, summarizing, and predicting) are appropriate for all content areas and forms of text and calls for participation by all students regardless of reading ability levels. Studies have shown that Reciprocal Teaching was more effective than basic skills instruction alone (Palincsar, David, & Brown, 1989) and has contributed to students' increased abilities to consider multiple perspectives, debate ideas, and reflect on more global aspects of their reading (Marks, Pressley, Coley, Craig, Gardner, DePinto, & Rose, 1993).

Findings from the Third International Mathematics and Science Study (TIMSS) contributed to the research base for the development of the mathematics program and curriculum. Recommendations from this distinguished study advocated a balanced approach to mathematics instruction; specifically, the integration of computational and procedural skills with problem-solving and conceptual understanding. This "balanced approach to mathematics" concept is the foundation of the Learning Tree Mathematics Program.

Learning Tree also integrates a professional development component and an assessment component. The professional development component consists of pre-program and follow-up training for teachers implementing the program as well as school-site coaching and monitoring to provide instructional and program-related support. The assessment component includes curriculum-aligned pre- and post-tests for each course, scanning and scoring of assessments, data analyses, and reporting.

The Learning Tree Program in LAUSD

The Learning Tree Intervention Program as offered through BTB provided two distinct California standards-based curriculums, Mathematics and English/ Language Arts, for grades six through eight.[1] Typically, there were three full sessions of Learning Tree that ran during the course of the school year. Each LT session consisted of a 30-hour program that took place on Saturday mornings across nine weeks. In 2004-05, LT was implemented in 11 LAUSD middle school sites by ALS-trained LAUSD teachers (or other credentialed staff) and supervised by school-site administrators who also worked with the BTB Branch. Depending on the individual school site, both curricular programs may have been implemented concurrently or have alternated from session to session.

Full implementation of the Learning Tree Program recommends that ALS staff provide a full day of pre-program training and subsequent follow-up trainings to include work with the instructional sequences and manipulatives not typically covered in the pre-program trainings. However, for this 2004-05 school year, LAUSD contracted for the initial pre-program trainings only. These trainings took place 1-2 weeks prior to the first day of instruction. Table 1 presents a list of items covered during the pre-program trainings.

[1] The Learning Tree English/Language Arts Program is also offered for elementary grade levels and has been implemented in the majority of schools in the Stockton Unified School District since 2003.

Table 1
Pre-Session Training Content

- Purpose, history, and research-base of the Learning Tree Program
- Roles and responsibilities of BTB and school-site administrators, , LT teachers, Students, and ALS coaches
- Week-by-week program schedule
- Administration and collection of assessments
- Results from prior sessions and intended student outcomes
- Scope and sequence of standards and how selected
- Training on research-based strategies (e.g. Reciprocal Teaching, Problem Solving)
- Step-by-step, hands-on instruction on how to facilitate the Instructional Sequences 1 and 2 and how to use the accompanying manipulatives
- Questions and answers

Learning Tree coaches were assigned by ALS to each school site. All of the Learning Tree coaches were credentialed teachers and most held concurrent classroom assignments during the traditional school week. The responsibilities of the Learning Tree coaches included answering questions related to curriculum and delivery of instruction, providing demonstration lessons or opportunities for co-teaching, observing instruction and offering feedback to teachers regarding program implementation, and maintaining positive relationships with administrators, teachers, and students. In addition, the coaches typically worked with school site administrators to organize materials, oversee testing, and collect pre- and post-test student data.

The coaching component of Learning Tree has had many positive effects across all aspects of the program. School sites that had originally opted out of this component and have since incorporated coaching into their programs have seen much improvement in their program organization, implementation, and student outcomes. Coaching support provided to classroom teachers has been consistently shown in the research literature to impact instruction and academic achievement. In fact, studies conducted in LAUSD over the past few years have revealed that direct coach-teacher interaction and ongoing engagement of coaching activities were “strong predictors of instructional changes in the classroom” (Rivera, Burley, & Sass, 2004).

The two LT curricular areas offered at the middle school level are Mathematics and English/Language Arts. Each is described in the following sections.

The Mathematics Program

The Mathematics program consisted of four different courses: Number Sense Class A, Number Sense Class B, Algebra and Functions Class A, and Algebra and Functions Class B. Each "Class" or set of sequences was based on 6th and 7th grade California standards in the Number Sense and Algebra and Functions strands. The Algebra strand of the LT program provided a foundation for the Algebra 1 content standards. At the eighth through twelfth grade levels, California mathematics standards are organized around disciplines (i.e. Algebra) and do not require students to complete all standards by a specific grade level. Each of the math sequences in the Learning Tree program incorporated computational skills, conceptual understanding, and problem solving. Research-based strategies and manipulatives were the cornerstones of each sequence.

The English/Language Arts Program

The Learning Tree English/LA program consisted of eight instructional sequences. Each sequence was based on the types of texts that were reflected in the standards and found on the California Standards Test (CST): Autobiography, Time-Sequence, Episode, Definition, Biographical Narrative, Compare-Contrast, Cause-Effect, and Problem-Solution.

Each sequence followed a similar pattern. It began with a lesson around grammar. Next there was a pre-reading activity designed to build background and connect to students' prior knowledge. They then read a piece of text using reciprocal teaching. After they finished reading, they used a graphic organizer to help them comprehend the text. Students used a graphic organizer as a pre-writing activity to aid them in writing an essay in the style they just read. For example, students in the Compare-Contrast sequence read an essay comparing two different things and then write their own comparative essay. The students would then edit each other's essays in three different ways—idea edits, sentence edits, and word edits.

PURPOSE OF EVALUATION

The purpose of this evaluation was to provide evidence of the Learning Tree Program's effectiveness within LAUSD's BTB middle school population. This evidence was gathered through the systematic study of program implementation, student outcomes as measured by

academic assessments and teacher perceptions. It is through the integration of all evidence that we present the evidence of the Learning Tree Program's effectiveness.

Evaluation Design

The strongest, most rigorous design to study program effect is the randomized control group design. However, implementation of this type of research design is not always feasible in educational settings. Although the BTB program staff initially acknowledged that this design was preferable, the logistics were simply too unrealistic for Beyond the Bell and its staff to accomplish. The main reasons were as follows:

- Since BTB serves populations of students in need of academic intervention, denying students this intervention for the sake of instituting a true "non-treatment control group" compromised the BTB Branch's mission.

- Testing randomly selected, comparable students from the larger middle school populations presented additional problems of space and time (for testing students during the school day at the home school sites) and attendance (the reality of "control group" students attending school on a Saturday to take a test).

- Funding is linked to student enrollment and attendance in the academies and implementation of this design may have negatively affected funding.

After much deliberation with the BTB program staff and technical input from the LAUSD's Program Evaluation and Research Branch (PERB), the revised research design was approved. The design utilized for the quantitative portion of this evaluation was quasi-experimental with non-equivalent groups. The structure is very similar to the pre-/post-test design without the random assignment. However, the limitations of this design included internal validity threats of selection, regression to the mean and pre-test measurement error. In order to create comparable groups for analyses, we followed a multivariate propensity score matching method as outlined by Quigley (2003) in the *Center for the Study of Evaluation* (CSE) Technical Report. This method is described in the data analysis section of this report.

Research Questions

The key questions posed to guide this evaluation were as follows:

1. To what extent was the Learning Tree Program fully implemented in the Beyond the Bell Extended Learning Academies? What were the effects of implementation on middle school student outcomes?

2. To what extent was there a program effect on student outcomes for students participating in a Learning Tree curricular area (i.e. Mathematics) as compared to those who did not participate?

3. Did program effects vary by student sub-group?

4. What was the relationship between attendance and student outcomes? Did attendance vary across the two curricular areas (English/LA and Mathematics)?

5. What were the perceptions of the Learning Tree Program?

METHODOLOGY

This study was conducted across two Learning Tree sessions during the 2004-05 school year. Each session served as the context for study of a specific curricular program. The first of the two sessions (Session A) was dedicated to the study of the Learning Tree Mathematics Program, whereas Session B was dedicated to English/Language Arts. The methods employed across both sessions were very similar. Any differences in methodology were specifically noted.

Sampling and Group Assignment

There were three sampling and/or group assignment methods that took place over the course of this study. These included the selection of school sites, the placement of students into curricular areas, and the formation of experimental and comparison groups.

Sample Selection

PERB staff randomly selected four BTB middle school sites as the study sample for this evaluation. During the 2004-05 school year, all four sites implemented both curricular programs concurrently. Each school site began and ended the academy sessions on the same calendar dates. All participating students were in grades six through eight.

The evaluation team collected data from both the School Profile information web pages (part of LAUSD.net) and the CDE's STAR data websites to determine the similarity of the study schools. According to the data for the 2003-04 school year, the demographics of the schools were parallel. These similarities served as the rationale for not differentiating students by school. Table 2 highlights the similarities across the four selected school sites.

Table 2

Demographic Similarities across Selected Schools

Calendar	3-Track
Teachers	Average 52% Permanent Teachers 2-5 Median Years Teaching at same School
Students	Average 86% Hispanic Average 54% English Learners
CST Scaled Scores	
English/LA	All three grade levels (6 – 8) average at low end of Basic level
Mathematics	Grade levels 6 and 7 average at low end of *Basic* level
Algebra I	Grade 8 averages at higher end of *Below Basic* level

Student Assignment by Curricular Area

Learning Tree students were typically referred for enrollment by school-site administrators and language arts and/or mathematics teachers, based on low performance in mathematics, language arts, or both. As requested by BTB, school staff assigned all enrolled students into one of the two curricular areas for each study session. These assignments were non-random and based on area of student need.

Since students could only participate in one curricular area per session, it was determined that students in the "target" area (i.e. Mathematics) would serve as the experimental group and students in the "non-target" area (i.e. English/LA) would serve as the comparison group. This was

reversed for the second academy session. The non-equivalence of these intact groups led to the use of analytic methods (propensity score matching) to form statistically comparable groups in the absence of random assignment.

Formation of Experimental and Comparison Groups

The evaluation team employed multivariate matched sampling methods to form experimental and comparison groups from the intact non-equivalent groups. As described by Quigley in the CSE Technical Report (2003), the use of propensity score matching is an accepted and widely used method for creating comparable groups where random assignment is not feasible. Quigley cites several works by such authors as D.B. Rubin, P.R. Rosenbaum, and R. H. Dehejhia as support for this mode of analyses.[2] Explanation of this matching process appears in the data analysis section of this report.

Data Collection

Data collection included both qualitative and quantitative methods. These methods were employed across both academy sessions.

Observations

Direct classroom observations of the implementation of the Learning Tree program were conducted over three visits during each academy session. Observation protocols were specifically designed to collect data on the implementation of Learning Tree program strategies including modeling, cooperative groups, and other program-specific strategies. Per session, only classrooms that implemented the "target" curricular area were observed. In other words, only Mathematics classrooms were observed in the first session. Two trained observers independently visited each of the classrooms three times during the session: Day 3, Day 5, and Day 8.

Student Data

Various types of student data were collected for this study. The first type consisted of student assessment data. All students were given pre-tests and post-tests specific to their curricular

[2] There were several authors cited by Quigley as support for this methodology. The inclusion of specific names and not others does not imply any judgment of importance. The full list of cited researchers and authors is available in the CSE Technical report. See Reference section in this report for full citation.

area at the start and end of each Learning Tree program session. The ALS Mathematics pre- and post-tests consist of 25 multiple-choice items that have been aligned to the specific program content and standards. The ALS English/Language Arts pre- and post-tests consist of 35 multiple-choice items. These assessments each take about 40 – 50 minutes to administer.[3] In addition, students were given the *Multilevel Academic Survey Test – Short Form (MAST)* in Math or Reading. This norm-referenced test has been described as appropriate for use with students in grades 3 through 12 (Howell, Zucker, Morehead, 1985). Each MAST assessment added approximately 10-15 minutes to the test administration.

The testing schedule for both academy sessions is presented in Table 3. As may be seen, the curricular area of focus for Session A was Mathematics. Students in both the experimental and comparison groups were tested in Math. Session B followed the same procedure with the English/Language Arts tests.

Table 3

Learning Tree Program and Testing Schedule for Both Academy Sessions

Session A (Nov 2004 – Feb 2005)				Session B (March 2005 – May 2005)			
	Pre-test MATH	Teach MATH	Post-test MATH		Pre-test Eng/LA	Teach Eng/LA	Post-test Eng/LA
Experimental Group (Math students)	Yes	Yes	Yes	Experimental Group (English/LA students)	Yes	Yes	Yes
Comparison Group (students participating in English/LA program)	Yes	No	Yes	Comparison Group (students participating in Math program)	Yes	No	Yes

BTB and school-site staff supplied attendance rosters at the end of each session. The rosters contained attendance hours for each participating LT student. In addition, databases were reviewed to determine which students had previously participated in Learning Tree (during the session just prior to Session A) and how many hours they accumulated in the program. This served to "disqualify" students based on recent knowledge of the curriculum and exposure to the tests.

[3] Technical information for the ALS-developed assessments is presented in Appendix A.

LAUSD also provided student data from the Student Information System database. These data consisted of such student level variables as 2004 CST scaled scores in Math and English/Language Arts, English Language proficiency level (if applicable), ethnicity, and school track.

Teacher Survey

The evaluation team designed a teacher survey to capture information regarding the teachers' perceptions of and experiences with the Learning Tree program. The purpose of the survey was to elicit information from LT teachers regarding their training, program support, perceptions of effectiveness, and prior experience with the Learning Tree program. Learning Tree coaches distributed the surveys over the last two weeks of each session. The survey contained both scaled response items and open-ended questions that took approximately 20 minutes to complete. Surveys were distributed with self-stick envelopes to protect the confidentiality of responses. Surveys were either returned by mail or were picked up by the ALS coaches.

Data Analysis

As with the data collection methods described above, the modes of analyses were both qualitative and quantitative.

Qualitative

Classroom observation data were analyzed and scored by trained coders on a 4-point implementation rubric (ranging from 4 = "exemplary" to 1 = "not faithful replication"). A copy of the full rubric is available in Appendix B. The scores from the three classroom observations were combined to produce a holistic score representative of classroom implementation. Qualitative analyses of field notes were used to support rubric-based scores. Open-ended survey responses were analyzed for emergent patterns and themes.

Quantitative

As previously described, this evaluation employed a quasi-experimental design with experimental and control groups to study program effect. The groups of students designated as the experimental and comparison groups were intact groups and were neither randomly assigned nor

randomly selected. The technique of propensity score matching was used to create statistically comparable groups based on specific variables of interest. Propensity scores are probabilities that are calculated through logistic regression analyses. All students, regardless of their actual group membership, become part of the analyses. Simply put, a student's propensity score is the probability that he or she would be participating in a Learning Tree program based on relevant criteria, such as CST scores, grade level, track, and their pre-test scores. Once propensity scores were calculated, students in the intact comparison group were matched score by score to students in the intact experimental group. Students that did not match were left out of the group formations. Statistical tests for equality of distributions (Kolmogorov-Smirnov tests) and equality of means (*t* tests) were then conducted on the relevant criteria. The results of these tests would establish that the two groups did not differ statistically and would all be likely to participate in the Learning Tree program. This process was conducted for all four Mathematics courses and the English/Language Arts program.

The most appropriate analytic method for the quasi-experimental design is the Analysis of Covariance model (ANCOVA) where the program effect is indicated by the post-test scores after holding the pre-test scores constant. Analyses using the ANCOVA model as well as appropriate descriptive statistics were performed on the "statistically comparable" groups. Student variables, levels of program implementation, and attendance were included in these analyses. Statistical information and summary tables of this process are presented in their respective results sections.

Data Limitations

Student assessment data was the primary outcome measure used for this study. For the majority of the analyses, student data was used only if there were matching pre- and post-test scores. This resulted in a huge drop in the number of eligible students. Additional analyses may have required the exclusion of certain students due to other missing data. Students that were removed as a result of the propensity score matching process are discussed in subsequent sections. Tables 4 and 5 display the numbers of students that were lost because of missing pre- or post-test scores and the resulting numbers of matched-score students.

Table 4
Summary of Students Tested for Session A (Mathematics)

	Total Number of students tested (Pre and/or Post)	Total Number of Pre/Post Matches*	% Match
Experimental Group			
Number Sense Class A	419	251	60%
Number Sense Class B	111	41	37%
Algebra/Functions Class A	226	153	68%
Algebra/Functions Class B	71	31	44%
Total	827	476	58%
Comparison Group			
Number Sense Class A	349	224	64%
Number Sense Class B	34	28	82%
Algebra/Functions Class A	91	53	58%
Algebra/Functions Class B	40	29	73%
Total	514	334	65%
GRAND TOTAL	1341	810	60%

*Learning Tree Assessments

As shown in Table 4, only 60% of the students had matching pre/post Mathematics scores. Absence and/or late enrollment were the primary reasons for this low percentage. In some cases, students were moved from one course to another.[4] In at least two classrooms, students were administered the wrong post-test. Table 5 displays similar numbers.

Table 5
Summary of Students Tested for Session B (English/Language Arts)

	Total Number of students tested (Pre and/or Post)	Total Number of Pre/Post Matches*	% Match
Experimental Group	548	284	52%
Comparison Group	628	256	41%
TOTAL	1176	540	46%

*Learning Tree Assessments

[4] Students that were moved from Math to English, or vice-versa, in the middle of a session were removed from the analyses. School administrators had total control of these movements.

RESULTS
THE MATHEMATICS PROGRAM

This section presents all findings specific to the Mathematics program and its four component courses of Number Sense (Class A and B) and Algebra and Functions (Class A and B). Results are presented in the order that best corresponds to the evaluation questions.

Program Implementation

In our sample for Session A, 45 classrooms were designated for Mathematics instruction. Coding and rubric scoring of the observation field notes yielded the implementation ratings presented in Table 6. As may be seen in the table below, a large number of classrooms (49%) were observed as "attempting to faithfully replicate" the curriculum. This level indicated an observed effort, though not always effective, at implementing the strategies, techniques, and management skills necessary for faithful program replication.[5] Forty-three percent of the observed classrooms demonstrated faithful and exemplary levels of implementation. The four classrooms that received a score of "1" were observed engaging in activities that were not part of the Learning Tree curriculum. During at least two observations, observers were told that Learning Tree instruction had ended for the day – with at least 30 to 45 minutes left in the period.

Table 6

Summary of Levels of Program Implementation as Observed in Mathematics Classrooms

Implementation Level	Description	Number of Math Classrooms	%
4	Exemplary	12	27%
3	Faithful Replication	7	16%
2	Attempting to Faithfully Replicate	22	49%
1	Not Faithful Replication	4	9%
		45	100%

[5] See Appendix B for the full rubric used for scoring implementation levels.

Successful implementation of the Learning Tree Program is an essential requirement of program success. Using post-test scores on the LT tests as the outcome measure of student performance, an ANCOVA analyses was run with pre-test scores as the covariate and the classroom implementation level as a main effect. The results showed that student performance on the LT post-tests varied significantly with the level of program implementation, F (3, 475) = 7.132, p=.000. Although the statistical magnitude of this effect was small (η^2 = .04), the educational significance is much more relevant and supports the fundamental requirement for program success.

Student Outcomes

As described earlier, the investigation of program effects was quasi-experimental, comparing the test scores of two intact groups of students (experimental vs. comparison) for each of the four Mathematics courses. The following section describes the propensity score matching process used to form the "comparable" groups for analyses.

Propensity Score Matching

Prior to performing the techniques for propensity score matching, we removed all students that had participated in the same Mathematics course just prior to the current session. The rationale was that these students had very recent exposure to both the curriculum and the assessments. For each of the four courses, we conducted logistic regressions to calculate the propensity score for each student.[6] Predictor variables included 2004 CST Math scaled scores, grade levels, the students' track, LT pre-test scores, and MAST pre-test NCE scores. The propensity scores for students already in the intact comparison group were matched to the nearest number to the scores of those students already in the experimental group. As a result of this procedure, we found that the Number Sense Class B and the Algebra and Functions Class B group sizes were too small for analyses. Table 7 displays the resulting group sizes for all four courses.

[6] Refer to page 11 for additional explanation of the propensity score matching process.

Table 7

Session A Group Sizes After Propensity Score Matching

	Total Prior to Matching	Unmatchable/Missing Data	Resulting Experimental/ Comparison Groups
Number Sense Class A	475	203	136/136
Algebra/Functions Class A	206	122	42/42
The following groups were too small for comparative analyses:			
Number Sense Class B	69	35	17/17
Algebra/Functions Class B	60	34	13/13

The newly formed groups for the Number Sense Class A and Algebra and Functions Class A were tested for equality of distributions and equality of means. Table 8 presents the results for both of these tests on the two math courses. The large *p* values indicated that on all of these predictor variables, the resulting groups were not statistically different from each other. In other words, based solely on the criteria provided, the groups were statistically comparable.

Table 8

Test of the Equality of Distributions and Means for Experimental Group Students and Comparison Group Matches

Number Sense Class A				
	Equality of distributions (Kolmogorov-Smirnov Test)		Equality of means (2-sided)	
Variable	K-S *Z*	*p*	*t* statistic	*p*
Propensity	0.315	1.000	0.027	0.979
LT Pre-test	0.591	0.876	0.073	0.942
MAST Pre-test	0.815	0.520	-1.026	0.306
2004 CST Math	0.451	0.987	-0.534	0.594
Grade	0.261	1.000	0.244	0.807
Track	0.286	1.000	0.406	0.685

Algebra and Functions Class A				
	Equality of distribution (Kolmogorov-Smirnov Test)		Equality of means (2-sided)	
Variable	K-S *Z*	*p*	*t* statistic	*p*
Propensity	0.327	1.000	-0.097	0.923
LT Pre-test	0.873	0.431	-0.607	0.546
MAST Pre-test	0.655	0.785	0.292	0.771
2004 CST Math	0.546	0.927	-0.347	0.730
Grade	0.109	1.000	-0.349	0.728
Track	0.436	0.991	-0.537	0.592

It is important to note that these newly formed groups were only comparable based on the variables listed. There were many other unmeasured sources of variation that were not accounted for in this analyses which may have influenced their comparability for analyses.

Analyses of Program Effect

Using the LT post-test scores as the primary measure of student outcomes, ANCOVA procedures were performed. Although the average pre/post-test difference in scores for the experimental group were greater than those from the comparison group in Number Sense Class A, this difference was not statistically significant (see Table 9). As may be seen, the treatment effect (*B = 2.11)*, had not achieved a .05 level of significance (*p* = .242). Analyses with the MAST NCE data produced similar results.

Table 9

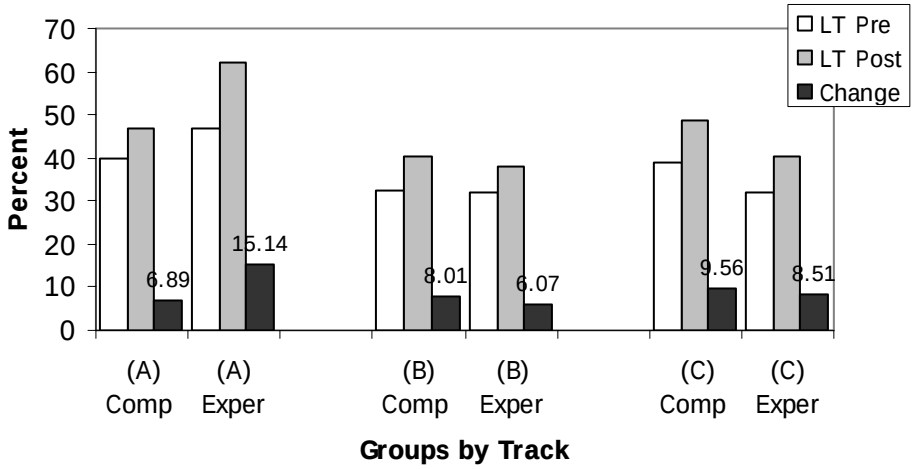
Analysis of Covariance (ANCOVA) Model Summary for Number Sense Class A

	n	Pre-test *M(SD)*	Post-test *M(SD)*	
Experimental	136	37.57 (19.29)	48.06 (20.73)	
Comparison	136	37.75 (21.58)	46.07 (20.32)	
	B	*SE*	*p*	η^2
Intercept	22.00	2.09	.000	
Pre-test	0.69	0.04	.000	
Group[1]	2.11	1.80	.242	.005

[1]This variable was dummy-coded as 1 = Experimental group, 0 = Comparison Group

When disaggregating students by school track and grade level, the ANCOVA procedures revealed a highly significant program effect among Track A students, B (1, 98) = 10.61, p = .000. In other words, Track A students in Number Sense Class A demonstrated a significantly greater gain in test scores over their counterparts (see Figure 1). Disaggregating by grade level revealed a smaller program effect between the groups at the sixth grade level, B (1, 177) = 3.25, p = .14; however, statistical significance fell just below the .10 level. Interestingly, grade level appeared to make a difference in all test takers, regardless of group designation. In other words, sixth graders overall did significantly better on the Number Sense Class A tests than did students of other grades. A possible reason for this is the fact that the Number Sense Class A curriculum is based largely on 6th grade standards and serves to reinforce regular weekday instruction. Since 7th and 8th grade students typically receive curriculum geared to their respective grade-level standards, it is possible that this intervention instruction of Number Sense curriculum reduced the observable program effect. The opposite effect was observed for the MAST test results. Although 6th graders scored higher on their MAST pre-tests than the other grades, their scores did not improve over the course of the program as did 7th and 8th graders.

Figure 1
Program Effect by Track (Mathematics)



Analyses of the Algebra and Functions Class A test data revealed no statistically significant difference between students scores (either LT tests or MAST) in the experimental and

comparison groups. The small group sizes prevented more disaggregated analyses. However, similar patterns of higher scores (as described for the Number Sense students) were observed among students on Track A.

Student Attendance

Student attendance in the LT program was recorded by staff at each site and submitted to our evaluation team. As described earlier, student attendance varied greatly over the course of the 9-week program. Students entered the program and exited at different points. In order to study the effects of attendance on student performance in the Mathematics program, we examined the relationship between hours attended and LT test scores of all students who participated in a mathematics course (n = 991).

The average number of attendance hours was 20.81 hours (SD = 7.51). Students with matching pre- and post-tests (n = 476) had a higher average attendance with less variability (M = 24.82, SD = 4.88). Analyses revealed a significant positive relationship between student attendance hours and LT post-test scores (r = .14, p = .001). A similar positive relationship was seen with MAST post-test scores (r = .10, p = .03). In other words, the more time students spent in the program the better their mathematics performance as evidenced by their post-test scores. These findings align to the multitude of research linking increased learning time to achievement.

Mathematics Teacher Survey Results

Our evaluation team distributed surveys to all LT Mathematics teachers (n = 45) in Session A. The response rate was 47%. Because of this small number, these survey responses were combined with those of the LT Mathematics teachers from the seven non-study schools. Although this only increased the overall response rate to 49%, it increased the actual number of completed surveys to fifty. Preliminary review of the surveys did not reveal any differences among responses from teachers at study vs. non-study schools. About 90% of these LT Mathematics teachers concurrently held traditional classroom assignments at the middle school level. All but one teacher attended the Learning Tree training prior to the start of Session A. The average number of years of overall teaching experience was 8.5.

Teachers were asked to rate the sequences within their respective courses. For both the Number Sense and Algebra and Functions courses, the clear majority of responses described the

sequences as “strong” or “adequate” (85%). The most often cited reasons for these ratings were the clarity of the directions and the “easy to follow” curricula. Those teachers who rated sequences as “weak” described some activities as “redundant” or “too few.”

The teachers were asked to respond to a series of scaled-response questions regarding implementation and training/support. Table 10 presents the average responses to questions regarding implementation frequency of various program components. These questions were aligned to the classroom observation rubric. The possible responses were as follows: 4 = “all of the time,” 3 = “more than half of the time,” 2 = “half of the time,” 1 = “less than half of the time,” and 0 = “not at all.”

Table 10

Average Responses to Program Implementation

Survey Items	*M (SD)*
Over the 9 weeks of the Learning Tree Program, how often did you …	
place students in cooperative learning groups or pairs?	3.05 (0.95)
model the use of manipulatives?	2.90 (1.07)
provide manipulatives for students to use?	2.78 (1.09)
use *direct instruction* as part of your teaching?	2.96 (1.08)
use the activities as demonstrated?	3.16 (0.91)
connect the instructional activities to the standards?	2.83 (1.02)
model the activities for the students?	3.50 (0.87)
use different activities/strategies/materials to supplement the lessons?	2.52 (1.15)

As may be seen in the table above, teacher responses indicated that placing students in cooperative learning groups, using the activities as demonstrated, modeling the use of manipulatives, and using direct instruction strategies as part of the teaching occurred more than half of the time. In particular, teachers reported the modeling of activities as the most frequently implemented component. This was consistent with teachers’ descriptions of two activities that they found to be very effective with their students: activities involving group or pair work and any activity using manipulatives. Components which were not implemented as frequently included the explicit connection of program activities to the standards and the accessibility of manipulatives for student use. Teachers reported that “half the time” they supplemented the curriculum with

additional materials. This was particularly true of Algebra teachers who reportedly supplemented the program with additional "practice problems," and "problems that connect back to the classroom."

The survey also included items regarding training and coach support. Table 11 presents the items and average responses. For this series, the 5-point response scale ranged from" Strongly Agree" (5) to "Strongly Disagree" (1).

Table 11

Average Responses to Items of Training/Coach Support (Session A)

Survey Items	*M (SD)*
When asked, the coach effectively demonstrated lessons/strategies in my classroom.	3.85 (0.82)
I was able to successfully replicate strategies demonstrated by the coach and/or the trainer.	3.91 (0.76)
The coach was able to answer my questions about teaching the program.	4.02 (0.79)
The coach provided me with useful information and feedback.	3.91 (0.72)
I have received enough training to implement this program.	3.91 (0.88)
There was not enough time to fully implement the program as presented in the training.	3.53 (1.00)
In general, I would have liked more examples and explanations of the strategies.	3.64 (1.09)

Overall, teachers agreed the support provided by their ALS coach positively impacted their implementation. This coaching support included effective classroom demonstrations of lessons and strategies, availability, and ability to provide useful answers, program information, and constructive feedback on instructional delivery of curriculum. Teachers' positive responses of their ability to successfully replicate and implement strategies provided further evidence of the positive impact of the coaching component. The last three items specifically addressed the training. Teachers agreed that they received adequate training for overall program implementation. However, teachers also requested more examples and explanations of strategies. The 1-day training that teachers received prior to each session only covered the process of the program and explanations of specific strategies for the first two sequences only. This was very likely the reason for the response differences in these two survey items. Finally, teachers (specifically Number

Sense teachers) agreed that there was not enough time to implement the program as presented in the training.

THE ENGLISH/LANGUAGE ARTS PROGRAM

Study of the Learning Tree English/Language Arts program was conducted during Session B. All findings from this session are presented below in the same order as the mathematics section.

Program Implementation

At the start of the session, 50 classrooms were designated for English/Language Arts instruction. By the end of the session, 42 classrooms remained. Small class sizes were the primary reason for these closures. Coding and rubric scoring of the observation field notes yielded the implementation ratings in Table 12.

Table 12

Summary of Level of Program Implementation as Observed in English/Language Arts Classrooms

Implementation Level	Description	Number of Eng/LA Classrooms	%
4	Exemplary	0	0%
3	Faithful Replication	7	17%
2	Attempting to Faithfully Replicate	19	45%
1	Not Faithful Replication	16	38%
		42	100%

As shown in the table above, many of the classrooms (45%) were observed as "attempting to faithfully replicate" the program's curriculum.[7] This level of implementation was also observed in 49% of the mathematics classrooms. Seventeen percent of the classrooms were rated as faithfully replicating the program. None of the 42 observed classrooms exhibited exemplary replication. Sixteen classrooms (38%) received a score of "1," or "not faithful replication." In many cases, Learning Tree instruction was not observed at all. In addition, little evidence of

[7] See Appendix B for the full rubric used for scoring implementation levels.

cooperative groups, direct instruction, and student engagement were observed in these classrooms. Cross-referencing survey data also revealed that a majority of these English/LA teachers were new to the Learning Tree Program this session. This may have influenced the resulting implementation levels.

Successful implementation of the Learning Tree Program is an essential requirement of program success. Using post-test scores on the LT tests as the outcome measure of student performance, an ANCOVA analyses was run with pre-test scores as the covariate and the classroom implementation level as a main effect. The results showed that student performance on the LT post-tests varied significantly with the level of program implementation, F (2, 273) = 3.087, p=.05. Although the statistical magnitude of this effect is small (η^2 = .02), the educational significance is much more relevant and supports the fundamental requirement for program success.

Student Outcomes

Similar to the Mathematics program in Session A, we were challenged with intact groups of students participating in Session B. We again used propensity score matching techniques to form the most "comparable" groups using the criteria and data that were available to us. The following section presents the results of these methods.[8]

Propensity Score Matching

Prior to matching, all students who participated in English/Language Arts in Session A were removed from the pool. Exposure to the tests and the curriculum were the primary reasons. Logistic regression was performed to calculate the probability, or likelihood, that a student would participate in this particular Learning Tree program based on a set of criteria (predictor variables). Propensity scores of the comparison group scores were matched to the nearest score of the experimental group. Students with scores that did not match to the nearest .05 were removed. Table 13 displays the resulting group sizes.

[8] Refer to page 11 for more detailed description of propensity score matching.

Table 13

Session B Group Sizes after Propensity Score Matching

	Total Prior to Matching	Unmatchable/Missing Data	Resulting Experimental/ Comparison Groups
English/Language Arts	540*	278	131/131

*Of this group, 117 were removed because they attended this class in prior session

The newly formed groups were tested for equality of distributions and means. The large *p* values indicated that on all of the predictor variables put into the regression, the resulting groups were not statistically different from one another. In other words, based solely on the criteria provided, the groups were statistically comparable. Table 14 presents the results of these statistical tests of equality.

Table 14

Test of the Equality of Distributions and Means for Experimental Group Students and Comparison Group Matches

	Equality of distribution (Kolmogorov-Smirnov Test)		Equality of means (2-sided)	
Variable	K-S *Z*	*p*	*t* statistic	*p*
Propensity score	.185	1.000	-.008	.994
LT Pre-test	.741	0.642	.048	.962
2004 CST ELA	.432	0.992	-.021	.984
Grade	.620	1.000	.076	.940
Track	.247	1.000	.467	.641

Scores from the MAST Reading Test were not used as a predictor variable in this matching process. For unknown reasons, there were a substantial number of students that did not have a MAST pretest score. Inclusion of this variable would have decreased the resulting group size by almost 20%. A second process on the whole group (n = 423) was undertaken substituting the LT pretest for the MAST pretest for analyses.

Analyses of Program Effect

Using the LT test scores as the primary measure of student outcomes, ANCOVA procedures revealed that the experimental students demonstrated a statistically significant treatment effect over their comparison counterparts (see Table 15). Similar, yet not statistically significant, results were produced using the MAST as the outcome measure, B (2, 209) = 2.535, p = .19.

Table 15

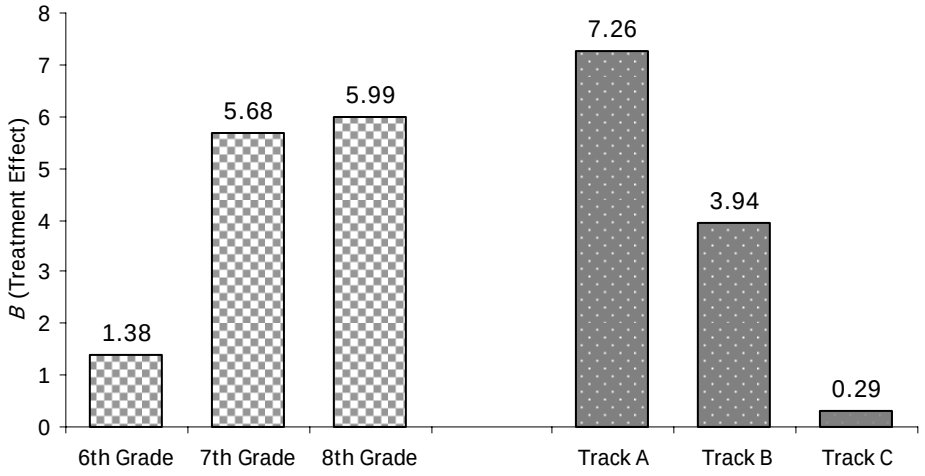
Analysis of Covariance (ANCOVA) Model Statistics for English/Language Arts

	n	Pre-test *M(SD)*	Post-test *M(SD)*	
Experimental	131	43.32 (19.43)	42.66 (17.80)	
Comparison	131	43.42 (17.23)	38.95 (17.39)	
	B	*SE*	*p*	η^2
Intercept	14.87	2.23	.000	
Pre-test	0.642	0.04	.000	
Group[1]	3.778	1.62	.020*	.021

[1]This variable was dummy-coded as 1 = Experimental group, 0 = Comparison Group

The data was also disaggregated by track and grade level. Similar to the findings of the Mathematics program, a significant program effect emerged for Track A students, B (1, 67) = 7.26, p = .07. In other words, Track A students participating in the English/Language Arts program outperformed their counterparts. Analyses by grade level also revealed significant program effect for 8th grade students, B (1, 66) = 5.99, p = .04, and 7th grade students, B (1, 78) = 5.68, p = .08. See Figure 2 for a graphic comparison of these results. Analyses conducted with the MAST matched groups did not reveal any treatment effects when disaggregated by grade level or school track.

Figure 2
Treatment Effects by Grade Levels and Tracks (English/Language Arts)



Student Attendance

Student attendance in the LT program was recorded by staff at each site and submitted to our evaluation team. As described earlier, student attendance varied greatly over the course of the 30-hour program. Students entered the program and exited at different points. In order to study the effects of attendance on student performance in the English/Language Arts program, we examined the relationship between hours attended and LT post-test scores (n = 509).

The average number of attendance hours was 19.03 hours (SD = 8.09). Students with matching pre- and post-tests (n = 261) had a higher average attendance with slightly less variability (M = 22.83, SD = 7.09). Analyses revealed a significant positive relationship between student attendance hours and LT post-test scores (r = .13, p = .03). A similar positive relationship was seen with the MAST post-test scores (r = .14, p = .01). In other words, the more time students spent in the program the greater their performance in English/Language Arts as demonstrated by their post-test scores.

English/LA Teacher Survey Results

Our evaluation team distributed surveys to 46 English/Language Arts teachers in Session B. Twenty-seven surveys (59%) were returned. Because of this small number, these survey responses were combined with those of the teachers from the seven non-study schools, thereby increasing the actual number of completed surveys to 53. Preliminary review of the

surveys did not reveal any differences among responses from teachers at study vs. non-study schools.

About 68% of the Learning Tree English/Language Arts teachers concurrently held traditional classroom assignments at the middle school level. Thirteen percent held assignments at the high school level. Ninety-three percent of the teachers attended the Learning Tree training prior to the start of Session B. Forty-one percent of the teachers reported that they were new to the Learning Tree Program. The average number of years teaching experience was 12.6.

Teachers were asked to rate the sequences within the English/LA program. The sequences were listed as follows: Autobiographical Sequence, Time-Sequence, Episode, Definition, Biographical Narrative, Compare-Contrast, Cause-Effect, and Problem-Solution. Autobiographical Sequence, Time-Sequence, and Definition received ratings of "strong" by over 50% of the teachers. The most often cited reasons for this were the organization of the sequences and the effective strategies. Those teachers who rated sequences as "weak" generally did not provide reasons for their ratings.

The teachers were asked to respond to a series of scaled-response questions regarding implementation and training/support. Table 16 presents the average responses to questions regarding implementation frequency of various program components. These questions were derived from the classroom observation rubric. The possible responses were as follows: 4 = "all of the time," 3 = "more than half of the time," 2 = "half of the time," 1 = "less than half of the time," and 0 = "not at all."

Table 16

Average Responses to Program Implementation (Session B)

Survey Items	*M (SD)*
Over the 9 weeks of the Learning Tree Program, how often did you …	
use *direct instruction* as part of your teaching?	3.00 (0.97)
model the activities for the students?	3.53 (0.83)
engage your students in the use of choral responses?	2.43 (1.15)
place students in cooperative learning groups or pairs?	2.81 (0.99)
implement the activities as demonstrated/explained in the curriculum materials?	3.43 (0.76)
supplement or substitute the program's activities or lessons with other material?	1.25 (1.31)
explicitly connect the instructional activities to the standards?	2.55 (1.24)

Average teacher responses indicated that modeling activities for students, implementing the activities as demonstrated or explained in curriculum, and using direct instruction strategies as part of the teaching occurred more than half of the time. Teachers also reported very little to no use of supplemental materials. Consistent with findings from classroom observations, teachers reported less frequent engagement of students in choral responding and placement if students in cooperative learning groups or pairs. Finally, explicit connections of instructional activities to the standards were made about half the time.

The survey also included questions regarding training and coach support. Table 17 presents the questions and average responses. For this series, the 5-point response scale ranged from "Strongly Disagree" (1) to "Strongly Agree" (5).

Table 17
Average Responses to Items of Training/Coach Support (Session B)

Survey Items	*M (SD)*
The coach assigned to our school was approachable.	4.39 (0.81)
When asked, the coach effectively demonstrated lessons/strategies in my classroom.	4.03 (0.92)
The coach was able to answer all of my questions about teaching the program.	4.29 (0.82)
The coach provided me with useful information and feedback.	4.17 (0.88)
When needed, the coach made herself/himself available to me.	4.36 (0.77)
I would like to have more training to implement this program.	2.98 (1.34)
In general, I would like more examples and explanations of the strategies included in the training.	2.98 (1.29)

Overall, teachers agreed that the coaching support provided by their school-site ALS coach was very positive. This coaching support included effective classroom demonstrations of lessons and strategies, availability, and ability to provide useful answers, program information, and constructive feedback on instructional delivery of curriculum. Teachers' responses were relatively neutral as to the need for more training and explanation. This was different from the Mathematics teacher responses which indicated a greater desire for additional training.

Finally, teachers were asked to describe, in their own words, their perceptions of their students' growth as a result of their participation in the Learning Tree Program. Forty-two teachers (79%) responded to this question. The patterns that emerged from these open-ended responses focused primarily on observed improvement in writing and reading comprehension skills and strategies. The following comments exemplified these themes:

> "Over the course of this program, my students have increased their ability to write complete sentences. They have also gotten better at using context clues to figure out the meaning of something they don't know."

> "Students that could write/speak in phrases or simple sentence at the beginning could speak/write compound and complex sentences at the end of 9 weeks."

> "They learned some useful writing skills when responding to literature – the use of transition words, the use of attention getting adjectives, the use of appropriate writing format."

Teachers also reported the development of student self-awareness, or the ability of the students to be aware of their own learning. The development of confidence was also reported:

> "I feel that students have started to think aloud and wondering about their own learning process."

> "Students are more willing to tackle tasks as they have new confidence in themselves."

> "They became more comfortable with their own progress and knowledge which then reflected back to the class as a whole."

Participation and attendance were areas that teachers believed influenced their students' progress.

> "For those that attended, a few 20%, seemed to gain momentum. Most were absent 50% of the session and I needed to review constantly."

> "Most kids didn't have their heart into learning or paying attention on Sat. AM. They went through the motions but I didn't see any growth."

SUMMARY AND DISCUSSION

The purpose of this evaluation was to provide evidence of the Learning Tree Program's effectiveness within LAUSD's Beyond the Bell middle school population. Five key questions were posed to guide this evaluation. Data were collected in response to these questions through both quantitative and qualitative methods. Summarized findings are presented below and discussed within the context of the five evaluation questions.

- *To what extent was the Learning Tree Program fully implemented in the Beyond the Bell Extended Learning Academies? What were the effects of implementation on middle school student outcomes?*

Quality and level of implementation of a program are strong contributing factors to its success. For the purpose of this study, quality and level of implementation were synonymous. The level of implementation of the Learning Tree Program was determined through observations of classroom instruction over the course of each session. These observations were coded and scored on program-aligned rubrics. Combining both sessions, the average implementation score fell at the "attempting to faithfully replicate" level. This level of implementation denoted an observed effort, although not always effective, at replicating the strategies, techniques, student engagement, and classroom management necessary for faithful program replication. The highest level of implementation, "exemplary," was only observed in twelve Mathematics classrooms. "Not faithful replication," or the lowest score, was observed in twenty classrooms, sixteen of which were English/Language Arts classrooms.

For both the Mathematics and English/Language Arts programs, level of implementation of the Learning Tree program was significantly related to how well students performed on post-test assessments. This finding holds great educational significance and supports educational research linking implementations to outcomes. One such study by Cooper (1998) of the Success For All Program found that quality of implementation not only impacted student outcomes but was also shown to influence school culture and weaken program resistance. Additionally, findings from a district-wide evaluation of LAUSD's reading program (Slayton, Oliver, & Burley, 2003)

strongly suggested that improving the quality of Open Court-related pedagogy and level of program implementation would lead to improved student performance.

- *To what extent was there a program effect on student outcomes for students participating in a Learning Tree curricular area (i.e. Mathematics) as compared to those who did not participate?*

Using propensity score matching techniques, we attempted to create statistically "comparable" groups to study the effect of participation in the Learning Tree program. The LT assessments were the primary outcome measure. This was largely due to the close alignment of the LT assessments to the curriculum. For students participating in the English/LA program in Session B, there was a statistically significant program effect. Results were similar, although not statistically significant, when using the MAST-Reading as the outcome measure. Participation in the Mathematics program did not produce a statistically significant program effect although patterns of scores did suggest differences between the groups. Program effects did vary by disaggregated sub-group (see next section.)

- *Did program effects vary by student sub-group?*

Studying program effects by student sub-group proved very difficult because of the small group sizes. In addition, the similarity of student characteristics across the four schools (i.e. language classification) also prohibited disaggregation. However, we were able to study differences in program effect by both student grade level and school track.

For both the Mathematics and English/Language Arts programs, differences in program effect were seen among grade levels and track. Track A students participating in the Number Sense Class A course significantly outperformed their non-math counterparts. Sixth graders also performed better, although the effect was not statistically significant at the .05 level. The weekday instruction of Number Sense standards at the sixth grade level to all sixth grade students, regardless of LT assignment, may have reduced the observable program effect between "experimental" and "comparison" groups. This interpretation was supported by the finding that, overall, sixth grade students outperformed students of the other grades, regardless of group

assignment. The educationally relevant implication is that the intervention not only added to but reinforced the learning that takes place during weekday instruction. Similarly, Track A students participating in the English/Language Arts course significantly outperformed their comparison group counterparts. As for differences by grade level, greater treatment effect was seen among seventh and eighth grade students.

- *What was the relationship between attendance and student outcomes? Did attendance vary across the two curricular areas (English/LA and Mathematics)?*

The maximum number of attendance hours of the Learning Tree Program was 30 hours per session. Across both programs, student attendance was positively related to performance. In other words, the more time students spent participating in the program, the better they performed on their post-test assessments. Students with matching pre/post-test scores had greater overall attendance (with little variability) than students with post-tests scores only. Because of this, we were unable to determine if greater attendance was linked to greater change in scores. The average attendance for students in the Mathematics program (Session A) was 20.81 hours and 19.03 hours for students in the English/LA program (Session B).

Although the average hours of participation fell below the full program recommendation, student attendance in voluntary out-of-school time intervention programs can vary greatly regardless of whether the program is externally or internally provided. In fact for 2004-05, LT attendance exceeded the average attendance for a similar type of district intervention program. Hodson (2005) reported in her evaluation of LAUSD's standards-based Extended Learning Program (ELP) that the average student attendance was 15 hours, of a possible 30, for both the 2002-03 and 2003-04 school years. Reasons for fluctuating attendance were not clear for ELP or Learning Tree and may require further study. Overall, since greater instructional time has been linked to achievement, the attendance patterns for the Learning Tree program are an indication of the potential benefits to all participating students.

- *What were the perceptions of the Learning Tree Program?*

Surveys were designed and administered to gather teacher background information,

perceptions of the quality and effectiveness of course components, information on program implementation and perceptions of training and coach support. Surveys were administered to all Mathematics teachers during Session A and all English/LA teachers during Session B. Because of low response, surveys from teachers in all eleven BTB schools were compiled.

The key to an effective Learning Tree program is the quality of implementation and replication in the classrooms. As we have seen, the observed classroom implementation was linked to student performance. Teachers perceptions of the program and their own implementation helped to provide an inside view of their experiences with Learning Tree. Overall, teachers were very positive about both programs and its effectiveness with their students. They reported an observed growth in specific comprehension skills, writing, and fluency. They also described their students' growing awareness of their own learning and improved confidence. Teachers rated the majority of sequence components for both programs as "strong" and "adequate." Common reasons included good organization, clarity of directions, ease of implementation, and effectiveness of strategies. Comparisons of responses across the two programs did reveal some small differences in program implementation. Mathematics teachers reported greater use of cooperative learning groups or pairs. English/LA teachers reported greater implementation of activities as demonstrated and/or explained in the curriculum materials. These teachers also reported fewer occurrences of supplementing or substituting within the LT curriculum.

The training and coach support that were also part of the Learning Tree program also received very positive feedback from teachers. Across both programs, teachers were very positive about the support, availability, and feedback that their school-site coach provided. The desire for greater explanations of strategies among the Mathematics teachers was very likely tied to the limited pre-program training that only addressed the first two sequences.

CONCLUSIONS AND RECOMMENDATIONS

The purpose of this evaluation was to study the effectiveness of the Learning Tree Program in the middle school academies of LAUSD's Beyond the Bell Branch. The integration of all evidence collected through the systematic study of program implementation, student outcomes, and teacher perceptions have led to very positive and constructive conclusions. The Learning Tree

Program provided students in need of academic intervention with opportunities such as increased instructional and learning time, reinforcement of learning that takes place in the regular week-day classroom, and the potential for smaller group interactions. Overall, the evidence presented has supported the Learning Tree Program as an effective out-of-school-time intervention program for the Beyond the Bell middle school population.

This study has shown that the quality of Learning Tree implementation was linked to student performance. This was the most significant finding of this study. Clearly, as LT teachers form the direct connection to the students, their training and on-going support to improve quality of implementation was the key to program success. The LT teachers have expressed very positive feedback about the support they received from their on-site ALS coaches. Continued coaching support and opportunities for follow-up training are strongly recommended.

As we have seen, the analytic results of program effect were mixed. However, the extent to which the limitations of the design and the comparability of the groups have influenced the score differences must be considered. Students in the English/Language Arts program did demonstrate a significant program effect. Students in Number Sense also demonstrated a similar pattern, yet did not achieve the traditionally accepted levels of statistical significance. There was no program effect demonstrated in the Algebra and Functions class. Subgroups of students, particularly Track A "experimental" students, significantly outperformed their "comparison" counterparts in both programs. Future study would be necessary to identify potential reasons for these track differences. Grade level differences also emerged and were very likely tied to the connections between Learning Tree standards of focus and the standards taught at specific grade levels during the regular school day. This connection may have blurred differences between the experimental and comparison groups. However, reinforcement of student learning was educationally relevant. Again, this would necessitate further study to examine the connections between traditional school instructional content and LT content.

Student attendance varies greatly in out-of-school-time programs. This reality impacts program effectiveness on several levels: low attendance may cause classes to close; the instructional pacing may be affected because teachers may have to repeat lessons for absent students; students not receiving benefits of the full program; and missing or unreliable assessment data. As we have seen, student attendance in Learning Tree was positively linked to student performance. Although average student attendance for the Learning Tree program was adequate

(almost 70%), as compared to other district out-of-school-time programs, incentives for increasing attendance should be considered.

REFERENCES

Brown, A., Palincsar, A., & Armbruster, B. (1984). Instructing comprehension fostering activities in interactive learning situations. In H. Mandl, N. Stein, & T. Trabasso, (Eds.), *Learning and comprehension of text*. Hillsdale, N. J.: Erlbaum.

Camilli, G., Vargas, S., & Yurecko, M. (2003). Teaching children to read: The fragile link between science and federal education policy. *Education Policy Analysis Archives*, *11(15).*

Cooper, R. (1998). *Socio-cultural and within-school factors that affect the quality of implementation of school-wide programs.* Baltimore, MD: Center for Research on the Education of Students Placed at Risk. (ERIC Document Reproduction Service No. ED426173).

Hodson, C. L. (2004). *Evaluation of extended learning program implementation in elementary schools during 2003-04*. Unpublished manuscript. Los Angeles: Los Angeles Unified School District.

Howell, K.W., Zucker, S. H. & Morehead, M. K. (1985). *MAST - Multilevel Academic Survey Test. Examiner's Manual*. San Antonio, TX: The Psychological Corporation.

Linn, R. L. & Gronlund, N. E., (2000). *Measurement and assessment in teaching, Eighth edition*. NJ: Prentice-Hall, Inc.

Marks, M., Pressley, M., Coley, J., Craig, S., Gardner, R., DePinto, T., & Rose. (1993). Three teachers' adaptations of reciprocal teaching in comparison to traditional reciprocal teaching. *The Elementary School Journal, 94(2), 267-283*.

National Institute for Literacy. (2003). *Put reading first: The research building blocks for teaching children to read*. Available online: www.nifl.gov/partnershipforreading/publications/ PFRbooklet.pdf.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: NICHD.

Quigley, D.D. (2003). *Using multivariate matched sampling that incorporates the propensity score to establish a comparison group*. (CSE Technical Report No. 596). University of California, Los Angeles. Retrieved December 8, 2004 from http://cresst96.cse.ucla.edu/reports/TR596.pdf.

Rivera, N.V., Burley, K., & Sass, J.S. (2004, February). *Evaluation of school-based professional development (2002-03)*. Unpublished manuscript. Los Angeles: Los Angeles Unified School District.

Slayton, J., Oliver, D. & Burley, K., (2003, October). *Evaluation of the district reading plan, 2002-2003*. Unpublished manuscript. Los Angeles: Los Angeles Unified School District.

Snow, C.E., Burns, M.S., & Griffin, P. (Eds.) (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Varma, S. (n.d.). *Computing and using point biserials for item analysis*. Retrieved November 2004 from www.eddata.com/resources/publications.

APPENDIX A
Technical Information for ALS Assessments

The following table presents item statistics (p-value, pt. biserials) and reliability estimates (Kuder-Richardson 20, Cronbach's alpha) for the ALS-developed Learning Tree assessments used in this study. This information has been calculated with assessment scores from Beyond the Bell students.

	# Items	N	Mean P-Value	Mean Pt. Biserial	Kuder-Richardson 20	Cronbach's Alpha
Mathematics Assessments						
Number Sense Class A Pre-Test	25	1541	.36	.40	.74	.74
Number Sense Class B Pre-Test	25	170	.30	.26	.41	.41
Algebra/Functions Class A Pre-Test	25	644	.35	.31	.61	.61
Algebra/Functions Class B Pre-Test	25	166	.33	.34	.65	.65
Number Sense Class A Post-Test[a]	25	257	.46	.43	.79	.79
Number Sense Class B Post-Test[a]	25	29	.31	.29	.58	.58
Algebra/Functions Class A Post-Test[a]	25	65	.39	.34	.67	.67
Algebra/Functions Class B Post-Test[a]	25	33	.41	.31	.62	.62
English/Language Arts Assessments						
ELA Benchmark Form 1 (Pre-Test)	35	1233	.40	.38	.79	.79
ELA Benchmark Form 2 (Post-Test)	35	826	.44	.35	.77	.77

[a] The post-test assessment scores used in these analyses do not include students that participated in these courses.

P-values are the proportions of students that respond correctly to a test item. Pt. Biserial correlations describe the relationship between the student scores for a given item and the student scores for the whole test. Positive pt. biserials above .20 generally indicate that the items are performing as anticipated (Varma, n.d.).

This study also provided the opportunity to perform additional reliability testing (equivalent forms method over time) which is one of the soundest methods for estimating reliability (Linn & Gronlund, 2000). The pre-tests and post-tests for each course were constructed as "equivalent" forms. Comparison groups of students that did not participate in the particular curricular area were used for this analysis.

Equivalent Forms		**N**	**Correlation**
Number Sense Class A Pre-Test	Number Sense Class A Post-Test	206	.52**
Algebra/Functions Class A Pre-Test	Algebra/Functions Class A Post-Test	49	.66**
ELA Benchmark Form 1 (Pre-Test)	ELA Benchmark Form 2 (Post-Test)	540	.67**

**p<.01

Tests for the Class B assessments for both Number Sense and Algebra/Functions were not conducted at this time due to extremely small sample sizes.

APPENDIX B
Quality of Implementation Rubric

4	Exemplary	• Students are highly engaged in the lesson • The activities are implemented exactly as written/described in curriculum and/or have been demonstrated • Students are in cooperative groups or pairs during activities • Many opportunities have been provided for students to share information and ideas • Classroom and student behavior is well-managed • Each activity is modeled using a structured, interactive conversation with the students (direct instruction) • There is modeling of how to use manipulatives and opportunities are provided for students to use them • Standards are emphasized and instruction/activities are being connected to the standards
3	Faithful Replication	• Students are following teacher's direction during lesson/activity • The activities are implemented in the correct order in a way that the students are responding to • Students are in cooperative groups or pairs during activities • Some opportunities have been provided for students to share information and ideas • Classroom and student behavior is adequately managed • Each activity is modeled in a lecture format • Opportunities are provided for students to use manipulatives • Standards are emphasized during instruction
2	Attempting to Faithfully Replicate	• Students are not engaged in the lesson although effort at engagement is observed • The activities are implemented in a way that is observed as confusing to students • Students are in cooperative groups or pairs some times • Few, if any, opportunities have been provided for students to share information • Classroom and student behavior is not managed effectively • Activities are modeled but not effectively (sometimes incorrectly) • Opportunities are provided for students to use manipulatives (although not often correctly) • Standards are rarely emphasized
1	Not Faithful Replication	• Students are not engaged in the lesson (no effort is observed) • The activities/instruction not implemented correctly (observed use of unrelated materials) • Students are working individually on all/most of activities • No opportunities have been provided for students to share information • Classroom and student behavior is not managed (Teacher may be disengaged) • No modeling is occurring (activities used as worksheets) • Manipulatives are not used • Standards are not discussed – no evidence of students understanding reasons for working