

REPORT

The Leader in Me Evaluation Report

April 2015

**Prepared for
United Way of Acadiana**

**Prepared by
Ray Biggar Jr., Ph.D.
Steven Dick, Ph.D.
Alisha Bourque M.S.
Cecil J. Picard Center for Child Development and Lifelong Learning
University of Louisiana at Lafayette**

Acknowledgements

United Way of Acadiana

Margaret Trahan
CEO

Jason Huffman, Ph.D.
Director of Impact Studies

University of Louisiana at Lafayette

Ramesh Kolluru, Ph.D.
Acting Director, Picard Center
Vice President of Research

James Henderson, Ph.D.
Provost and Vice President of Academic Affairs

Cooperating School District Staff

Dr. Donald Aguillard
Lafayette Parish Superintendent

John Bourque
Acadia Parish Superintendent

Tom Spencer
Director of Accountability, Lafayette Parish

Adele Carr
Pre-K and Elementary Education, Acadia Parish

Bryan Alleman
Technology Coordinator, Acadia Parish

Contents

Contents	3
Figures	4
Executive Summary	5
Background Information	5
Study Description	6
Research Design:	6
Evaluation Questions:	7
Population	7
Academic Achievement.....	8
iLEAP Math	8
iLEAP ELA	11
LEAP Math	12
LEAP ELA	13
DIBELS.....	14
Behavioral Changes	15
Attendance	15
Suspensions	16
Expulsions.....	17
Office Discipline Referrals (ODR's)	17
Limitations.....	19
Conclusions	20
Recommendations for Future Studies	21

Figures

Figure 1: iLEAP Math by Cohort and Test Year.....	9
Figure 2: iLEAP Math by Cohort and Years for Students in Poverty.....	10
Figure 3: iLEAP Math by Cohort and Years for Male Students.....	10
Figure 4: iLEAP Math by Cohort and Years for African American Students.....	11
Figure 5: iLEAP ELA by Cohort and Years	11
Figure 6: LEAP Math by Cohort and Years.....	12
Figure 7: LEAP Math by Cohort and Years for African American Students.....	12
Figure 8: LEAP ELA by Cohort and Years	13
Figure 9: LEAP ELA by Cohort and Years for African American Students.....	13
Figure 10: LEAP ELA by Cohort and Years for Students in Poverty	14
Figure 11: DIBELS Results by Cohort and Year	15
Figure 12: School Attendance by Cohort (Percent)*.....	16
Figure 13: Percentage of Students Suspended by Cohort	18
Figure 14: Number of Expulsions by Cohort	19
Figure 15: Office Referrals by Cohort (Actual Count)*	19

Executive Summary

The purpose of this report was to begin the process of understanding the effect of The Leader in Me (TLIM) program on local schools. Two cohorts of schools were considered. First, most of the evaluation involved changes in Cohort 1 schools that have been participating in the program for two or three years. Second, this report sought to provide a baseline for schools just starting the program. This initial report provides some encouraging results on academic measures. In particular:

- TLIM schools consistently outperformed Control schools in Math on both LEAP and iLEAP.
- TLIM schools moved from underperforming to outperforming Control schools on DIBELS.
- African American students were more likely to reach benchmark in TLIM schools.
- TLIM schools were more likely to be on benchmark in ELA (LEAP and iLEAP) both before and after the introduction of the program.
- Attendance at TLIM schools has remained fairly stable for both cohorts.
- Trends in suspensions and expulsions are inconclusive due to data inconsistency. Future evaluations will require student level data to ensure analysis and outcomes are generated at similar or the same aggregated level.

TLIM outcomes on behavior indicators are somewhat encouraging but less conclusive than academic indicators. The inconclusive outcomes related to behavior can be attributed to poor data quality provided for analysis. Student level data was used for achievement indicators. Therefore the level of confidence related to outcomes on achievement is much higher than findings on behavior indicators. Overall, there is no upward or downward trend on behavior measurements at the cohort level. However, much of the variation can be attributed to one school in Cohort 1 and one or two schools in Cohort 2. Please read the section on recommendations for future evaluations to aid in reporting findings on behavior indicators.

Background Information

TLIM is an initiative for schools developed by FranklinCovey. The program asserts that it positively impacts school culture, academics, and student skills necessary to be successful in the **21st** century. In addition, aspects of TLIM can have a profound impact on the community as students, teachers and school administration infuse program components in areas beyond school walls. The foundation for change is based on the integration of Steven Covey's 7 Habits of Highly Effective People into all aspects of the school experience. The TLIM model also provides students with a conduit to discover, develop or refine leadership skills that each possesses.

TLIM program, in Acadiana, was initially implemented at two schools in Lafayette parish and Martin Petitjean Elementary school in Acadia parish during the 2010-11 school year. Within three years Martin Petitjean was recognized as one of only 50 schools nationwide (first in Louisiana) to be recognized as a "lighthouse" school among

all schools implementing the TLIM model. Since 2010 the program has expanded to schools in Acadia and Lafayette Parish. In spring of 2013, FranklinCovey and the Picard Center co-sponsored a symposium in Lafayette on TLIM. During this time, the Lafayette Parish School System (LPSS) expressed interest in district wide implementation of TLIM. The United Way of Acadiana has entered into a partnership with LPSS to assist in funding schools to implement TLIM. Additional schools in the Acadiana region are currently implementing TLIM, with the support of other local nonprofit organizations such as Lafayette Education Foundation.

Study Description

Acadiana area schools that have been implementing TLIM prior to the 2013-14 school year were designated as Cohort 1:

- J. Wallace James Elementary (2 + years)
- Judice Middle (3 + years)
- Martin Petitjean (3 years)
- Plantation Elementary (3 + years)

During the 2013-14 school year, the United Way of Acadiana created a cohort (Cohort 2) of nine schools in Acadia and Lafayette Parishes that began implementing TLIM. Cohort 2 is comprised of the following schools:

- Central Rayne Kindergarten (Acadia)
- Ross Elementary (Acadia)
- Carencro Heights Elementary
- Ernest Gallet Elementary
- Ossun Elementary* (not funded by UWA)
- Ridge Elementary
- Milton Elementary
- Woodvale Elementary
- Youngsville Middle

The purpose of this evaluation was to collect, analyze and report on data relevant to outcomes of TLIM at the grade, school and cohort levels. It should be noted, program fidelity is a key factor when analyzing outcomes; however, an instrument for measuring this variable (fidelity) has not been developed. Therefore implementation and its potential impact on outcomes cannot be reported on at this time. The goal of this evaluation is to analyze trends on academic and behavioral indicators in schools that have agreed to implement TLIM initiative.

Research Design:

This project was designed to evaluate the effects of a large and growing group of students. It was designed to accommodate a group of students that will quickly outgrow a

traditional matched comparison methodology. The data was compared with district and state trends in the first year of program implementation at the school. Starting in the second year of implementation, the primary methodology was an interrupted time series in which the trend in evaluation scores prior to program introduction will be compared to the trend after the evaluation. The reporting for this evaluation is composed of two groups: Cohorts 1 and 2. The students comprised of each cohort will serve as the sample for this study. Outcomes for individual schools are not included in this evaluation. Primary evaluation scores include:

- DIBELS (when available)
- iLEAP/LEAP (when available)

Intervening variables include:

- Gender
- Ethnic Group
- Free or Reduced Meal Plan Status (poverty).

As noted in the previous section, the lack of an instrument to measure program fidelity severely hampers the evaluation of TLIM. Implementation of any program can be fluid and fluctuate from year to year. For example, changes in administration or educators can have an adverse impact on implementation, particularly if new staff does not buy into a program. Measuring program fidelity is a hallmark of any program evaluation. It provides valuable information on the current status of a program and evaluators with a metric to generate inferences related to level of implementation and outcomes. Without a measure of implementation it is difficult to connect outcomes to TLIM. With respect to this evaluation, reporting by cohort (length of time) implies, schools that have been implementing TLIM longer should have markedly different outcomes than schools just beginning to implement. Unfortunately, we have no way to measure trends in implementation among schools in each cohort or for individual schools.

Evaluation Questions:

1. What are the trends in student achievement in English Language Arts (ELA) and Math before and after school entrance into TLIM initiative?
2. What are the trends in student behaviors such as office discipline referrals (ODR's), attendance and discipline before and after implementation of the TLIM initiative?

Population

As much as possible, the selected population constitutes a census of the students in the cohort schools. In addition, non-TLIM schools (same grade levels in schools not implementing TLIM) in the two districts are included as the Control cohort. The Control

group represents students attending schools that are not currently implementing TLIM. Therefore, Control group and non-TLIM schools should be interpreted synonymously when reading this report.

Major differences are expected and tested between Cohort 1 and the Control group. Cohort 2 is included in the following analysis to establish a baseline for later research. Given that this is the first year of TLIM for Cohort 2, no substantial effects are yet expected.

TLIM is a non-prescriptive; therefore program effects may take longer to generate outcomes. Staff and students must develop goals, procedures, and systems to institute the program. As such, the introduction period is more of a gradient than an absolute line.

Academic Achievement

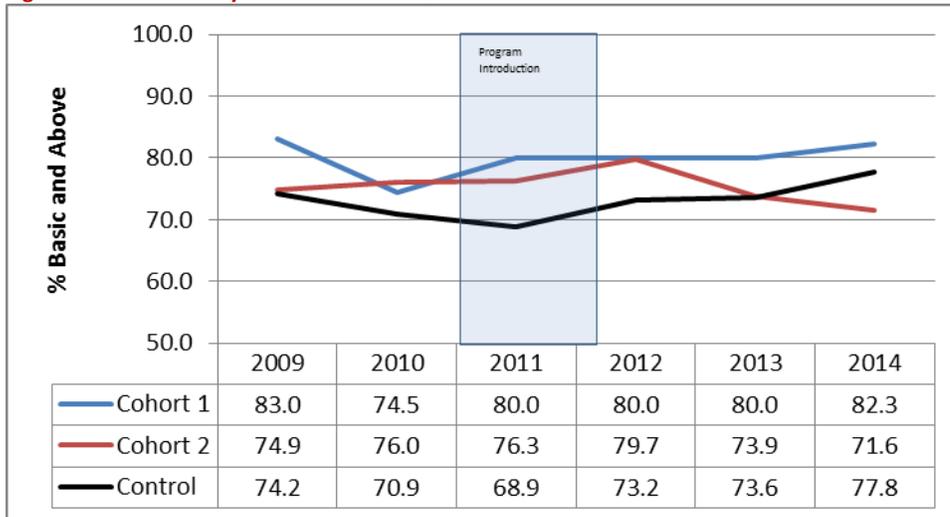
Academic achievement is one area potentially impacted by TLIM. As students take control of their own goals and feel responsible for achievement the value of succeeding academically becomes more intrinsic. As discussed in the prior section, most of the analysis concentrates on the difference between Cohort 1 and the Control group. The Control group represents students that are not attending schools currently implementing TLIM.

iLEAP Math

The iLEAP test is given at 3rd, 5th, 6th, and 7th grade. The analysis in this section deals with the Math portion of the evaluation instrument. It will highlight those subgroups where significant or important effects were found. In the following analysis, the term “benchmark” is used to refer to levels basic or above on the iLEAP assessment. Significant differences are tested using an ANOVA. The iLEAP is given in the Spring so the year refers to the test year (e.g., 2011 is for the 2010-11 school year).

Figure 1 (below) summarizes the long term trend in iLEAP scores across years. Cohort 2 has been included in this figure to establish a baseline measure on this indicator, but the program is not yet mature for evaluation. The important change was between Cohort 1 and non-TLIM implementing schools on the same measure. Prior to 2010, the two groups’ percent on benchmark was nearly identical. After the 2010 school year (test in spring of 2011) the two groups exhibited significantly different percent of iLEAP test takers performing at basic or above (ANOVA $F= 41.2$, $p=0.002$). The percentage of Cohort 1 students achieving benchmark on this indicator has been remarkably consistent (minimum 80%) since TLIM implementation began in 2010. Although the percentage of students achieving benchmark in non-TLIM schools has increased, they remain 4-6% below Cohort 1 during the same reporting period.

Figure 1: iLEAP Math by Cohort and Test Year¹



A slightly different dataset was used to look at the effect on subgroups. Student level data was not available for the 2014-15 school year; therefore, Cohort 2 is not included in the following subgroup analysis. The analysis presents the results when the data produced interesting or significant results. First, students in poverty (as measured students eligible for the free and reduced lunch program) indicated a nearly identical performance between Cohort 1 and the non-TLIM (Control) group prior to the Spring 2011 iLEAP (see Figure 2). Following that year, there was a significant difference in the percent on benchmark (ANOVA $F=35.4$, $p=0.001$). Poverty students attending TLIM implementing schools were more likely to be on benchmark than those in the non-TLIM schools (Control group). Although the percentage of students scoring at basic or above is lower for this subgroup (poverty), students achieving benchmark at TLIM schools are 5-6% higher when compared to non-TLIM schools. This trend was similar when comparing all students to students attending non-TLIM schools. The consistency of percentages of students scoring at basic or above over three reporting (school years) periods suggest, at a minimum, TLIM may have a positive impact on math scores. However, more time is needed to see if this trend continues longitudinally.

¹ Data comes from an evaluation of state reports of school level data for 36 qualifying schools in the two districts.

Figure 2: iLEAP Math by Cohort and Years for Students in Poverty²

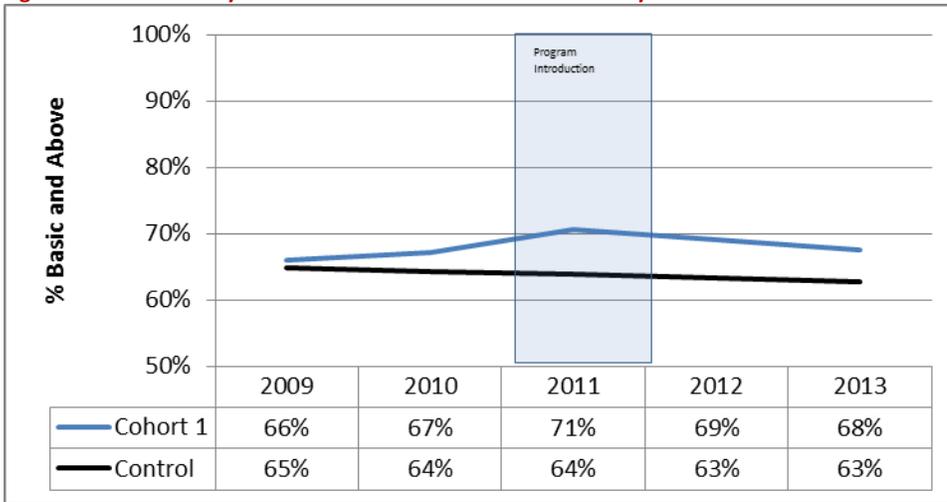
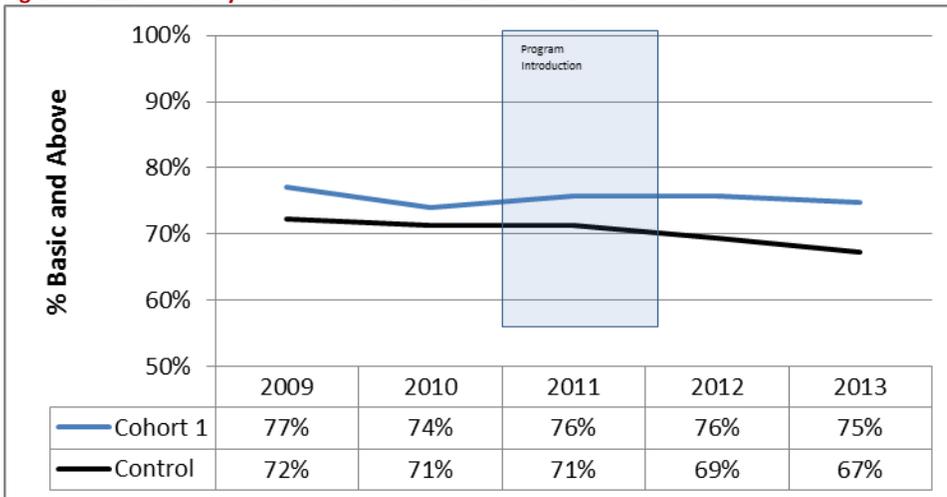


Figure 3: iLEAP Math by Cohort and Years for Male Students



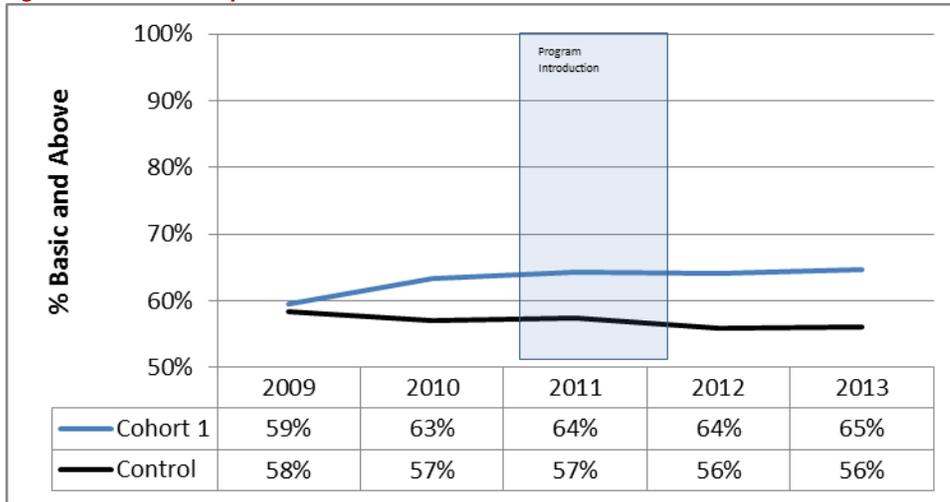
Virtually the same effect was seen for male students (see Figure 3 above, ANOVA $F=25.4$, $p=0.002$). However, the significant difference started with the iLEAP test in the Spring of 2010. Cohort 1 males are 5%-8% more likely to achieve benchmark, when compared to males in non-TLIM schools. Additionally, the percentage of Cohort 1 males that score at basic and above has remained consistent since the program introduction, while males in the Control group have declined.

African American students displayed the same increase in percent on benchmark with the Spring 2011 iLEAP (ANOVA $F=237.2$, $p<0.0001$). The one difference is that African American students in Cohort 1 were out performing the Control group prior to TLIM program (ANOVA $F=10.3$, $p=0.006$). However, the percent on benchmark for the African American students in Cohort 1 has grown by 11% over the study period as opposed to a 2% decline for the Control group. Overall, students at TLIM schools

² Data for subgroup analysis was produced from student level data across the affected years, and school districts N=82,119. The same dataset was used for ELA analysis.

(Cohort 1) were more likely to score at basic or above on iLEAP math than students attending schools not implementing TLIM.

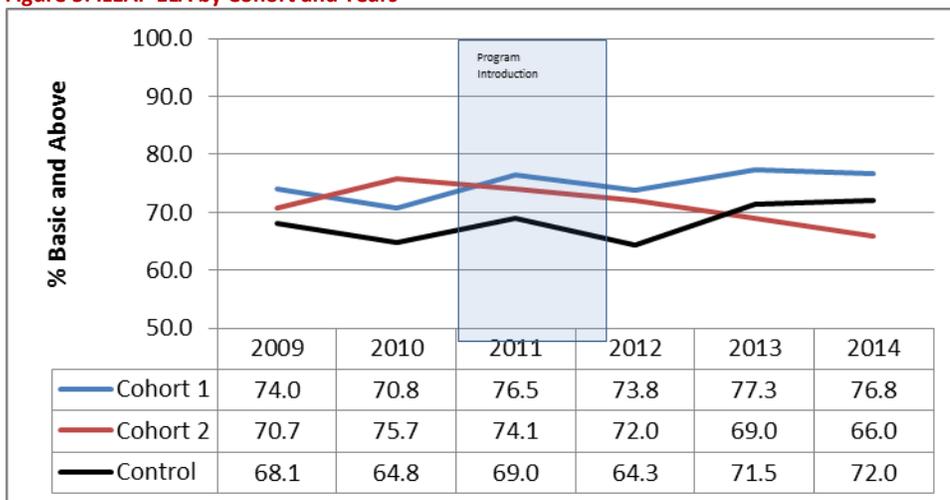
Figure 4: iLEAP Math by Cohort and Years for African American Students



iLEAP ELA

Turning to the English language arts (ELA) section of the iLEAP (see Figure 5), a similar overall difference between Cohort 1 and the non-TLIM schools was evident (ANOVA $F=15.4$, $p=0.002$). Visually, Cohort 1 and the Control experienced similar shifts in performance. Cohort 1 had about 6% more students on benchmark across the years available. Overall, there was not strong support for an effect of TLIM on ELA scores. Subgroups also showed nearly identical percent of students on benchmark with mostly non-significant findings.

Figure 5: iLEAP ELA by Cohort and Years



LEAP Math

LEAP Math performance may see more definitive results in the years to come. Just before the introduction of TLIM, Cohort 1 schools surged to 18% more on benchmark than the Control group and then closed quickly to a difference of less than 6%. Right after the introduction of TLIM, Cohort 1 Schools rebounded and continued to outpace the Control group (ANOVA, $F=26.4$, $p=0.0009$). Subgroup analysis consistently showed the same large advantage of Cohort 1 schools over the Control both before and after the introduction of TLIM. As seen in Figure 7, African American students were significantly more likely to be on benchmark in Cohort 1 schools than in the Control group (ANOVA $F=70.8$, $p<0.0001$). Visually, it is not yet certain that TLIM schools will widen the gap but they are clearly outperforming Control schools. Other demographic subgroups showed the same strong performance in LEAP math.

Figure 6: LEAP Math by Cohort and Years³

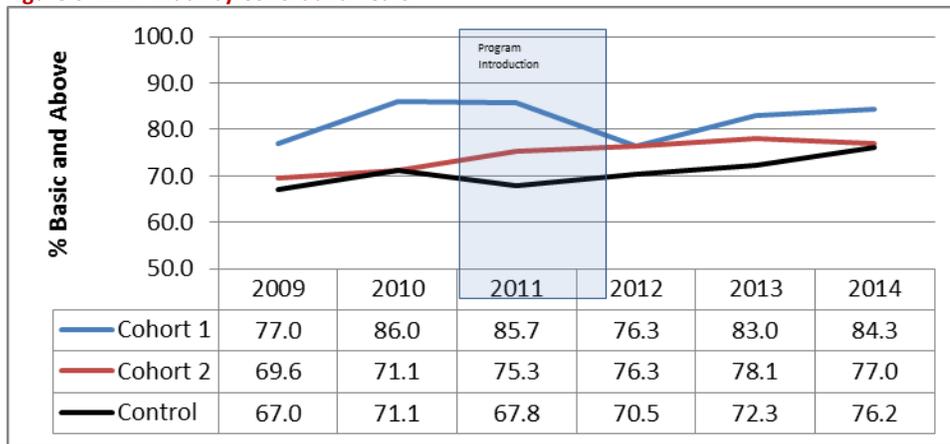
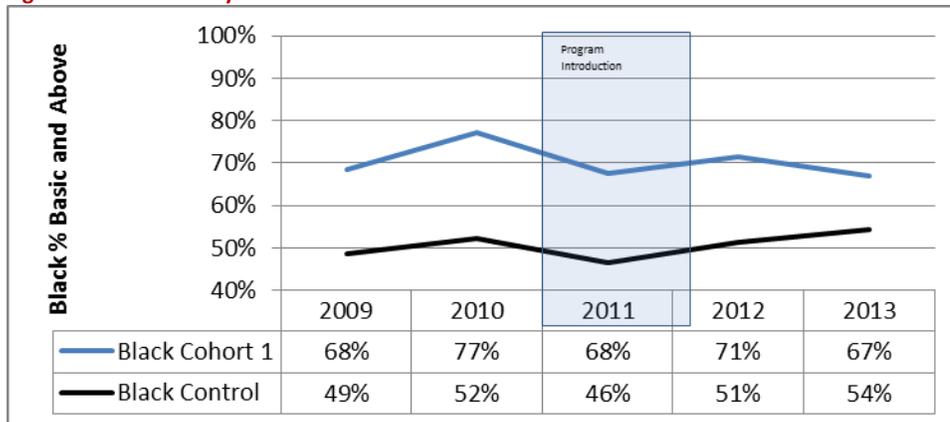


Figure 7: LEAP Math by Cohort and Years for African American Students⁴



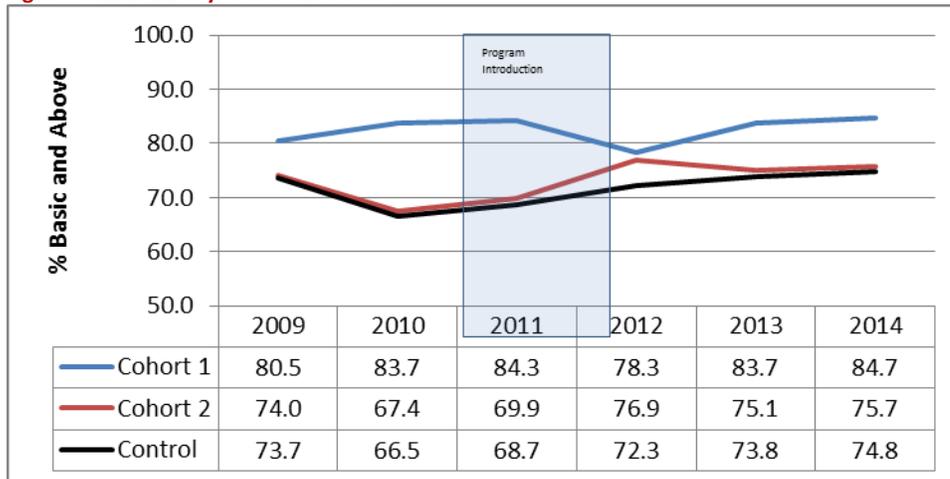
³ Data comes from an evaluation of state reports of school level data for 36 qualifying schools in the two districts.

⁴ Data for subgroup analysis was produced from student level data across the affected years, and school districts N=42,334. The same dataset was used for ELA analysis.

LEAP ELA

Just as in LEAP Math, LEAP ELA included a slight reduction in Cohort 1 schools percent on benchmark just as TLIM program started. Clearly, Cohort 1 schools were more likely (8-16%) to be on benchmark than non-TLIM schools (ANOVA $F=41.0$, $p<0.0001$).

Figure 8: LEAP ELA by Cohort and Years



Subgroup performance echoed the entire cohort with TLIM Cohort 1 consistently more likely to be on benchmark. As displayed in Figure 9, African American students in TLIM program were 16% to 24% more likely to be on benchmark than Control group schools (ANOVA $F=67.4$, $p<0.0001$). The same was true for students in poverty (see Figure 10, ANOVA $F=94.6$, $p<0.0001$). Overall, TLIM effects on academic indicators are promising and aligned with existing research that indicates TLIM has some effect on academic performance indicators.

Figure 9: LEAP ELA by Cohort and Years for African American Students

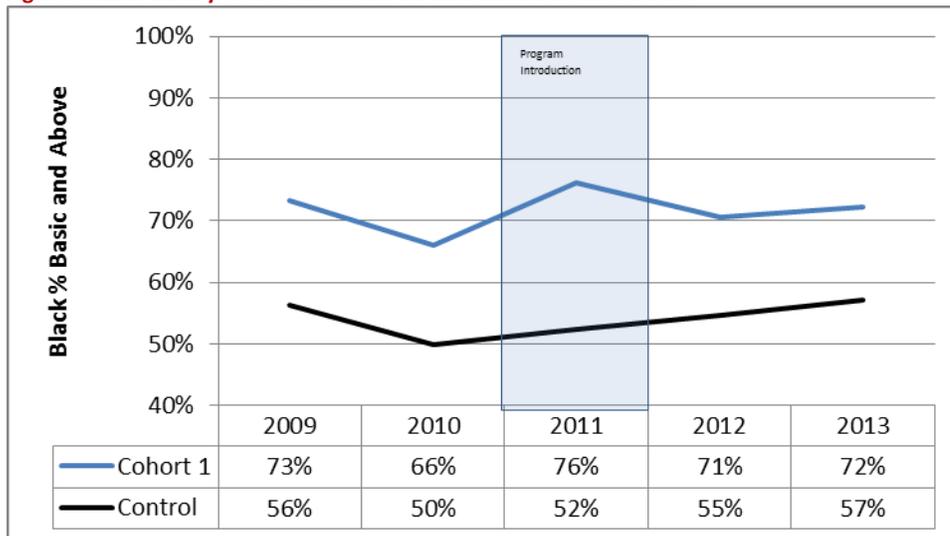
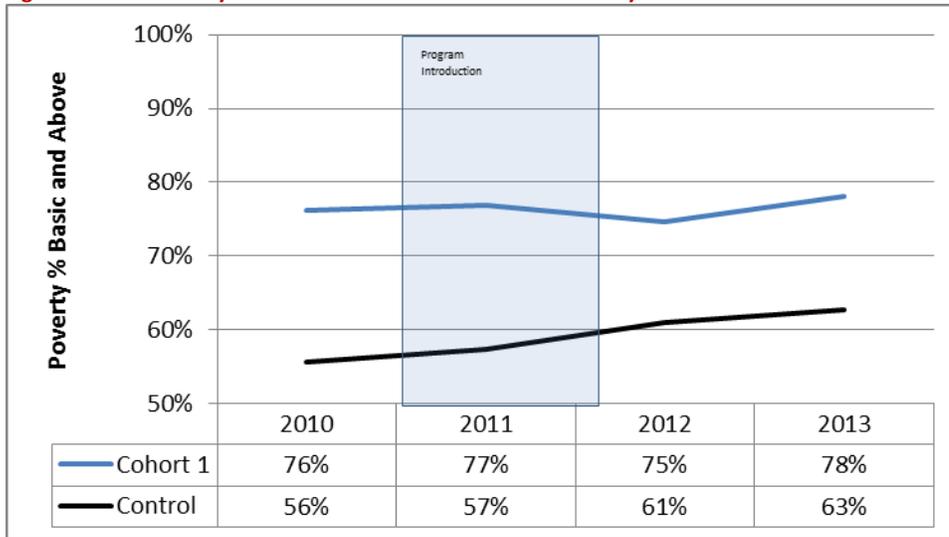


Figure 10: LEAP ELA by Cohort and Years for Students in Poverty

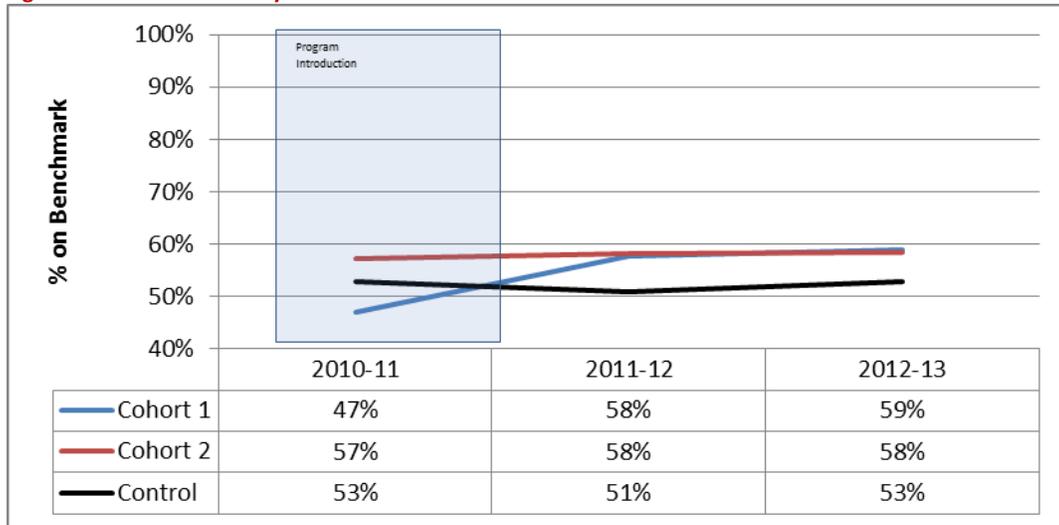


DIBELS

DIBELS is a state mandated evaluation of literacy and pre-literacy skills given to all kindergarten to third grade students. Similar results were found when analyzing DIBELS data, as results on other achievement indicators. The analysis indicates TLIM schools are outperforming Control (non-TLIM) schools and that a large increase occurred once the program was implemented. Benchmark in DIBELS is defined as being on the top of three levels. In some versions of DIBELS, benchmark is called “less risk.” In others, it is called “core.” The evaluation used instructional recommendation or comprehensive need for support as a summary measure.

Cohort 1 schools experienced a dramatic shift after the introduction of TLIM. Cohort 1 schools moved from 6% below Control Schools to 6% above for a significant difference (ANOVA $F=27.9$, $p=0.03$). The impact of TLIM on this performance indicator is very promising. Additionally, DIBELS scores in Cohort 1 increased significantly by 11% in the first year of TLIM program and maintained that high level (increasing by another percent) in the second year of the program. A DIBELS subgroup analysis is not provided due to data quality concerns with the student level data of the participating schools. Since the population for this sample is very young, there is some indication TLIM has some impact on achievement very early in education.

Figure 11: DIBELS Results by Cohort and Year⁵



Behavioral Changes

Attendance

The following figures present data from both Cohorts 1 and 2. The changes should be more pronounced in Cohort 1, given the length of time TLIM has been implemented for this group. As the data allows, analysis was performed at the school level (aggregated to cohort) using a regression discontinuity design or ANOVA to compare attendance across school years beginning with the 2007-08 school year and ending with the 2013-14 school year. It should be noted that implementation of TLIM begins prior to student exposure to the program. In addition student exposure to the program varies across schools. For example, a school may begin implementation in the fall, but student exposure may not start until the spring. This could have an impact on attendance outcomes.

Attendance is a key behavioral measure for TLIM. However, as demonstrated in Figure 12, Cohort 1 schools did not display a significant change in attendance. The slight downward trend in Cohort 1 did not appear to be significant. Visually there was an aggregated decline in attendance in Cohort 2 of 1% between baseline and year one of implementation. However, it is too soon to suggest this is a trend for Cohort 2. Historically, attendance tends to remain fairly stable indicator over time, with small variations in change. It may be prudent to inquire about the significance (not statistical) of these small variations and what is an acceptable percentage for this measurement. This may add to the robustness of this performance measure.

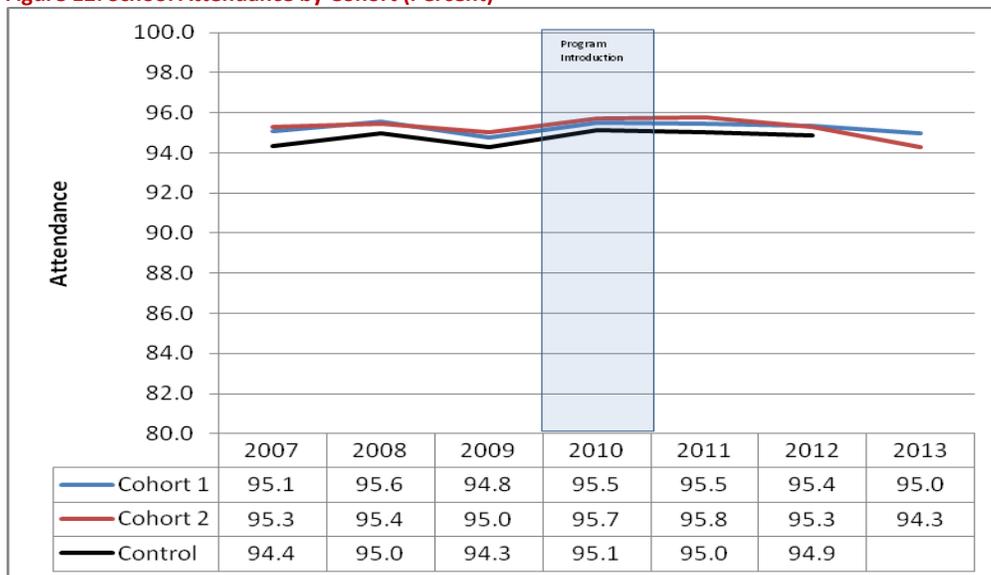
It's important to also look at the individual schools and variation in attendance. Although this report does not report at the school level, it's clear that outliers in both

⁵ Data for this figure comes from student level data of schools in participating parishes over the years (N=37,548).

cohorts are having a negative impact when reporting on the cohort level. The impact is more salient in Cohort 1 which has only 4 schools. For example, a school in Cohort 1 had a 3% increase in attendance during the first year of TLIM implementation, followed by a 2% (or more) decrease for the 2012-13 and 2013-14 reporting periods. The other schools in the cohort remained fairly stable. In essence, attendance has remained fairly unchanged from year to year (figure 12). Since students at each school are exposed to TLIM at different periods, even within the same school year, it may be more informative to report this indicator by school in future evaluations.

Available data on tardiness was suspiciously inconsistent and was dropped from analysis due to unreliability. Tardiness may be a more robust behavioral indicator for TLIM, particularly if it captures punctuality while in school. Personal responsibility is a leadership attribute that could be measured by tardiness. However, this data may be too granular to capture reliably and consistently.

Figure 12: School Attendance by Cohort (Percent)*



Suspensions

Due to inconsistency and incomplete data sets, behavior indicators for this report were difficult to analyze. Most of the behavioral data sent for evaluation was aggregated to school level (e.g. occurrences). At this level, the analysis of the data is limited to school level only. Duplicates, or individuals sent to the office more than once cannot be removed. Non-duplicate data was provided for suspensions therefore, the percentage of students (by cohort) suspended at least one time could be calculated. However, analysis by subgroup(s) was not possible without student level data. The inconclusive outcome on indicators related to behavior underscores the importance of using school level data for reporting. Additionally, adding a control group was not possible with the aggregated data provided on behaviors.

Due to differences in data reporting; suspensions, expulsions and office referrals are reported here beginning in 2010-2011 and ending 2013-2014 school year. Reliable comparative data was not available prior to 2010 (the start of Cohort 1). Additionally, figures 14 and beyond do not include data from schools in Acadia parish. Behavioral data from the parish was not made available for this evaluation. Later reports can and will expand upon the indicators if data is provided for evaluation.

Figures 13 and 14 represent the percentage of students (non-duplicate) suspended or expelled at least once during the school year. A change (e.g. policy) in suspensions and expulsions in Lafayette parish schools occurred during the 2012-2013 school year that cannot be fully explained by the data. Both dropped sharply for both cohorts. The data was triangulated with other data sources which produced the same or similar numbers. In Cohort 1 the number of students suspended declined at every school in Cohort 1 and all but one school in Cohort 2. Overall suspensions dropped by 8% in Cohort 1 during the 2012-2013 school year but returned almost to previous levels the following year. Cohort 2 suspensions increased slightly from between baseline (4.28%) and year 1 (6.05%) of implementation. However, the increase is still lower than the percentage of students suspended in the two years prior to program introduction.

A closer examination of school level data revealed the similar findings as attendance. Cohort 1 has an outlier school that contributes heavily to its suspension rates. One school in Cohort 1 accounted for approximately 75% of all suspensions in the cohort. Although there is variation among schools suspensions in Cohort 2, it is not to the degree that is seen in the first cohort. Also, Cohort 2 has twice as many schools as Cohort 1. Therefore, changes (positive or negative) for the older cohort are more pronounced. Reporting by school in place of cohort would be more informative for this indicator.

Expulsions

Expulsions in both cohorts are virtually nonexistent. This can only be explained by a change in policy or data not being collected by the parish. A dramatic reduction in expulsions due to policy generally causes a spike in suspensions. However, both indicators showed sharp declines in 2012-2013. The data in figure 14 represents the total number of expulsions (occurrences) for each cohort. Therefore, the numbers prior to the 2011-12 school year was inflated by duplicates. Subsequent evaluations should provide some clues, if a change in policy and/or reporting is the cause of the decline in expulsions. This indicator should be monitored for relevancy in future evaluations.

Office Discipline Referrals (ODR's)

Perhaps the most sensitive behavioral indicator for TLIM is office discipline referrals (ODR's). This indicator captures all behavior issues, including those that do not lead to suspensions and/or expulsions. It also taps into school climate and classroom management. Theoretically, both are areas TLIM should positively impact. The data supplied for this evaluation provides the total number of ODR's by cohort. Since the evaluator was not provided with student level data, there was very little we could do with this indicator. Duplicates could not be eliminated, subgroups could not be analyzed,

major and minor infractions separated and analyzed and number of referrals stratified could not be performed. The numbers in figure 15 represent the total number of ODR's (occurrences) by cohort. Further analysis was not possible without student level data.

Similar to the previous behavior indicators, the numbers for Cohort 1 are inflated by one school. However, there was a 40% reduction in Cohort 1 ODR's between 2012-2013 and 2013-2014 school years. This was the first year a reduction in ODR's was captured for this cohort. A smaller (4.2%) decrease occurred in Cohort 2 during year one of implementation (2013-2014).

It should be noted; reporting on single occurrences for this indicator is not the optimal measurement. ODR's are categorized as minor and major offenses. It would be more informative to report referrals as they are categorized. Additionally, some school districts classify a referral as "major" once a student has three cumulative referrals. Student level data would allow the evaluator to report ODR's as categorized and cumulative. The number of referrals may dramatically drop after one or two incidences. This can be a more accurate measure of TLIM impact on behavior at the school level. However, data must be provided at the appropriate level that allows for analysis at this level.

Figure 13: Percentage of Students Suspended by Cohort

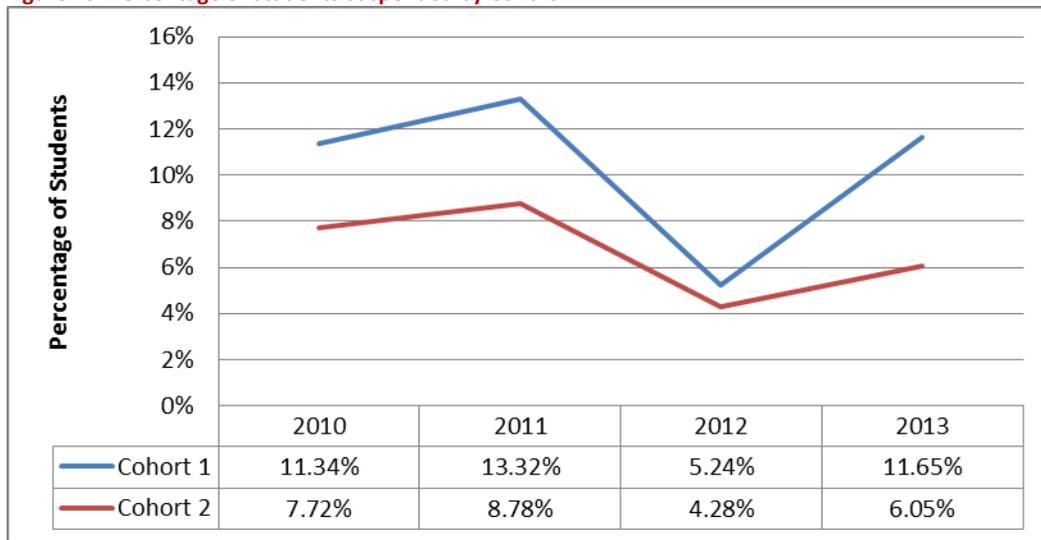


Figure 14: Number of Expulsions by Cohort

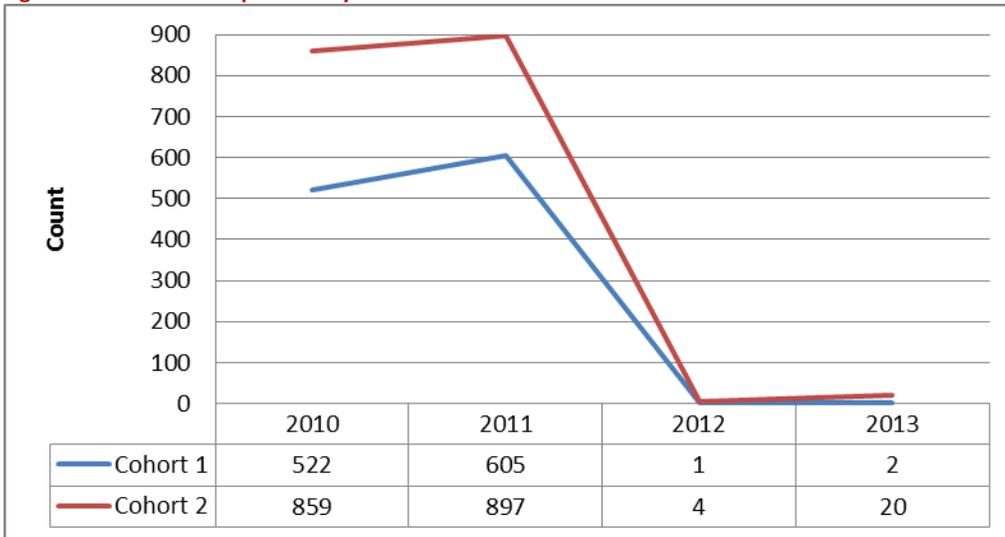
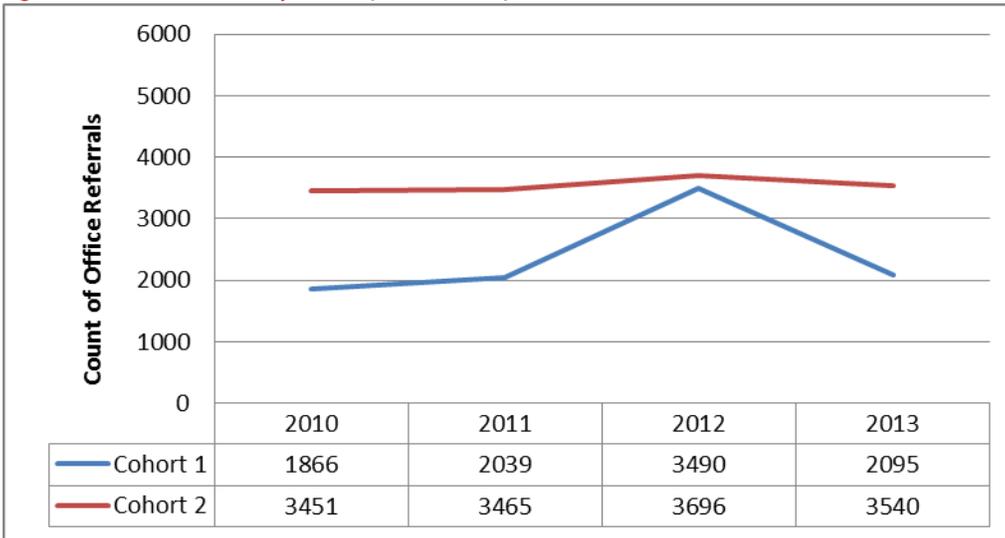


Figure 15: Office Referrals by Cohort (Actual Count)*



Limitations

This study is the first evaluation of TLIM in Cohort 1 and 2 schools. Due to changes in data availability and/or the quality of the data provide, this report is limited to a meta-analysis technique that brings comparative measures in from various sources. Researchers had to choose valid comparison data from districts and state data. In addition, data consistency from year to year proved to be a challenge. In particular, the school year starting in 2011 seemed to have some unusual demographic differences in one school district and 2012 surged with discipline data. However, the evaluators were limited as to what could be done with the data sources provided for analysis, this was particularly problematic when reporting on behavior indicators. School level data

provides evaluators with the flexibility to report outcomes at different levels of granularity.

The lack of an instrument or measurement of program fidelity severely limits the power of this evaluation. One of the hallmarks of program evaluation and effectiveness is implementation of program components as intended. Although TLIM is fairly non-prescriptive, compared to many school based programs, connecting current or future outcomes to TLIM will be difficult without a measure of fidelity. Additionally, evaluations of the program will rely heavily on the quality and level of data provided for measuring program outcomes.

Additional years of this study should give researchers more evidence of an effect of the program with longer trends to rule out alternative causes of behavioral and academic changes. Until then, the reader should be aware that at least some of the results outlined in this report may be the results of other programs or strong leadership in the TLIM schools. Future studies on program effectiveness may want to include qualitative data related to program experiences such as leadership. Although this cannot replace measuring program fidelity, it can add more information related to program effectiveness to the evaluation.

Conclusions

The interpretation of the results in this report should be interpreted with some degree of caution, specifically when reviewing behavior indicators. Student level data would allow a richer analysis of the effects of TLIM particularly on measures of behavior. However, there is reason to believe there is a positive effect of the program as demonstrated by changes in the academic measures of students. TLIM's effect on behavior is inconclusive at this time. The data provided and used for this evaluation on behavior indicators limited the Picard Center's ability to conduct an in-depth analysis and sub-analysis. Additionally, the lack of a measurement of program fidelity restricted this evaluation from correlating program implementation to program effectiveness and/or outcome.

Overall, the results suggest a strong effect on young children (from DIBELS), in Math studies, and for certain traditionally at-risk sub-groups. Consistently larger gains were achieved by TLIM students who are African American, in Poverty, and at lower grades. There seemed to be gender differences but it is unclear if there is a consistent pattern. Clearly, the results indicate a promising program both overall and for demographic subgroups most in need. Behavioral data seems to suggest that student behavior is holding steady and is not yet significantly changed by TLIM program. The academic findings in this evaluation are consistent with a 2012 case study of TLIM conducted by Johns Hopkins University. Interestingly, the case study used qualitative measures to report on behavior.

Recommendations for Future Studies

The following recommendations are suggested for future evaluations of The Leader in Me:

- Districts should provide student level data to allow for the analysis of behaviors required to evaluate program effectiveness. Allocation of future funding should be considered.
- A mixed method or qualitative evaluation/case study is more appropriate for this program. Since no measure of program fidelity has been developed; observations and focus groups can be utilized to tap into implementation. Perceptions and experiences related to TLIM and outcomes could be more robust and a valid representation of the program.
- Qualitative data (i.e. focus groups, observations, etc) should be collected to capture the perceptions and/or experiences of positive changes in behavior as a result of TLIM. This can add contextual information to quantitative changes, found or absent on behavior indicators.
- Explore reporting evaluation outcomes at the school level. This will be a more valid and informative representation of TLIM. School names should be replaced when reporting outcomes. A separate measure for the school district should be created as the number of schools increase.