



*Does High School Matter?*  
**An Analysis of Secondary Outcomes for a  
Cohort of Ninth Graders**

*Authored by: Amy Wiseman, Ph.D.  
Susan Dawson*

January, 2018

---

## Original Plan for Creating Datasets

The KLE Foundation wanted access to a campus level dataset that could be compared to the datasets from the Texas Higher Education Coordinating Board (THECB) that are used to produce their secondary to postsecondary student outcomes metrics. To be comparable, these datasets would need to be cohorts, where groups of students were followed over several years, rather than yearly snapshots. To create these cohorts, access to student level data would be required, thus E3 Alliance utilized our Education Research Center (ERC) access to produce these datasets.

Though the THECB metrics are based on 8<sup>th</sup> grade cohorts, E3 and KLE decided to create 9<sup>th</sup> grade cohorts to track students starting in Central Texas high schools. Separate cohorts were created for 2005, 2006 and 2007 ninth graders in Texas. These cohorts were chosen so that we could compute a 6-year college completion rate 10 years later. Similar to the THECB, we chose to use all students in the cohort, whereas historically when E3 Alliance creates cohorts, they are limited to students who were in 9<sup>th</sup> grade for the first time. These students were followed across ten years, and several different pieces of information about their high school and higher education outcomes were collected over this time period. If a student had a high school outcome at any time during the first seven years, they were counted as having the outcome, and higher education outcomes were collected across the entire ten years. See Appendix for the list of variables and definitions originally included in the dataset.

These outcome results were aggregated at the school level, where any campus with 9<sup>th</sup> graders was included in the school-level dataset. Even though the data were aggregated, they were still not FERPA compliant until small cells (any count from zero to four) were masked. For variables that stand alone and are not directly related to other variables (such as higher education enrollment count), that suffices for masking. However, masking has to account for imputation when there are variables that depend on other variables (such as higher education enrollment count for high school graduates, rather than all 9<sup>th</sup> graders) or when there are variables that add up to other variables (such as low income high school graduate count and non-low income graduate count equals the total graduate count).

The issue of variable dependency rarely occurred in the KLE datasets; where possible key outcomes were based on the count of 9<sup>th</sup> graders, so, for example, the higher education enrollment rate was entirely unrelated to the high school graduation rate. And in any future campus level datasets, this would need to be the case. The datasets cannot be used effectively to track student attrition across an educational pipeline.

In contrast, the issue of counts adding up across variables rendered the first attempt at the datasets nearly useless. About 80 variables were included in an attempt to provide as much detail as possible about campus outcomes, including splitting out results for each outcome by whether students were low income and whether they were mobile or stable. Additionally, more detail was provided for many variables, such as providing counts of postsecondary credential type (certificate, 2 year degree, 4 year degree), along with the total count of credentials. All of these variables could be interesting breakdowns to evaluate at a regional level, where the aggregated counts rarely need masking. However, at a school level, there were many small cells that led to significant imputation such that conclusions are essentially unusable. For example, any time the mobile

group had to be masked for an outcome, the non-mobile group had to be masked as well, even if it had a large count. Or, any time the low income count for an outcome had to be masked, the non-low income had to be masked as well. This led to significant data loss unless we included far fewer variables than originally planned in the analysis.

## Updated Plan for Creating Datasets

Clearly, due to masking/FERPA requirements, it was not going to be possible to provide the level of detail about secondary and postsecondary outcomes we had intended. In consultation with KLE, we simplified the plan for creating the dataset significantly. This led to no more information at all on mobility – now all student outcomes from 10 years later would still be assigned to the campus from 9<sup>th</sup> grade, even if the student changed campuses during the high school years. It also meant no more including both breakdowns and totals for outcomes. Because the question that prompted the dataset creation in the first place was about low income student higher education completion rates, we included counts split out by low income and non-low income students but did not include total counts of 9<sup>th</sup> grade students at campuses.

Ultimately, KLE chose to include 10 pieces of information for each campus for each of the cohorts in the simplified datasets (see Table 1 for data definitions). These are separate counts for low income and non-low income students overall on the campus, and separate counts by income breakdown for high school graduation, high school dropout, higher education enrollment, and higher education credential. Because totals were excluded, if the non-low income count for one variable had to be masked for a given cohort, the low income version of the variable did not need to be masked.

*Table 1. Data definitions used in the simplified dataset*

Number of Low Income students in the 9th grade cohort
Number of Low Income students from the 9th grade cohort that Graduated from any Texas Public High School across 7 years
Number of Low Income students from the 9th grade cohort that Dropped out from any Texas Public High School across 7 years
Number of Low Income students from the 9th grade cohort that Enrolled in any Higher Education (during or after high school) across 10 years
Number of Low Income students from the 9th grade cohort that Attained any Higher Education Credential across 10 years
Number of Non-Low Income students in the 9th grade cohort
Number of Non-Low Income students from the 9th grade cohort that Graduated from any Texas Public High School across 7 years
Number of Non-Low Income students from the 9th grade cohort that Dropped out from any Texas Public High School across 7 years
Number of Non-Low Income students from the 9th grade cohort that Enrolled in any Higher Education (during or after high school) across 10 years
Number of Non-Low Income students from the 9th grade cohort that Attained any Higher Education Credential across 10 years

Note that two standard pieces of information are missing that could theoretically have been included. Traditionally the count of graduates, dropouts and leavers adds up to the total count of students in the cohort. Although in this dataset a student could be coded as belonging to more than one of these groups, many times these did fully add up, and for that reason, KLE decided to leave it out. And higher education persistence is missing. By definition to calculate persistence one must know a student's enrollment status the year prior, so if enrollment is masked, persistence will also be masked. However, KLE decided that completion was the only higher education outcome of interest; though it implies prior enrollment, it is not evaluated based on enrollment.

## Data Preparation for Central Texas Analysis

The Michael and Susan Dell Foundation requested an analysis of outcomes from the school level datasets that had been released from the ERC for KLE. There were still several steps that had to be completed for the data to be ready for analysis. School level variables were added to the datasets to designate whether schools were Alternative Education Campuses (AECs) and to provide geographic location information. The percent low income rate was calculated for each campus for each cohort, and this information was used to designate a school as low income (if 40% or more of 9<sup>th</sup> graders were low income). Additional school level demographics such as ELL status and ethnicity would have been added if there had been time to do an even more in-depth analysis for schools by demographics in addition to income.

An initial check of how many schools had variables masked showed that AECs were usually masked for most if not all variables, such that analysis on the few remaining AECs would be unrepresentative of this class of campuses. Thus the set of schools in the dataset and analysis was limited to the 55% of campuses that were not AEC campuses that accounted for 97% of 9<sup>th</sup> graders.

Most of the non AEC campuses (hereafter: campus) were traditional high schools, but a few were 9<sup>th</sup> grade centers, specifically for Georgetown ISD, Lockhart ISD and for Stony Point in Round Rock ISD. Not limiting the dataset to first time 9<sup>th</sup> graders meant that where students are typically in a special 9<sup>th</sup> grade center, a small set of students were instead found in the high school. Their results suggesting that they were not first time 9<sup>th</sup> graders, so we chose to merge the data from these high schools with their 9<sup>th</sup> grade centers where possible and otherwise remove these campuses from the dataset. This left 52 high school campuses in Central Texas included in this postsecondary outcomes analysis.

## Data Quality and Imputation

Even after removing the AECs, we still found quite a few campuses with masked variables, especially dropout counts (which is good) and higher education completion counts, especially for low income students (which is not). Not only were the rates of masked data high for consideration of data quality, but any attempt to aggregate campuses across time (lumping the three cohorts together) or by demographic characteristic would be inaccurate if these masked values were treated as missing data. And the degree of inaccuracy would vary by how many students were at the campus.

In the most time intensive part of working with the data after it was released from the ERC, multiple methods of imputation were attempted. We ultimately settled on a method that balanced the accuracy of imputing a value, versus leaving it blank, that took into account the count of students (low income, non-low income or total

count), and the number of years at the campus that there were non-masked values for the outcome in question. This method of imputation meant that averaging data across the three cohort years would yield more stable estimates than looking at the data for each cohort individually, but it also meant that more imputation was possible. For the lowest rate of non-masked data from the ERC, higher education credential counts for low income students, prior to imputation, only 63% of the counts across campuses and cohorts were not masked. After imputation and aggregation, 73% of campuses had a higher education credential rate for low income students, and 83% had an overall credential rate.

## Analysis

### *Regional Trends Over Time*

Even though the data were not reliable enough to look school by school at how their 9<sup>th</sup> grade outcome rates varied across the three cohorts, it was possible to assess these trends for the 9<sup>th</sup> graders at a regional level. The average rates by year for Central Texas can be found in Table 2. First we examined whether the campus level percent of low income students at campuses changed over time. Even though E3 Alliance has previously shown increasing low income student rates across the region over time, in this dataset, the average campus level low income rates for 9<sup>th</sup> graders were highly consistent from 2005 to 2007.

*Table 2. Average rates per year for secondary and postsecondary outcomes in Central Texas*

9th Grade Cohort Year	High School Graduation Rate	Dropout Rate	Higher Ed Enrollment Rate	Higher Ed Completion Rate
2005	70%	14%	58%	20%
2006	72%	14%	57%	20%
2007	75%	10%	54%	20%

Next, we examined whether trends could be found across the three years for the outcomes of interest overall. These analyses were conducted two ways: by comparing the raw rates, and by comparing the rates after adjusting for campus 9<sup>th</sup> grade income status. Overall graduation rates showed a numerical increase across the three years that was statistically significant when income was taken into account. Dropout rates were consistent in 2005 and 2006, but dropped in 2007, which was a significant drop both for the raw rates and with the income status adjustment. In contrast, raw higher education enrollment rates did not differ, but when the rates were adjusted for income, unfortunately the enrollment rate was lower in 2007. Lastly, there were no differences in higher education completion rate over time, regardless of whether income status was taken in to account.

### *Relationship of Campus Income Status and Outcomes*

For all remaining analyses, data are averaged across the three cohorts to provide the most accurate imputed values and most reliable results. The list of campuses included in the analysis are in Appendix B. Note that much of the addition of charter schools to the educational landscape in the region has occurred in more recent years,

thus this report is best viewed as being about Central Texas high school campuses as defined from 2004-2005 to 2006-2007, which may or may not generalize to the current set of campuses in Central Texas.

Given the degree to which income status affected the results, we quantified this relationship by correlating the low income rate for 9<sup>th</sup> graders at a campus with the overall outcome rates for those students. What we found is that 65% of why high school graduation and dropout rates vary between campuses can be explained by the campuses' low income rate (based on correlations of -.81). A student's family income has an even stronger relationship with higher education outcomes (correlations of -.94). *The rate of low income 9<sup>th</sup> graders at a campus explains 88% (!) of the variability in higher education enrollment and 89% of the variability in higher education completion for these 9<sup>th</sup> graders.* Basically, this means that knowing about income status at a campus (at least for 9<sup>th</sup> graders) means that you can predict the proportion of those students who will enroll and complete higher education with a very high level of certainty. With more time for analysis, these data could likely be used to create such a prediction model.

#### *Relationship of Geography and Outcomes for Low Income Students*

To better understand why income status is such a strong predictor of outcomes, an analysis was conducted that compared students at campuses inside and outside of the 'Eastern Crescent' around Austin. Because all 7 of the high schools located in the Eastern Crescent are low income schools (where 40% or more of the students at the school are low income), this entire analysis was limited to low income schools. Numerically, students in the Eastern Crescent of the Austin area were worse on all 4 outcomes compared to students at low income schools in the region but not the Crescent (see Table 3 for means). Though these were meaningful differences, only the difference in higher education credential rates rose to the level of statistical significance (likely due to having only 7 high schools in the Eastern Crescent).

*Table 3. Rates of secondary and postsecondary outcomes for low income schools by geography*

Geography of Low Income Schools	Low Income Rate	HS Graduation Rate	HS Dropout Rate	Higher Ed Enrollment Rate	Higher Ed Credential Rate
<b>Schools in Eastern Crescent</b>	69%	56%	23%	39%	7%
<b>Other Low Income Schools</b>	53%	67%	17%	47%	13%

The reason for the difference is nearly entirely due to the fact that the low income rate for low income schools in the crescent is higher than for low income schools not in the Crescent. In fact, controlling for 9<sup>th</sup> grade school level income status entirely removed the differences that had been found for high school graduation, dropout, and completion rates. Schools in the crescent looked no different than the low income schools outside of that area when taking into account relative population of low income students at the campus; for example, the average graduation rates for low income campuses both in and out of the Eastern Crescent was 64%. Interestingly, a statistical difference was found for higher education enrollment, but in the opposite direction – once school level income status was accounted for, students at the schools in the Eastern Crescent had a *higher* enrollment rate than students at low income schools outside the crescent (49% in the Crescent and 43% outside it).

### *Relationship of Student Income Status to School Income Status and Outcomes*

Another factor that may help explain difference in performance by income status across schools is whether students of a particular income status are in the majority or minority at the school. Another way to ask the question is whether there is a protective effect of being a low income student at a primarily non-low income school. Thus we conducted an analysis that evaluated how low income and non-low income students fared in schools where students in their grade were primarily low income (40% of ninth graders or higher) or non-low income.

For all outcomes, students at non-low income schools fared better than at low income schools, and within a school non-low income students fared better than low income students. In fact, at non-low income schools, non-low income students had better outcome rates that ranged from higher education enrollment and completion rates that were 21 and 30 percentage points higher, respectively, than non-low income students at the same schools. Even at low income campuses, non-low income students enrolled in higher education at an 18 percentage point higher rate than low income students.

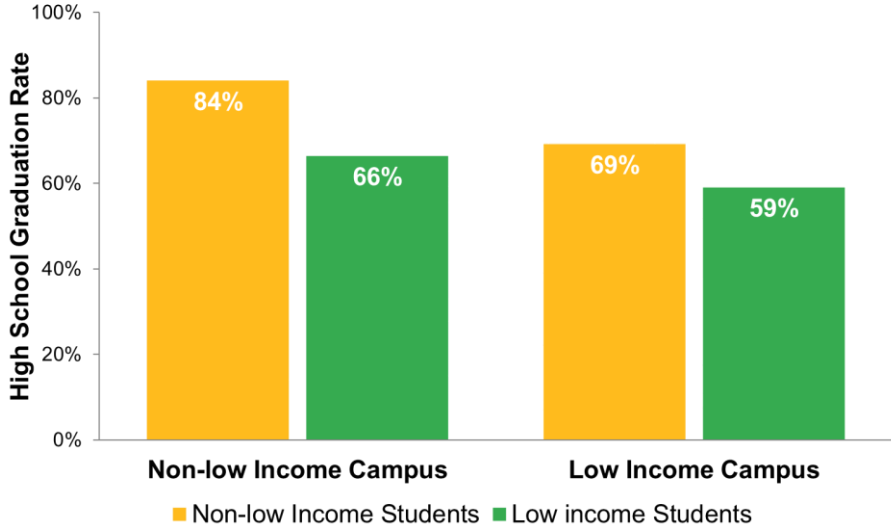
In regards to the question of whether there is a protective effect for low income students when they are in non-low income schools, for all four outcomes, low income students had better outcomes if they were at non-low income campuses as compared to low income campuses. In fact, on average, low income students at non-low income schools had outcome rates that were 7 percentage points better than low income students at low income schools.

Though this result appears to suggest a protective effect for low income students at non-low income schools, it may better reflect the level of poverty of low income students at these schools. Low income students at non-low income schools are likely less poor than low income students at low income schools, given the degree of economic geographical segregation in the Austin area. It is likely that low income students at Bowie are less poor on average than low income students at Del Valle High School, for example. The way that students' family income status is measured – as a simple threshold, does not allow us to differentiate between degrees of poverty.

A more stringent way of assessing whether school income status was actually protective is to see if low income students at non-low income schools perform at least as well statistically as non-low income students at low income schools, because in this case the comparison group is has the same income status or higher income status. In other words, if we can show that students that fall at least just below the income eligibility cutoff (i.e. low income students) at non-low income schools fare as well statistically as students who are at least on the higher side of that eligibility cutoff (i.e., non-low income students) at low income schools, then it really looks like a protective effect.

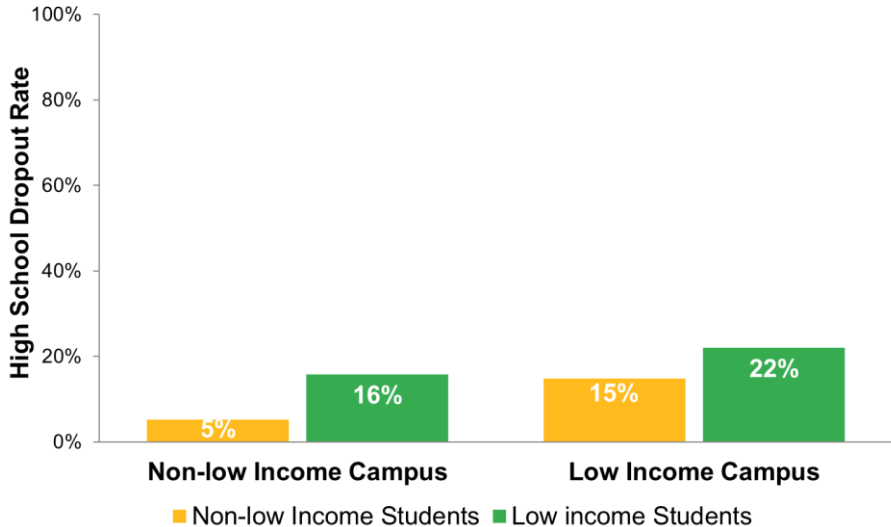
The interesting story emerges for high school graduation rates (see Figure 1), when comparing low income students at non-low income schools (66% graduated) with non-low income students at low income schools (69% graduated). These rates are not statistically different, suggesting a protective effect of being at a non-low income school. Additionally, non-low income students in non-low income schools have much higher graduation rates than the other three groups.

Figure 1. High School Graduation Rates by Campus and Student Income Status



The story for dropout rates is the same as for graduation rates, with the data showing the inverse pattern because non-low income students at non-low income schools showed the lowest dropout rates, as expected (see Figure 2). Again, was no statistical difference between low income students at non-low income schools and non-low income students at low income schools.

Figure 2. High School Dropout Rates by Campus and Student Income Status



Thus for high school outcomes, being in a non-low income school appears to provide a protective effect for students of all incomes. However, this interpretation should be viewed with caution because of the arbitrary cut-off of “low income” based on federal free and reduced lunch status, which is based on family income being 135%, or 185%, respectively, of the federal poverty level. Previous analyses by E3 Alliance have found large



differences in outcomes of reduced lunch versus free lunch students; in fact, for some outcomes reduced-lunch students are more similar to their non-low income peers than to free-lunch students. As previously discussed, low income students at non-low income schools may be generally better off than most low income students at low income schools. In fact, low income students at the non-low income school may be similar in income status to the non-low income students at the low income school, and so findings by income are largely the result of students falling on the opposite side of an arbitrary income cutoff.

For both higher education enrollment and completion (see Figures 3 and 4, respectively), again, non-low income students in non-low income schools have much higher outcomes than the other groups. But for these higher education outcomes, there does not appear to be a protective effect for low income students at non-low income schools, in that these students have numerically and statistically lower enrollment and completion rates than do non-low income students in low income schools: family income is a larger determinant of postsecondary outcomes because of affordability.

*Figure 3. Higher Education Enrollment Rates by Campus and Student Income Status*

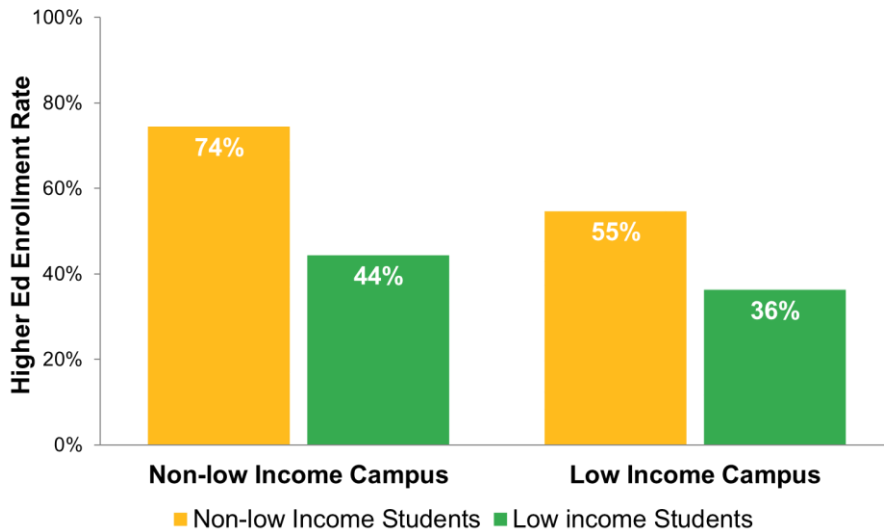
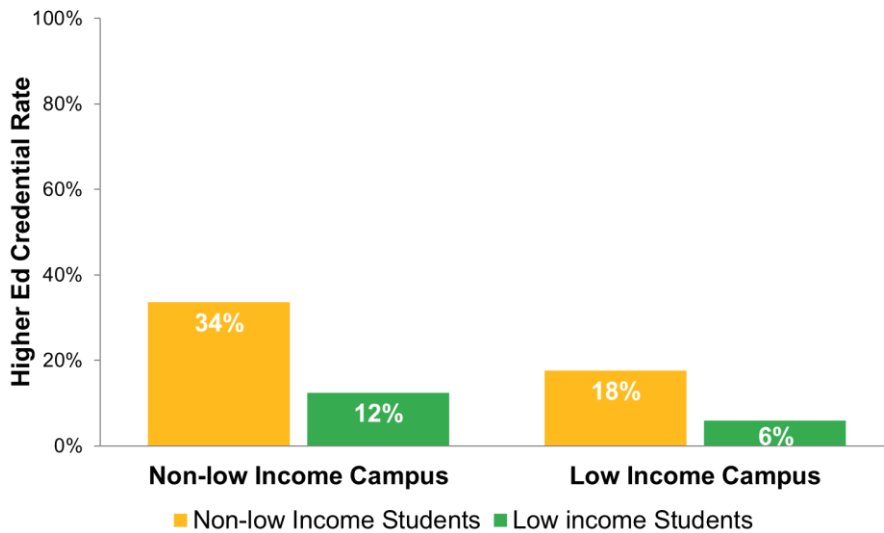


Figure 4. Higher Education Credential Rates by Campus and Student Income Status



## Conclusions

Masked school-level datasets provide unique challenges for data quality that are ameliorated enough to allow for analysis. These challenges have been partially overcome by limiting the outcomes reported in the datasets and utilizing three cohorts of data for reporting campus outcomes.

Overall conclusions that can be drawn from the 2005-2006 to 2006-2007 cohorts for this analysis include:

- Across the cohorts studied, high school graduation rates increased slightly, but college enrollment rates did not significantly change.
- Family income is a very strong driver of high school outcomes and a particularly strong driver of higher education outcomes.
- Students graduating from high schools in the 'Eastern Crescent' have poorer outcomes than students from other low income schools in the region, but these outcomes appear to be explained by the relatively greater population of low income students at those schools.
- Both campus income status and student income status have effects on these outcomes, whereby non-low income students at non-low income schools show far and away the best outcomes.
- Being in a non-low income school shows protective effects for low income students, at least for high school outcomes.
- Low income students coming from non-low income schools do not appear to have the same protective effect in higher education - their higher education credentialing outcomes do not statistically differ from their low income peers from poorer schools.

*The conclusions of this research do not necessarily reflect the opinions or official position of the Texas Education Agency, the Texas Higher Education Coordinating Board, or the State of Texas.*

## Appendix A

The definitions of variables that were intended to be in the original datasets for KLE are included here. Each variable, it is the count of students at each campus that meet the variable description. So for example, 'Dropped out from any campus' would be the number of students who were at a particular campus in 9<sup>th</sup> grade who were dropouts from any Texas school in that year or at any time in the next 6 years (7 years total).

Variable Definitions
Campus CDC ID
No follow-up data on graduation, dropout, leaver
High school graduate from any campus
Received GED
Had a leaver code not reflecting graduation, dropout, or GED
Dropped out from any campus
Graduated from same campus as their 9th grade snapshot campus
Dropped out from same campus as their 9th grade campus
Graduated from different campus than their 9th grade snapshot campus
Dropped out from different campus than their 9th grade campus
Graduated from any campus in 4 years or less
No follow-up data on graduation, dropout, leaver or received GED
Show any college enrollment during or after high school
Show any college enrollment during high school (all refers to info from TEA or NSC data)
Show any college enrollment after high school
Record of college enrollment for those students graduating from their 9th grade campus is in NSC data
Graduated from snapshot campus in 4 years or less
Graduated from non-snapshot campus in 4 years or less
Graduated from snapshot campus and enrolled in college during high school
Graduated from non-snapshot campus and enrolled in college during high school
Graduated from different campus than their 9th grade snapshot campus and enrolled in college during or after high school
Graduated from same campus than their 9th grade snapshot campus and enrolled in college during or after high school
Graduated from non-snapshot campus and enrolled in college after high school
Graduated from snapshot campus and received bachelors or advanced degree
Graduated from snapshot campus and received certificate or completer
Graduated from snapshot campus and received associates degree
Graduated from non-snapshot campus and received bachelors or advanced degree
Graduated from non-snapshot campus and received certificate or completer
Graduated from non-snapshot campus and received associates degree
Total number of associates degrees earned as highest degree
Total number of bachelors or advanced degrees earned as highest degree

Total number of certificates/completes earned as highest degree
Total number of unknown degrees earned as highest degree (unknown reflects data from NSC that doesn't specify degree attained)
Total number in cohort who attained any type of postsecondary credential
Total number who graduated from snapshot campus who attained any type of postsecondary credential
Total number who graduated from non-snapshot campus who attained any type of postsecondary credential
Total number in cohort who graduated or dropped out of snapshot campus
Total number in cohort who graduated or dropped out of non-snapshot campus
Low income high school graduate from any campus
Non-low income high school graduate from any campus
Low income and had a leaver code not reflecting graduation, dropout, or GED
Non-low income and had a leaver code not reflecting graduation, dropout, or GED
Low income high school dropout from any campus
Non-low income high school dropout from any campus
Low income students with any college enrollment during or after high school
Non-low income students with any college enrollment during or after high school
Low income students who graduated from same campus than their 9th grade snapshot campus and enrolled in college during or after high school
Non-low income students who graduated from same campus than their 9th grade snapshot campus and enrolled in college during or after high school
Low income students who graduated from diff campus than their 9th grade snapshot campus and enrolled in college during or after high school
Non-low income students who graduated from diff campus than their 9th grade snapshot campus and enrolled in college during or after high school
Low income student receiving any higher education credential
Non-low income student receiving any higher education credential
Low income student graduating from snapshot campus receiving any higher education credential
Non-low income student graduating from snapshot campus receiving any higher education credential
Low income student graduating from non-snapshot campus receiving any higher education credential
Non-low income student graduating from non-snapshot campus receiving any higher education credential
Low income student receiving bachelors or advanced degree as highest
Non-low income student receiving bachelors or advanced degree as highest
Low income student receiving associates/certificate/completer as highest
Non-low income student receiving associates/certificate/completer as highest
Number of low income students who graduated or dropped out from snapshot campus

Number of non-low income students who graduated or dropped out from snapshot campus

Number of low income students who graduated or dropped out from non-snapshot campus

Number of non-low income students who graduated or dropped out from non-snapshot campus

## Appendix B

Campuses included in secondary and postsecondary outcome analyses.

Campus	District	Campus Low Income
		Rate
Akins H S	Austin ISD	60%
Anderson H S	Austin ISD	22%
Austin H S	Austin ISD	36%
Bowie H S	Austin ISD	10%
Crockett H S	Austin ISD	58%
L B Johnson H S*	Austin ISD	57%
Johnston H S*	Austin ISD	85%
Lanier H S*	Austin ISD	82%
McCallum H S	Austin ISD	41%
Reagan H S*	Austin ISD	84%
Travis H S	Austin ISD	84%
Bastrop H S	Bastrop ISD	49%
Blanco H S	Blanco ISD	43%
Harmony Science Academy*	Charter	40%
Katherine Anne Porter School	Charter	43%
NYOS Charter School	Charter	14%
Star Charter School	Charter	7%
Texas Empowerment Academy*	Charter	44%
Del Valle H S*	Del Valle ISD	67%
Dripping Springs H S	Dripping Springs ISD	11%
Westlake H S	Eanes ISD	3%
Elgin H S	Elgin ISD	53%
Florence H S	Florence ISD	38%
Georgetown 9th Grade	Georgetown ISD	30%
Granger School	Granger ISD	38%
Jack C Hays H S	Hays CISD	30%
Lehman H S	Hays CISD	48%
Hutto H S	Hutto ISD	25%
Jarrell H S	Jarrell ISD	41%
Lyndon B Johnson HS	Johnson City ISD	28%
Lago Vista H S	Lago Vista ISD	10%
Lake Travis H S	Lake Travis ISD	14%
Cedar Park H S	Leander ISD	11%

<b>Leander H S</b>	Leander ISD	32%
<b>Vista Ridge H S</b>	Leander ISD	19%
<b>Liberty Hill H S</b>	Liberty Hill ISD	21%
<b>Lockhart High School Freshman Campus</b>	Lockhart ISD	49%
<b>Luling H S</b>	Luling ISD	56%
<b>Manor H S*</b>	Manor ISD	66%
<b>Hendrickson H S</b>	Pflugerville ISD	30%
<b>John B Connally H S</b>	Pflugerville ISD	47%
<b>Pflugerville H S*</b>	Pflugerville ISD	29%
<b>Prairie Lea School</b>	Prairie Lea ISD	66%
<b>McNeil H S</b>	Round Rock ISD	17%
<b>Round Rock H S</b>	Round Rock ISD	29%
<b>Stony Point Ninth Grade Campus</b>	Round Rock ISD	37%
<b>Westwood H S</b>	Round Rock ISD	10%
<b>San Marcos H S</b>	San Marcos CISD	59%
<b>Smithville H S</b>	Smithville ISD	44%
<b>Taylor H S</b>	Taylor ISD	57%
<b>Thrall H S</b>	Thrall ISD	22%
<b>Wimberley H S</b>	Wimberley ISD	14%

\*in Eastern Crescent

Note: campuses that opened in 2007-2008 or more recently are not included. Thus, what is currently Eastside Memorial HS is listed here at Johnston HS, and LASA did not exist as a separate campus from LBJ HS. There were fewer charter schools open, and the list did not include charter districts such as Harmony, KIPP and IDEA.