

Analysis of Student Retention at UTK

Prepared by The Faculty Senate Budget & Panning Committee, Scott Gilpatric, Chair

Overview

This document describes the key findings of an analysis of the retention and graduation rates among UT undergraduates. OIRA provided data for 2005 to 2011 cohorts from which we estimated how student characteristics are associated with retention from first to second year at UT and with graduating within six years. These data include measures of student academic performance (core high school GPA and ACT/SAT scores), race and gender, family income, county of permanent residence for Tennessee students, and receipt of various scholarships. We focus on retention of Tennessee students from the 2008-11 cohorts. Data on core GPA is capped at 4.0 prior to the 2008 cohort. Since we find that the additional variation above 4.0 is highly predictive of retention we focus on the cohorts where this is available (six year graduation rates of course cannot yet be studied for these cohorts). Our attention is restricted to Tennessee students because we do not have information on county of permanent residence for out-of-states students (who are also likely to be a systematically different population in many respects). Key findings include:

- For cohorts 08-11, for whom uncapped GPA is available, a higher high school GPA has a strong positive effect on the probability of retention which persists throughout the observed range, whereas the effect of a higher ACT score is relatively small.
- Higher family income has a strong positive effect on the probability of retention and this persists into the highest category (i.e. retention of students from families with income above \$200,000 is significantly higher than for students from even those families in the \$150,000-199,999 range).
- Controlling for other factors, Asian and black students at UT have a higher retention probability than white students; Hispanic students have lower probability.
- Distance of permanent address from UTK is *not* significantly associated with reduced retention.
- Controlling for other factors, students from urban counties (defined as counties having a population over 100,000) have higher retention probability than students from rural counties

Methodology

We employ simple Probit regressions that estimate the effect of various student characteristics on the probability of a student's retention from first to second year and on the probability of graduating within six years. The Probit model is appropriate for this data in which the variable to be explained is binary (either a student returns/graduates or does not) and it is reasonable to assume that unobservable factors impacting the decision follow a normal distribution. Complete regression results are reported in Appendix A for four specifications: 1) Retention for cohorts 08-11 utilizing uncapped core GPA scores; 2) Retention for cohorts 05-11 utilizing core GPA scores capped at 4.0; 3) Graduation within six years for cohorts 05 and 06 utilizing capped core GPA scores; 4) Retention for cohorts 05 and 06 utilizing capped core GPA scores. Specifications (3) and (4) together indicate that the effect of student characteristics on retention is generally very similar to the effect on graduation rates. For this reason we focus our analysis on the results from the retention specification (1) employing uncapped core GPA.

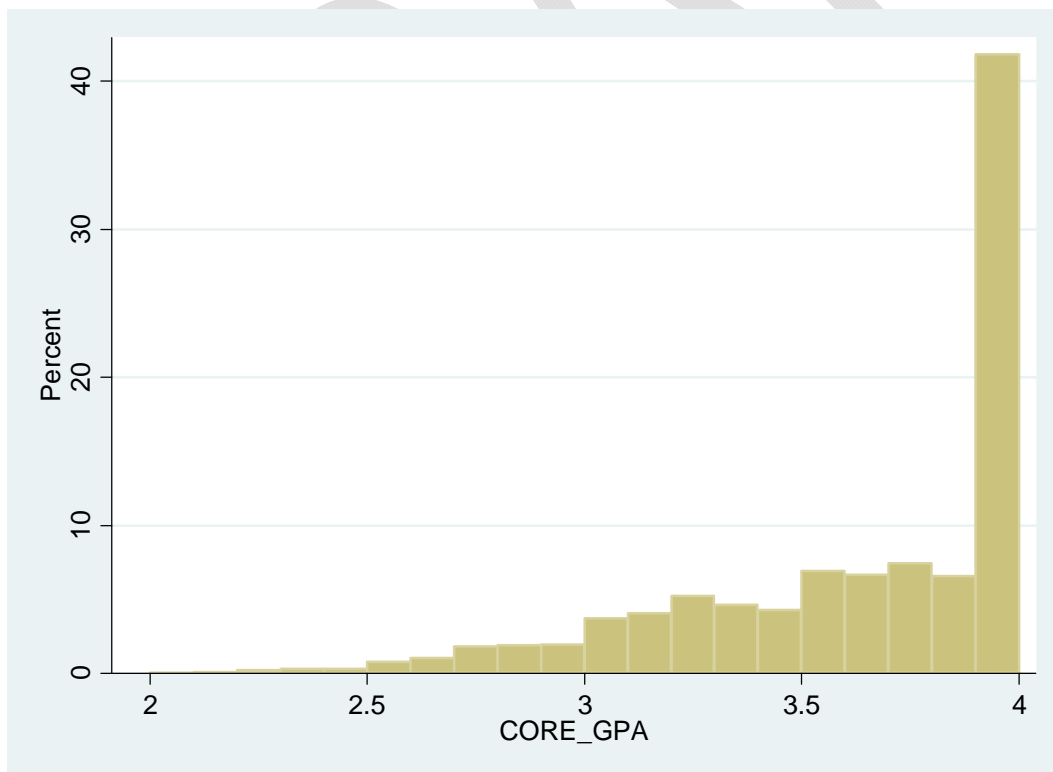
All specifications include the following right-hand-side variables. Academic preparation is captured by core GPA and composite ACT scores. Racial background is captured with indicator variables for Asian, Black, Hispanic, Multi-race, or None (race not reported), with White as the omitted category. Gender is captured with an indicator for Female. Household income is reported in the data as belonging to one of six categories: below \$50K, \$50-100K, \$100-150K, \$150-200K, more than \$200K, or not available. We include indicator variables for each category except the omitted category which is income over \$200K. The data record whether each student received upon entering each of four financial scholarships: Hope, Pell, Pledge, and Promise. We include indicators for each of these. Utilizing the information on each student's county of permanent residence we generated two additional variables. First, an indicator for being a resident of a county which is significantly urban, defined as having a population over 100,000. Second, we employ the U.S. Census population-weighted centroid of each county to identify the distance from this point to UTK. We include in the regressions the natural log of this distance. Finally, the regressions also include cohort fixed effects.

Findings

Core GPA

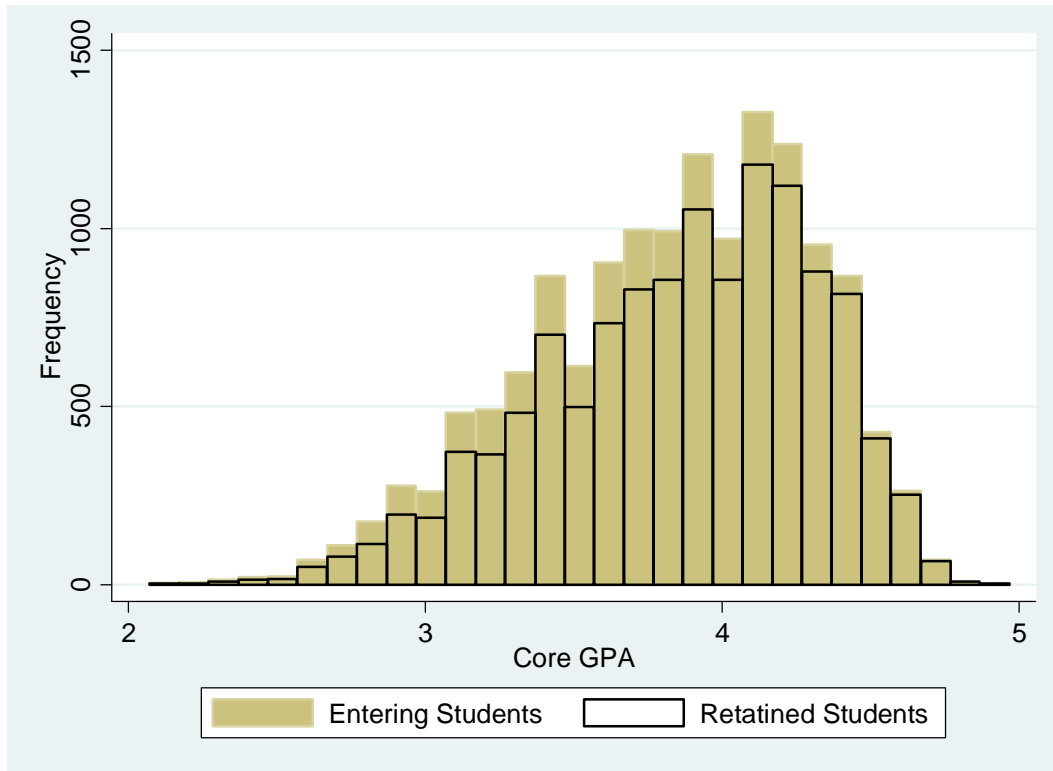
It is illustrative to first look at the distribution of core GPA within the population. First, note the censoring effect of having GPA capped at 4.0, which is the only measure available for cohorts prior to 2008.

Histogram of Capped Core GPA, 2005-2011 cohorts



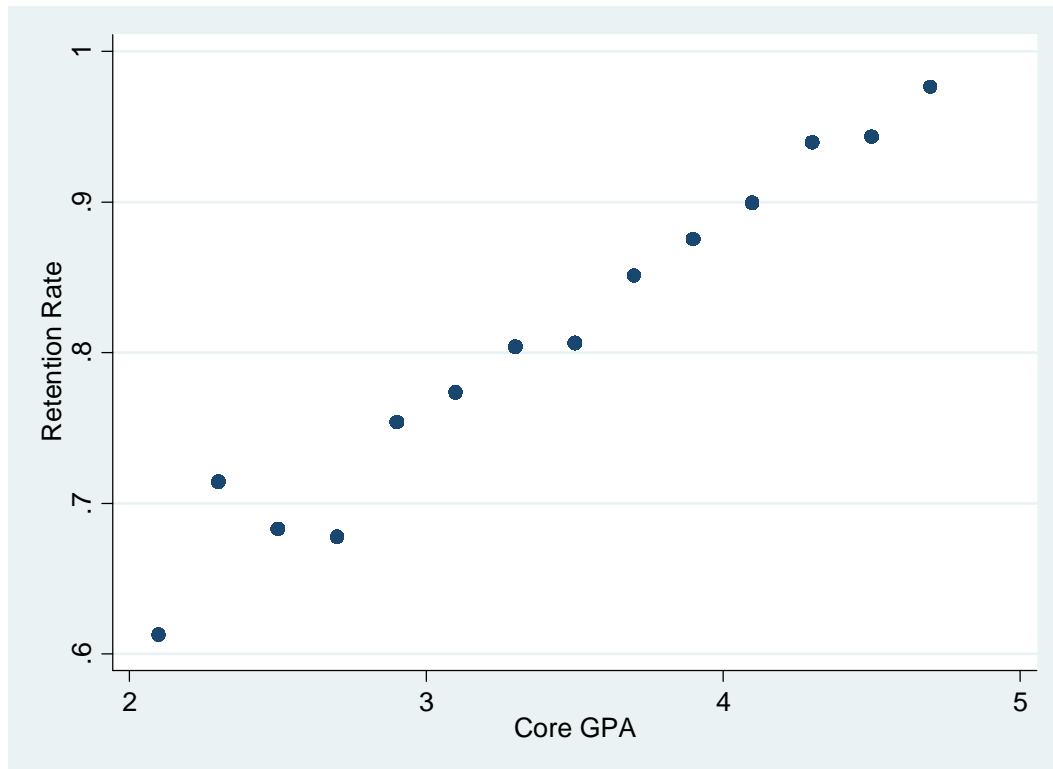
The following figure shows the distribution of uncapped core GPA for admitted students and for those who return in their second year. This figure illustrates that a significant number of students do not return after their first year across nearly the full range of GPA scores. A much larger share of those with low GPAs fail to return, but the largest absolute numbers of non-returning students are found in the area of the mode of the distribution around 4.0.

Histogram of Uncapped Core GPA for entering students and retained students, 2008-2011 cohorts



The graph below simply plots the retention rate for students as a function for their core GPA (not controlling for any other factors). Note that the overall retention rate in the sample is 85%.

Graph of Retention Rate vs. Core GPA

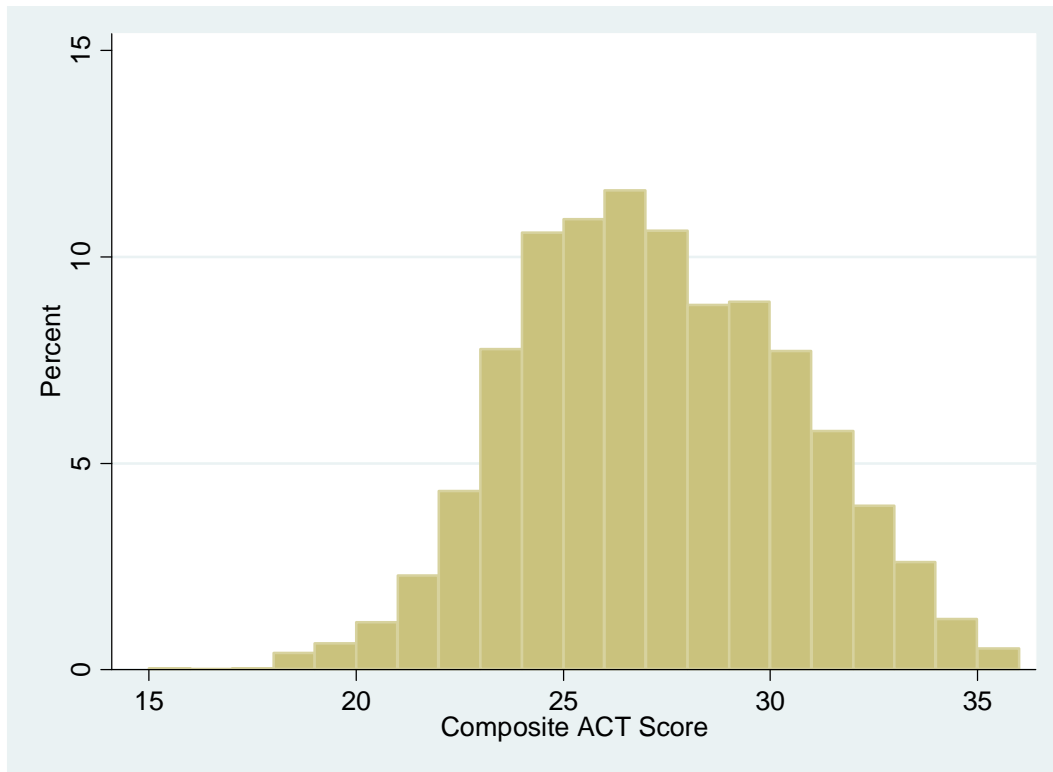


The strong positive relationship between core GPA and retention is unsurprising, but what is rather remarkable is how linear and persistent this relationship appears. One could hypothesize that for students above some sufficiently high level of academic preparedness factors other than academic readiness and success drive retention, which would create a plateau with differences in Core GPA above some level not being associated with differences in retention. We do not observe a plateau occurring at any level. Our regression results confirm what is apparent in this figure, which is that the effect of a higher Core GPA on retention is similar whether moving from 3.5 to 4.0 or 4.0 to 4.5.

The marginal effect of a one point increase in core GPA is estimated to be .126 (or an increase of 12.6 percentage points in retention probability), which is very statistically significant with a 95% confidence interval of .112 to .140. The standard deviation of core GPA in the sample is .46, and a one point increase of course represents a very significant difference. A one standard deviation increase in core GPA is estimated to increase the retention probability by 5.8 percentage points.

ACT Scores

Histogram of Composite ACT Scores, 08-11 cohorts



One of the most surprising findings of our analysis is the weak effect and significance of composite ACT scores on retention when the uncapped core GPA is also included in the regression. **We find that the marginal effect of a one point increase in a student's composite ACT score is .0022, with a 95% confidence interval of .0001 to .0042.** The standard deviation of composite ACT score in the sample 3.30, so a one standard deviation increase raises the retention probability by .0072, or less than one percentage point. The results indicate that composite ACT scores have little power in predicting retention when controlling for (uncapped) core GPA.

Race and Gender

Racial Mix of Sample

Asian	2.56%
Black	7.92%
Hispanic	2.3%
Multi-race	2.4%
Not Reported	0.8%
White	83.7%

Relative to the white population we find a statistically significant positive association with retention for Asian and Black students, and negative association for Hispanic students. The estimated marginal effects relative to white students (the omitted category in the specification) are as follows:

	Marginal Effect	95% Confidence Interval
Asian	.037	.004, .070
Black	.031	.009, .054
Hispanic	-.044	-.084, -.003

Note that these estimated marginal effects are controlling for all other factors. For example, *ceteris paribus*, Blacks students are estimated to have approximately 3 percentage points higher retention probability despite the fact that the simple overall retention rate for black students is nearly 3 percentage point *lower* than the retention rate for the full sample (82.3% vs. 85%). This is because Black students on average have other characteristics associated with reduced retention probability.

In our sample 49.2% of students are female, and the estimated marginal effect on retention of this indicator is .017 with a 95% confidence interval of .006 to .029. The estimated difference in retention probability between males and females controlling for other characteristics is therefore statistically significant, but small relative to other factors such as race, income and urban status.

Family Income

Distribution of Family Adjusted Gross Income in Sample and Effects on Retention

AGI Category	Share of Sample	Marginal Effect	95% Confidence Interval
Less than \$50,000	25.3%	-.089	-.123, -.0544
\$50,000-99,999	28.0%	-.087	-.111, -.063
\$100,000-149,999	21.9%	-.038	-.061, -.013
\$150,000-199,999	10.2%	-.039	-.068, -.010
More than \$200,000	13.4%	-	-
Not Reported	1.1%	.045	-.006, .097

Relative to the top income group of students from families with AGI greater than \$200,000, all lower categories of reported income are associated with a statistically significantly lower probability of student

retention. The effects are quite large. Students from families with AGI below \$50,000, or \$50,000 to \$99,999, are affected very similarly with the probability of retention nearly 9 percentage points below students from the top income group. The negative marginal effect is approximately half as great for students in both the \$100,000 to \$149,999 range, and the \$150,000 to \$200,000 range. Thus the data indicate that income is a significant factor in retention not only for students from low income families, but throughout the distribution. Controlling for other factors, even students from fairly affluent families have a significantly lower probability of retention than those from very affluent families.

Scholarships

Share of Sample Receiving Various Scholarships and Effects on Retention

Scholarship	Share of Sample	Marginal Effect	95% Confidence Interval
Pledge	11.9%	.007	-.013, .027
Promise	4.1%	-.042	-.077, -.005
Hope	97.6%	.144	.083, .205
Pell	26.7%	-.028	-.051, -.004

Federal Pell grants are awarded to students with exceptional financial need. Receipt of a Pell grant is associated with a negative effect on a student's retention probability.

Any student who is admitted and enrolls at the University of Tennessee from one of the 32 eligible high schools in Tennessee will receive the Promise scholarship valued at up to \$7,382 per year plus a \$1,200 award for other educational expenses. The list of Promise high schools is in Appendix B, and consists of schools serving low-income, mainly urban, populations. As reported above, receipt of this scholarship is associated with a negative effect of 4.1 percentage points on the retention probability.

It is important to note that the negative effects associated with these need-based financial awards do not imply that receiving the award makes a student less likely to be retained than if he or she had enrolled without receiving the scholarship. Rather, the negative effect almost certainly arises because the eligibility criteria for the awards captures students for whom obtaining a college education is unusually challenging financially and quite possibly in other respects as well. This is true despite the fact that the regressions include family adjusted gross income because these income categories likely do not fully control for differences in student and family circumstances that may determine both scholarship eligibility and retention.

Unlike the need-based scholarships, the estimated effect of receiving the Hope scholarship (also known as the Lottery scholarship) is estimated to be significantly positive. Hope scholarship eligibility is primarily determined by receiving a high school GPA of 3.0 or above or ACT score of 21 or above (a standard which nearly all students admitted to UTK meet). Students must also complete the Free Application for Federal Student Aid (FAFSA). Note that only 2.4% of students in the sample *did not* receive the Hope Scholarship and the large positive estimated effect on retention is best interpreted as an estimated negative association with the small number of students who failed to obtain this scholarship.

Distance and Urban County

Utilizing the information on each student's county of permanent residence we generated two variables. First, an indicator for being a resident of a county which is significantly urban, defined as having a population over 100,000. The list of these counties is in Appendix C. In our sample 74.0% of students are designated as urban. Second, we employ the U.S. Census population-weighted centroid of each county to identify the distance from this point to UTK.

Residency in an urban county has a positive marginal effect on retention of .045, with a 95% confidence interval of .032 to .059. This is large effect. Put differently, controlling for other factors, a student from a rural (non-urban) county is estimated to be 4.5 percentage points less likely to be retained than if the student is from an urban county, an effect approximately half as large as belonging to the lowest income category.

Our specifications include the natural log of the distance from a student's county of residence to UTK, which was found to be a better fit than entering distance directly. Although the marginal effect is estimated to be negative as expected, the effect is not found to be statistically significant at the 90% confidence level. The effect was found to be statistically significant at this level in some of our other specifications that employed capped GPA. Nevertheless, there is not strong evidence that distance from a student's permanent residence to UTK is associated with a reduced retention probability.

Appendix A: Estimated Average Marginal Effects from Probit Regressions

	Specification 1	Specification 2	Specification 3	Specification 4
Core GPA	.1265087 (.0068954)	.140487 (.0059032)	.2376895 (.0130983)	.1343462 (.0103621)
Composite ACT	.0022007 (.0010698)	.0037034 (.0007798)	.0040425 (.0018608)	.003286 (.0014895)
Female	.0174783 (.0058889)	.0130643 (.0045247)	.0325835 (.011047)	.0062548 (.0087354)
Native American	.044834 (.0460048)	.0425514 (.0119292)	-.1980958 (.0852723)	-.1461001 (.0590669)
Asian	.0371037 (.0167341)	.0378941 (.007957)	.0741817 (.0326236)	.0365988 (.0261348)
Black	.0315562 (.011219)	-.0434568 (.0173139)	.00535 (.0201212)	.0512051 (.016321)
Hispanic	-.0441857 (.0207645)	-.035274 (.0208625)	-.0366321 (.0438591)	-.026439 (.0332162)
Multi-Race	-.0352746 (.0205547)	.0106518 (.0225987)	(omitted)	(omitted)
Race Not Reported	.0121854 (.028825)	.0425514 (.0119292)	-.069194 (.0777448)	-.0224293 (.058466)
AGI below 50,000	-.088775 (.0175056)	-.0879109 (.012775)	-.1079789 (.0247946)	-.0698824 (.0196297)
AGI 50,000 to 99,999	-.0871896 (.0122076)	-.0786496 (.0092501)	-.0873487 (.0197491)	-.0539804 (.0160684)
AGI 100,000 to 149,999	-.0376088 (.0121658)	-.0277093 (.0114702)	-.0300681 (.0208694)	-.0224293 (.058466)
AGI 150,000 to 199,999	-.0391327 (.0148564)	-.0349777 (.0093822)	.0121076 (.0265063)	-.0064555 (.0214801)
AGI not reported	.0456784 (.0266863)	.0030104 (.0222782)	-.0710926 (.0474802)	-.0246665 (.0355573)
Pledge	.0072744 (.0104968)	.0044937 (.0085483)	-.0372627 (.0250859)	-.0113662 (.0188032)
Promise	-.0415288 (.0184631)	-.0467274 (.0154954)	(omitted)	(omitted)
Hope	.144402 (.0312408)	.0843894 (.0182265)	.0818616 (.0296528)	.0411357 (.0221706)
Pell	-.0282961 (.0119656)	-.0314265 (.0092472)	-.0514959 (.0219116)	-.0297631 (.0166106)
Ln(distance)	-.0051295 (.0033579)	-.0058528 (.0025547)	-.0004787 (.0062342)	-.0060215 (.0048674)
Urban County of Residence	.045321 (.0071714)	.0522183 (.0054619)	.0797868 (.0120095)	.048719 (.0093118)
Number of Observations	14241	25039	7127	7127
Pseudo R-Squared	.0688	.0637	.0755	.060

Notes: Robust standard errors are reported in parentheses. All regressions include a constant and cohort fixed-effects which are not reported. The omitted racial category is White, and effects of racial indicators are relative to this population. Similarly, the omitted income category is AGI of \$200,000 or above, and coefficients on other categories are effects relative to this population.

The specifications are as follows:

- 1) Retention for cohorts 08-11 utilizing uncapped core GPA scores
- 2) Retention for cohorts 05-11 utilizing core GPA scores capped at 4.0
- 3) Graduation within six years for cohorts 05 and 06 utilizing capped core GPA scores
- 4) Retention for cohorts 05 and 06 utilizing capped core GPA scores

DRAFT

Appendix B: Promise-Eligible High Schools

School	City
Antioch High School	Antioch
Austin East High / Magnet	Knoxville
B T Washington High School	Memphis
Brainerd High School	Chattanooga
Carver High School	Memphis
Central High School	Memphis
East High School	Memphis
East Ridge High School	Chattanooga
Fairley High School	Memphis
Fayette Ware Comprehensive High School	Somerville
Frayser Middle/ High School	Memphis
Fulton High School	Knoxville
Glencliff Comp High School	Nashville
Hamilton High School	Memphis
Howard School Of Academics Technology	Chattanooga
Kingsbury Middle/ High School	Memphis
Kirby High School	Memphis
Manassas High School	Memphis
Maplewood Comp High School	Nashville
Melrose High School	Memphis
Mitchell Middle/ High School	Memphis
Northside High School	Memphis
Oakhaven Middle/ High School	Memphis

Pearl Cohn Magnet High School	Nashville
Raleigh Egypt High School	Memphis
Sheffield High School	Memphis
Stratford Comp High School	Nashville
Trezevant High School	Memphis
Tyner Academy	Chattanooga
Westwood Middle/ High School	Memphis
Whites Creek Comp High School	Whites Creek
Wooddale High School	Memphis

DRAFT

Appendix C: Counties Designated as Urban, having population of 100,000 or more

Nashville Area

Davidson, Rutherford, Williamson, Sumner, Wilson

Knoxville Area

Knox, Blount

Memphis Area

Shelby

Chattanooga Area

Hamilton

Bristol Area

Sullivan

Johnson City Area

Washington

Clarksville Area

Montgomery

DRAFT