DISCUSSION PAPER SERIES

IZA DP No. 3265

Why Are Hispanic and African-American Dropout Rates So High?

Magnus Lofstrom

December 2007

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Why Are Hispanic and African-American Dropout Rates So High?

Magnus Lofstrom
University of Texas at Dallas
and IZA

Discussion Paper No. 3265
December 2007

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 3265
December 2007

ABSTRACT

Why Are Hispanic and African-American Dropout Rates So High?

The proportion of students who do not graduate from high school is dramatically higher among the two largest minority groups, Hispanics and African-Americans, compared to non-Hispanic whites. In this paper we utilize unique student-level data from the Texas Schools Microdata Panel (TSMP) in an attempt to determine what factors contribute to the higher minority dropout rates. We show that poverty is a key contributor. Lack of English proficiency among Hispanic student is linked to the higher Hispanic dropout probability. Our results also suggest that neighborhood characteristics may be important in explaining the high African-American dropout rates. We also address the issue of the surprisingly low official dropout rates reported by the Texas Education Agency (TEA) and show that the GED program explains some of the discrepancy.

JEL Classification: I21

Keywords: dropout rate, educational attainment

Corresponding author:

Magnus Lofstrom
University of Texas at Dallas
School of Economic, Political & Policy Sciences
P.O. Box 830688, GR 31
Richardson, TX 75083-0688
USA
E-mail: lofstrom@utdallas.edu

1. Introduction

A substantial proportion of students do not complete high school, a problem which is particularly pronounced among the two largest minority groups in the U.S. - Hispanics and African-Americans. Greene (2001) report that only 54 percent of Hispanic students graduate high school and only a slightly higher percentage of African-Americans, 56 percent, complete high school. Amongst white students, the graduation rate is significantly higher, 78 percent. Hispanic and African-American students in Texas are also substantially more likely to drop out of school than white students. Using unique longitudinal student data for the class of 1999 in Texas, we estimate that 40.3 percent of Hispanic, 38.5 percent of African-American and 26 percent of white public school students fail to graduate high school by the age of 20.

There are a number of reasons motivating a study of dropout rates, and particularly the high minority dropout rates. Chances for economic success among individuals who lack a high school diploma appear to be less today than at any other point in U.S. Evidence of this is widely reported in both media and academic journals. For example, Snower (1999) states that "since the mid-1970s in the U.S., the earnings of the less educated have fallen rapidly behind those of the more educated" (p. 4). The increasing importance of skills and education in the economy are apparent for economic outcomes such as employment and earnings. Perreira, Mullan and Lee (2006) report that in 2000 high school dropouts are almost twice as likely to be unemployed compared to high school graduates. Data from the 2000 U.S. Census also clearly show that high school dropouts fare relatively poorly in the labor market. Average annual earnings among male dropouts between the ages of 25 and 65 was $26,400 while male high school

graduates on average earned close to $35,000 per year. Male college graduates in the same age group earned on average close to $40,000. Among women, the differences across educational attainment groups are also large. Moreover, the types of jobs available to high school dropout rarely provide opportunity for significant upward mobility or benefits such as health insurance. In other words, a relatively certain road to economic long-term marginality is to not complete, *at least*, secondary schooling. It is clear that it is important for policy makers to know what factors and issues are related to poor educational outcomes, such as dropping out of high school.

This paper has two main objectives. Firstly and primarily, we seek to identify factors which explain the higher dropout rates among the two largest minority groups in Texas – Hispanics and African-Americans. Our study is based on unique longitudinal student level data which include all Texas public school students in the cohort studied, the class of 1999. Secondly, we generate dropout rates in Texas that arguably more accurately reflect student high school outcome than official statistics and explain some reasons for the discrepancies.

2. Previous Research

There is a large body of research focusing on identifying determinants of educational outcomes and student success (a review of this literature can be found in Haveman and Wolfe (1995)). Family background, income and parental education are factors frequently found to affect children's schooling outcomes. Other determinants are neighborhood and peer effects, as well as school characteristics. Surprisingly, research fails to find a consistent relationship between school resources and student achievement

(Hanushek, 2006). Research investigating ethnic differences in school outcome, much of which focus on differences in test scores, have also found the above characteristics to be factors which help explain minority-white differences in academic outcomes (Lavin-Loucks, 2006). Next, we discuss relevant economics of education research on ethnic differences in dropout rates.

The influential recent study by Cameron and Heckman (2001) includes an analysis of the high dropout rates, specifically the low high school graduation rates, among Hispanics and African-Americans. They find that family factors (e.g. family composition, parental education, family income) explain all of the black-white gap in high school graduation rates and most of the Hispanic-white gap. Cameron and Heckman (2001) also find that differences between white and minority scholastic ability, as measured by the Armed Forces Qualification Test (AFQT), play a prominent role in explaining the white-minority gap. Furthermore, controlling for differences in family background and AFQT, Hispanics and African-Americans are found to be more likely than whites to graduate high school. Based on their findings, they conclude that it "is early differences in resources and not later ones that matter more" (p. 492).

Given the large proportion of immigrants in Texas, particularly Mexican immigrants, research on high school completion among immigrants is also relevant. Perreira, Mullan and Lee (2006) find that differences in dropout rates are driven by differences in human, cultural, school and community capital. They also find that first generation Hispanic immigrants are less likely to be high school dropouts compared to their parents but the relative gains in schooling attainment, compared to their parents, decreases by the second and third generation. Noting that the average Hispanic immigrant

has fewer than 9 years of schooling, and has already been out of school for at least one year by the age of 16, Betts and Lofstrom (2001) suggest that the low levels of educational attainment among young Hispanic, particularly Mexican, immigrants may be due to the possibility that they do not "drop in" to high school when they arrive in the U.S.

In our analysis below we include several of the observable factors included in the Heckman and Cameron (2001) and the Perreira, Mullan and Lee (2006) studies as well as school and school district characteristics. Unlike the Cameron and Heckman study, but like the Perreira, Mullan and Lee study, we are able to identify GED holders and we treat them as dropouts. Additionally, we are able to control for student's English ability.

3. Data

We utilize data from the Texas Schools Microdata Panel (TSMP), a student-level data set made available under special arrangements with the Texas Education Agency (TEA). These unique data contain information on year, age, grade and school enrolled, gender, ethnicity, whether the student is economically disadvantaged, whether English is not the primary home language (ESL), whether the student has been identified as being of Limited English Proficient (LEP) and whether the student participates in bilingual, special or gifted education classes. For the high school graduates in our sample, we also have information on the graduation date. Furthermore, based on age, year and grade enrolled, we calculate the number of grades a student has been held back.

In constructing our analytic data we begin with a balanced panel of all students whom we can observe if they are enrolled in Texas public high schools between the ages

of 15 and 20. Since the TSMP data currently include student level information from 1990 through 2001, this necessarily means that we restrict the sample to students who belong to cohorts that would be graduating in the classes from 1993 to 1999. For example, students who would be expected to graduate in 1993 are 15 years of age in 1990 and they will not turn 20 until 1995, and so students in this graduating cohort are in our sample and represent the earliest cohort. Similarly, students expected to graduate from high school in 1999 are 20 in 2001, and hence students in this graduating cohort are in our sample and represent the latest available cohort in our data. For the sample of 15 to 20 year-old students who have expected graduation dates from 1993 through 1999, we can observe, for each year, whether the student was enrolled or graduated from high school. For our analysis and empirical models we restrict our sample to the latest cohort we can follow up until they turn 20, i.e. the class of 1999. Nonetheless, we will present dropout rates for all cohorts in order to show the trend in Texas dropout rates.

Given the structure of the data, we define a student to be a school dropout if we observe him/her enrolled in a Texas public school at the age of 15 and in some subsequent year between the age of 15 and 20 this person is not enrolled in a Texas public school nor have they graduated from high school. Put slightly differently, we use information on whether the student was enrolled in a Texas public school in a particular year and/or if the student was reported graduating high school in that year. If the student is not enrolled and has not graduated high school, but was observed enrolled at age 15, we define the student to be a school dropout.

Caveats – Our Definition of Dropout Rates

Our focus in the paper is students' secondary educational outcome by the age of 20. We define a person to be a dropout if he/she was observed enrolled in a Texas public school at the age of 15 and had not graduated high school by the age of 20. Students who were observed enrolled in a Texas public school at the age of 15 and who either transferred to a private school, entered home schooling or moved out of state cannot be tracked in our data and are incorrectly identified as dropouts in the event they actually graduate high school. Also, students who were observed enrolled at age 15 in a Texas public school and passed away before graduating high school are also incorrectly identified as a dropout in our data. Our dropout rates may then be viewed as "upper bounds". On the other hand, our dropout rates are underestimated by the fraction of students who dropout at age 14 or younger.

Our dropout rates are higher than the reported dropout rates by the Texas Education Agency (TEA) but quite closely correspond to the dropout rates implied by the high school graduation rates reported by Greene (2001). Greene estimates that 67 percent of the students in the cohort who should have graduated in 1998 in Texas actually did graduate, implying a dropout rate of 33 percent. Our calculation yields a dropout rate of 34.6 percent for the 1998 graduation cohort. We will discuss plausible factors explaining differences between the TEA reported dropout rates and our in the next section.

Data from other Sources – School, District and Local Information

In our empirical models we also incorporate information on school district characteristics, such as expenditure and revenues per pupil. These data are collected from

National Center for Education Statistics' (NCES) Local Education Agency Finance Survey for the years 1996 to 2001.[1] We also incorporate school characteristics information, including pupil teacher ratio, enrolment, ethnic/racial student composition and school location. The school level data are generated from NCES' Common Core of Data and are also for the years 1996 to 2001. Lastly, we generate annual local labor market conditions such as unemployment rate, employment and earnings growth and average weekly earnings. These data are derived from the Bureau of Labor Statistics' Quarterly Census of Employment and Wages (QCEW). The local labor market characteristics are generated at the Metropolitan Statistical Area (MSA) in which the school district is located. If the school district is located outside a MSA, we utilize the state average in that particular year.

Descriptive Statistics

We present summary statistics by race/ethnicity for the class of 1999 in Table 1. The data indicate that there are large differences across groups in student characteristics. For example, while approximately 17 percent of white students are economically disadvantaged (defined as eligible for reduced price or free lunch), more than 70 percent of Hispanics and close to 56 percent of African-Americans are economically disadvantaged. Not surprisingly, the data also show that Hispanic and Asian students are more likely to be immigrants, as well as LEP and ESL students than whites or African-American students.

[1] The period 1996 to 2001 represents the years in which the students in our analytical cohort, the class of 1999, were between the ages of 15 and 20.

There are also differences across groups in school and school district characteristics students attend. The data indicates that white students are more likely to attend smaller schools with lower pupil teacher ratio. The majority of Hispanic and African-American students attend schools located in central cities. Students in these two minority groups also attend schools in district with lower expenditure per pupil. The local labor market data indicate that among the four groups, Hispanics reside in areas with the lowest average earnings, earnings growth and employment growth and the highest unemployment rates. Overall, our data indicate that Hispanics and African-Americans attend schools, and reside in areas, that are less conducive to academic success than white and Asian students.

4. Dropout Rates in Texas

Why Are TEA Reported Dropout Rates So Low?

Dropout rates in Texas declined in the 1990s. Figure 1 shows that 38.2 percent of the class of 1993 did not graduate high school by age 20 while the dropout rate amongst students in the class of 1999 is 32.7%. TEA also reports a decline in the dropout rate during this period. However, the TEA reported dropout rates are dramatically lower than the ones shown in Figure 1. For example, for the latest cohort available in our data, and the cohort we will focus on in our analysis below, the class of 1999, TEA reports a 9 percent dropout rates (TEA, 2001). An obvious question is why do these dropout rates differ so much? There are a number of reasons for the discrepancy. For example, TEA does not define a student who leaves school, without graduating, as a dropout if the

student become incarcerated, been expelled for criminal behavior, or has completed all course work but did not pass the Texas Assessment of Academic Skills (TAAS) exam.

One important reason for the difference the dropout rate is the role of the General Educational Development (GED) program. The TEA definition of dropouts excludes all individuals who instead of graduating receive the GED credential. Furthermore, students who withdraw from school to enroll in an approved alternative program are not counted as dropouts. This includes work toward completion of the GED certificate, but not necessarily successfully passing the battery of GED exams.

The distinction between being a high school graduate and holding instead the GED credential, which is commonly referred to as "high school equivalency" diploma, would arguably not be important if GED holders did as well in the labor market as high school graduate and pursued higher education to a similar extent. Research, however, quite convincingly show that neither of these hold true. Cameron and Heckman (1993) find that GED holders fare consistently worse than regular high school graduates on any number of labor market outcomes. Also, research fails to find that the GED increases earnings among minorities relative to other minority dropouts (Tyler, Murnane, and Willett, 2000). Regarding higher education, the findings of Lofstrom and Tyler (2005) suggest that the GED is not a particularly effective route to postsecondary education relative to staying in school and obtaining a high school diploma. It appears quite clear that dropouts with the GED are not the equivalent of high school graduates and hence should not be treated as such in official educational attainment statistics.

To illustrate the importance of the GED in explaining the discrepancy between our dropout rates and the official TEA dropout rate, we generated the percent of dropouts

from the class of 1999 who attempted and passed the GED exam by the age of 20, obtained from the TSMP data. We present these statistics in Table 2. The data show that 20 percent of dropouts in this cohort obtain the GED credential. If we exclude students with the GED credential, using our data for the class of 1999, the implied dropout rate is 28 percent, implying that the official dropout rates is understated, at least, by close to 17 percent.[2] Table 2 also shows that slightly more than 29 percent of dropouts attempted the GED by the age of 20. If we exclude from our sample all of these individuals, the implied dropout rates is 25.6 percent, suggesting that that the official dropout rate is understated by more than 27 percent. It should be noted that TEA does report the number of students who graduated in each cohort. For the Class of 1999 cohort, TEA reports that 78.1 percent of this cohort graduated, implying a dropout rate of 21.9 percent (TEA, 2001). However, this statistic appears to receive substantially less attention than the reported, so called, dropout rate.

Although these calculations make clear that the GED plays an important role in explaining the low official dropout rate, they also show that other factors are very important. For example, as long as a student withdraws from school with the *intent* to enroll elsewhere, the student is an "official other leaver" and will not be included in the statistics used to calculate the official dropout rate. Put differently, these students are not part of the denominator used to derive the official dropout rate. An indication that this issue is an important factor in explaining the difference between our dropout rates and the official TEA dropout rates is the discrepancy in the reported size of the class of 1999. TEA reports that the class of 1999, grade 7 cohort, includes 240,865 students while the

[2] The unadjusted dropout rates shown in Figure 1 is 32.7 percent suggesting that the TEA dropout rates is understated by, at least, a factor of (32.7– 28)/28 = 0.169.

class of 1999 used in this paper contains 298,581 students. Again, the GED, and the number of students attempting the GED, is a factor explaining the discrepancy. The number of GED test takers in the class of 1999, according to our data, is 28,365, which leaves a difference of more than 29,000 students between the TEA reported cohort size and ours. It appears unlikely that 29,000 students of the class of 1999, about 10 percent of the cohort, who were enrolled in a Texas public school at the age of 15 enrolled in a private school, out-of-state school, out-of-country school or became home schooled (*and* eventually graduated). Hence, arguably the official dropout rates do not accurately reflect secondary school attainment among students in Texas.

Differences in Dropout Rates across Groups

The main objective of this paper is to analyze differences in dropout rates and particularly to identify factors explaining the high dropout rates among Hispanic and African-American students. Our calculated dropout rates vary substantially across ethnic groups, shown in Figure 2, and are especially high among Hispanics and African-Americans. According to our data, slightly more than 40 percent of Hispanic students in the class of 1999 who were enrolled in a Texas public school at the age of 15 did not graduate high school by the age of 20. The dropout rate is almost as high for African-Americans, 38.5 percent. Both white and Asian students dropout to a significantly lower extent, 26 and 23.4 percent respectively. In our empirical analysis below we will investigate the causes of these large differences.

Of obvious interest is also the timing of dropping out of school. That is, at what age and/or grade are students most likely to drop out? Figure 3 shows unconditional

dropout probabilities by age for each of our ethnic/racial groups, i.e. group specific cumulative density functions (cdf). The figure indicates that relatively few students are dropouts by the age of 15 and that there are only small differences across groups at this young age. This is consistent with the findings of Cameron and Heckman (2001), who note that "few males quit school before age 16" (p. 459). Furthermore, they note that "this finding is due in part to laws about compulsory schooling attendance and to the lack of labor market opportunities for people younger than 16" (p .461). The largest difference in dropout rates in Texas at age 15, of 3 percent, is between Hispanic and Asian students. However, the dropout gap between these two groups increases with age with a considerable increase during the year students turn 17. Overall, Figure 3 suggests that the changes in the cumulative dropout rates over age are quite similar and relatively low for whites and Asians at one end of the spectrum, with Hispanics and African-Americans on the other end, displaying alarmingly high dropout rates.

The longitudinal data utilized here can also be used to generate conditional dropout probabilities. The conditional dropout probability, the hazard rate, is the probability that a student will dropout by the end of the year, conditional upon being enrolled during the year. Unlike the unconditional dropout probability, the hazard rate does not need to be a monotonically increasing function with age. It also has the advantage of more clearly illustrating the timing of differences in dropout probabilities across groups. We present the group specific hazard rates in Figure 4.

It appears that the large differences across groups in the dropout rates start at age 16. The figure shows that, conditional on having stayed in school until age 16, Hispanics and African-Americans are approximately twice as likely as an Asian student to dropout

by the end of the year. This roughly holds true for students who continue to be enrolled the following year, when they turn 17. Approximately 8 percent of Asian, and white, students drop out during this year while 14 percent of Hispanic students. Among African-American students, 15 percent do not continue school further. Not surprisingly, staying enrolled until age 18, the age at which most students who graduate do so, the conditional dropout rates decline for all groups. Although the gap in the conditional dropout probability between Hispanics and whites, as well as the gap between African-Americans and whites, decrease to 5 percent, these minority groups are more than twice as likely to leave school, without graduating, at this age than white students. Overall, Figure 4 suggests that the higher dropout rates among Hispanics and whites, compared to whites and Asians, develop at ages 16 and 17.

Arguably, it is of greater relevance to determine at what grade, as opposed to age, the dropout differences arise. To shed light on this, we generated group specific conditional dropout probabilities, hazard rates, by grades. These are presented in Figure 5. The figure quite clearly shows that these differences are particularly generated in the 9^{th} grade and that relatively few students dropout in earlier grades. Approximately 12 percent of African-American students who enrolled in 9^{th} grade did not complete this grade. Among Hispanic students, an even larger proportion of students did not finish 9^{th} grade, slightly more than 15 percent. This is in stark contrast to white and Asian students. For these two groups, 6 percent and 4 percent respectively dropout in 9^{th} grade. Figure 5 also shows quite clearly that these differences decline in subsequent grades, with the exception of the difference between whites and African-Americans in 12^{th} grade. Surprisingly, the conditional dropout rates among African-American students who made

it all the way to 12th grade is higher than the hazard rate among African-American students enrolling in the 11th grade. In fact, the conditional dropout probabilities remain relatively constant among African-American students from grades 9 through 12 while they decline among Hispanic students in these grades.

The importance of the 9th grade as an explanation for the differences in dropout rates between the groups can also be illustrated by showing the group specific proportion of all dropouts who dropped out in 9th grade. About 16 percent of Asian dropouts dropped out in the 9th grade while about 23 percent of white dropouts enrolled in the 9th grade without completing it. Among Hispanic and African-American drop outs, 37 percent and 31 percent respectively experienced 9th grade as their last grade.

One plausible explanation for the substantially higher Hispanic and African-American 9th grade conditional dropout probabilities, compared to whites, is differential grade retention across ethnic groups. If a student is progressing through school at a normal rate, i.e. have not been held back a grade, we would expect him/her to be in the 9th grade by the age of 15. If however a student has been held back one grade, the student would be expected to be in 9th grade the year he/she turns 16. Importantly, as noted above, this is also the year in which students, given compulsory student attendance, may leave school. Furthermore, Cameron and Heckman (2001) report that minority groups are significantly more likely to have fallen behind in school by age 15 relative to whites.

To investigate whether normal grade progression, or lack thereof, is a key reason for the large minority-white difference in the 9th grade conditional dropout probability, we generate the grade specific conditional dropout probabilities by ethnic groups for the sub-sample of students who were on track at age 15, and were hence enrolled in 9th grade

or above, by this age. The hazard rates for this restricted sample are shown in Figure 6. The figure reveals that although the differences are smaller, large differences remain and the largest minority-white gaps, specifically, Hispanic/black-white gaps, are still observed for the 9th grade. It appears that other factors than differences in grade retention across ethnic groups are the main determinants. Although beyond the scope of this paper, further research is necessary to determine what specifically causes the increase in the dropout probability at this point and why minorities have a more difficult time transferring to high school.

5. Empirical Dropout Probability Model

Table 1 suggests that there are large differences in student, school and district characteristics across the four ethnic/racial groups. A key objective of this paper is to determine how these factors relate to the differences in dropout rates across the groups.

The educational outcome analyzed in this paper is whether or not a student of the class of 1999 who was observed enrolled in a Texas public school at the age of 15 is observed not completing high school by the age of 20, and hence is defined to be a dropout. Let the outcome variable y_{ijk} equal zero if student i in school j, in school district k is observed graduating from high school by age 20, and let y_{ijk} equal one if the student is observed dropping out between the ages of 15 and 20.[3]

We may think of the observed outcome y_{ijk} being the result of a latent process in which a student compares the marginal benefits to the marginal costs of continuing schooling. In this case, let the continuous latent variable, y_{ijk}, which represents the value

[3] Given our approach and data, the outcome analyzed is identical to not graduating high school by the age of 20. One implication, and advantage, of this is that students who leave school but return at a later point are only defined to be dropouts if they did not graduate by the age of 20.

an individual *i*, in school *j*, and district *k* receives from the particular decision in each school year during the period in which the students is still enrolled and between the ages of 15 and 20, be specified as:

$$y^*_{ijk} = \alpha_0 + \alpha_1 Hisp_{ijk} + \alpha_2 Black_{ijk} + \alpha_3 Asian_{ijk} + \mathbf{X}_{ijk}\boldsymbol{\beta} + \mathbf{W}_{jk}\boldsymbol{\delta} + \mathbf{Z}_k\boldsymbol{\gamma} + \varepsilon_{ijk}$$
$$y_{ijk} = 1(y^*_{ijk} > 0)$$

where 1(·) is an indicator function equal to one if the enclosed statement is true, i.e. the student opted to drop out and zero otherwise. The corresponding, linear dropout probability model is given by equation 1:

$$\mathbf{P}\left[y_{ijk} = 1 \,|\, \mathbf{X}, \mathbf{W}, \mathbf{Z}, \text{Race/Ethnicity} \right] = \alpha_0 + \alpha_1 Hisp_{ijk} + \alpha_2 Black_{ijk} + \alpha_3 Asian_{ijk} + \mathbf{X}_{ijk}\beta + \mathbf{W}_{jk}\boldsymbol{\delta} + \mathbf{Z}_k\boldsymbol{\gamma} + \varepsilon_{ijk} \quad (1)$$

Where *Hisp*, *Black* and *Asian* are indicator variables for Hispanics, African-American and Asian students respectively and matrices **X**, **W** and **Z** are defined as;

$\mathbf{X}_{ijk}$ = Matrix containing *student characteristics* such as gender, whether the student is designated as economically disadvantaged, reports English as a Second Language (ESL) , is designated as Limited English Proficiency (LEP), indicator variables for whether the student had been held back a grade or more by age 15, is identified as an immigrant, and whether he/she participated in special or gifted education.

$\mathbf{Z}_{jk}$ = Matrix containing controls for *school characteristics*, such as pupil teacher ratio, enrollment, school location, percent of students who received free lunch and percent of students who are white, Hispanic, African-American and Asian.

$\mathbf{W}_k$ = Matrix containing controls for school *district characteristics* and *local labor market conditions* by the MSA the district is located in. The matrix includes variables for expenditure per pupil and per

student federal Title 1 revenues. The local labor market characteristics included are employment growth, the unemployment rate, average weekly earnings and earnings growth.

The coefficients of particular interest in the above specification are α_1, α_2 and α_3 which represent the differences in the dropout probability between white and minority students. An alternative, and far more complex, approach would be to model the dropout probability as a dynamic process, such as Cameron and Heckman (1998) and Colding (2006). Given that our empirical approach does not account for possible dynamic selection bias, the results should be interpreted with some caution.

6. Empirical Results

The model specifications defined above are estimated using a sample of students who belong to the class of 1999 cohort. Our approach is to start with a parsimonious model specification and subsequently add controls for factors expected to affect the dropout decision and group differences in the dropout probability. This will shed light on what specific factors contribute to the white – minority gaps and to what extent. The estimated coefficients and t-statistics are shown in Table 3.[4]

We begin by presenting the Linear Probability Model (LPM) results from our first specification, Model 1, which only controls for gender and nativity. The objective here is to investigate whether the high dropout rates among Hispanic students is partially driven by the higher proportion of immigrants, who may face difficulties adjusting to U.S. schools. The results show that although immigrant students are significantly more likely to drop out of school, the higher proportion of foreign born students among Hispanics

[4] The standard errors used to derive the t-statistics are White's robust standard errors. Also, we relax the assumption that student observations within school are independent, i.e. they are clustered at the school level.

only explain a small proportion of this group's overall higher dropout rate. Native born Hispanic students are still about 13 percentage point more likely to drop out than white students, compared to the unadjusted difference of 14.4 percentage points. Not surprisingly, given the low proportion of immigrants among African-Americans, the controls for gender and nativity do not affect the estimated black-white dropout gap, which remains at 12.5 percentage points.

Minority students are more likely to be economically disadvantaged, as can be seen in Table 1, a factor likely to have an impact on educational attainment. To assess the role of poverty on group differences in the drop out probability, we add our measure of students coming from an economically disadvantaged family, shown as Model 2. The estimates clearly illustrate the influence of poverty on both the dropout probability and differences between minority and white students. A student who is economically disadvantaged is approximately 12 percentage points more likely to drop out of school than other students. Our estimate also show that, other factors constant, Hispanic students are approximately 7 percentage point more likely than white students to drop out of school, while African-American students are roughly 8 percentage points less likely to complete high school. These results imply that our simple poverty measure explains almost 1/2 of the Hispanic-white dropout probability difference and more than 1/3 of the black-white disparity.

Another potentially important factor contributing to the high minority dropout rates, mainly among Hispanic students, is English proficiency (Valenzuela, 2006). We add controls for English as a Second Language (ESL) and Limited English Proficiency (LEP) to our models, Model 3. Both of these variables are found to affect the dropout

probability. The results also show that the higher dropout rates among immigrant students is due to relatively low English proficiency - the estimated immigrant coefficient is now statistically insignificant. Limited English proficiency is also a contributor to the Hispanic-white difference in the dropout probability, which decrease to about 4.4 percentage points with the added ESL and LEP variables. The two language factors explain slightly more than 1/3 of the estimated dropout probability difference between Hispanic and white students, compared to our Model 2 specification estimate.

We next turn to the role of school characteristics and particularly pupil teacher ratio and school size, the results are shown as Model 4. Although the school average pupil teacher ratio is found to affect the dropout probability positively and significantly, school size is not found to significantly alter students' decision to leave school before graduation, in this model specification. Furthermore, these factors do not appear to explain much of the higher dropout probabilities among Hispanic and African-American students, relative to white students.

The next model specification, Model 5 is intended to investigate the role of the location of the school and the ethnic/racial composition of its students, the latter capturing to some extent peer effects. We show in Table 1 that Hispanic and African-American students are much more likely to attend schools that are likely to be located in the center of large cities than white students. Our data also show that Texas schools are segregated in the sense that white, Hispanic and African-American students are most likely to attend schools where their ethnic/racial group is the largest ethnic/racial group. Clearly, the school composition is a factor of the area, or neighborhood, characteristics,

and the effects of composition and location on educational outcome are difficult to disentangle.

The results for Model 5 indicate that, holding all other factors constant, students attending schools in central large and mid-size cities are more likely to drop out of school than students attending schools in any other location. The estimates also indicate that students attending schools with a higher proportion of economically disadvantaged students, ceteris paribus, are somewhat less likely to graduate high school. The results also point to a positive association between the dropout probability and the proportion of African-American and Asian students. Interestingly, the estimates also imply a negative correlation between the dropout probability and the proportion of Hispanic students. Importantly, our estimates show that these location related factors strongly affect the difference in the dropout probability between African-American and white students. Once we add these controls to our specification, and holding all factors constant, African-American students are no more likely to dropout of school than white students. However, Hispanic students are still found to be more likely to not complete high school compared to white students.

Other potentially influential dropout factors are the amount spent per student, i.e. per pupil expenditure, and revenues received from Title I. The latter funds are specifically intended to improve the academic achievement of the disadvantaged. It is possible that differences across school district in the amounts received from Title I, as well as per student expenditure, may also affect ethnic group differences in the dropout probability. Our results when we add these two factors to our specification, Model 6, suggest that higher spending and Title I revenues per student are associated with a lower

student dropout probability. However, we do not find any indication that differences in spending or revenues are factors explaining the high minority dropout rates. Similarly, when we add our controls for local labor market conditions, shown as Model 7, although they appear to be linked to the probability of completing high school, they do not seem to be strong contributors to group differences in the dropout probability.

The most general specification presented so far, Model 7, suggests that the higher dropout probability among African-American students, relative to white students, can entirely be explained by our set of observable factors. Observationally similar Hispanic students are however still predicted to be more likely to drop out of school than white students, by approximately 3.6 percentage points.

Our last model specification, Model 8, adds an indicator variable for whether the student had been held back a grade or more by the age of 15. This variable is based on whether the student was observed enrolled in 8^{th} grade, or lower, the year in which he/she turned 15. Grade retention is arguably simply a symptom, or predictor, of poor educational outcomes resulting from academic and personal difficulties experienced by the student. Viewed from this perspective, a finding that students who were held back a grade are more likely to drop out of school is neither surprising nor extremely useful since it is a manifestation of these negative experiences, without indicating which specific ones caused the student to dropout. Nonetheless, in the specification with the grade retention variable, Model 8, our estimated minority-white differences in the dropout probability will tell us what these differences are conditional on grade retention, i.e. whether a Hispanic student who has been held back a grade is more likely to drop out than a white student who was held back a grade, all other factors held constant.

Our estimates show that differences in grade retention between Hispanic and white students is a relevant factor. Students who have been held back a grade are substantially more likely to drop out, by close to 35 percentage points, compared to students who stay on track. In other words, grade retention is a relatively good predictor of dropping out of school. The Model 8 estimates suggest that lack of successful grade progression among Hispanic students result in a higher Hispanic dropout rates than whites. However, the estimates indicate that Hispanics students who made normal grade progression and were enrolled in at least 9^{th} grade by age 15 are still more likely to drop out, by about 3 percentage points, than a white student who had not been held back a grade at age 15.

Lastly, to determine whether other school characteristics, beyond the ones specified above, for which we do not have information help explain the difference in the dropout probability between Hispanic and white students, we also estimated a school fixed effects specification, shown as Model 9. Note that in this specification, we cannot include the observable school and school district variables, since they are perfectly multicolinear with the school fixed effects. The estimated Hispanic-white difference decreases only marginally to 2.8 percentage points. These results suggest that the school characteristics included in Models 7 and 8 are the ones most relevant for explaining the Hispanic-white dropout probability difference. Lastly, the fixed effects results also serve as a robustness check of our estimated individual student effects. The generally small differences in the estimates between Model 8 and 9, indicates that the student level estimates are not driven by unobserved school heterogeneity which may be correlated with the student characteristics.

7. Summary and Conclusions

This paper analyzes the high dropout rates among Hispanic and African-American students in Texas. Utilizing unique longitudinal student data, we show that Hispanic and African-American students are about 14 and 12 percentage points, respectively, more likely to drop out of school than white students. We show that the differences in the dropout rates appear to develop in 9^{th} grade.

We estimate multivariate dropout probability models to asses what factors affect the student dropout probability and differences across groups. Factors associated with the higher minority dropout rates differ somewhat between Hispanics and African-Americans. However, one determinant in common for these minority groups is poverty. Almost 1/2 of the difference in the dropout probability between Hispanic and white students stems from the greater prevalence of being economically disadvantaged among Hispanic students. More than 1/3 of the African-American and white student difference in the dropout probability is linked to our simple measure of student poverty. We also find that lack of English proficiency is a key factor in explaining the high Hispanic dropout rate. Our results suggest that neighborhood characteristics, here simply measured by the location of the school attended and student race/ethnicity composition of the school, contribute more to the high African-American dropout probability than school characteristics such as pupil teacher ratio and expenditure per pupil. This suggests the need for future research to not only focus on school characteristics but to also determine the role of neighborhood characteristics on student outcomes.

A secondary objective of the paper is to attempt to explain the discrepancy in official dropout rates reported by the TEA and the ones derived using appropriate and

unique student data such as the one utilized here. We show that the GED program is an important reason for the differences since dropouts with a GED or students who enroll in a GED preparation program are not included in the TEA dropout statistics. Given previous research findings strongly indicating that individuals with the GED credential are less successful in the labor market and have lower post secondary educational attainment than high school graduates, we suggest that these dropouts should not be treated and viewed as high school equivalents.

References

Betts, Julian R. and Magnus Lofstrom. (2000) "The Educational Attainment of Immigrants: Trends and Implications," in George J. Borjas ed. Issues in the Economics of Immigration, University of Chicago Press, for National Bureau of Economic Research, pp. 51-115.

Cameron, Stephen V. and James J. Heckman (1993) "The Nonequivalence of High School Equivalents." *Journal of Labor Economics*, 11, no. 1, 1-47.

Cameron, Stephen V. and James J. Heckman (1998) "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males" *Journal of Political Economy*, 106, 262-333.

Cameron, Stephen V. and James J. Heckman (2001) "The Dynamics of Educational Attainment for Black, Hispanic, and White Males" *Journal of Political Economy*, 109, 455-499.

Colding, Bjorg (2006) "A Dynamic Analysis of Educational Progression of Children of Immigrants" *Labour Economics*, 13, 479-492.

Greene, Jay P. (2001) "High School Graduation Rates in the United States", The Manhattan Institute for Policy Research Civic Report, November.

Hanushek, Eric A. (2006) "School Resources" in Eric A. Hanushek and Finis Welch (ed.), Handbook of the Economics of Education. Amsterdam: Elsevier.

Haveman, Robert and Barbara Wolfe (1995) "The Determinants of Children's Attainments: A Review of Methods and Findings" *Journal of Economic Literature*, 33, 1829-78.

Lavin-Loucks, Danielle (2006) "The Academic Achivement Gap", Williams Institute Research Brief, July.

Lofstrom, Magnus and John H. Tyler, (2005) "Is the GED an Effective Route to Postsecondary Education?" Paper presented at the 2005 Association for Public Policy Analysis and Management Fall Research Conference.

Perreira, Krista, Kathleen Mullan Harris and Dohon Lee (2006) "Making it in America: High School Completion by Immigrant and Native Youth" *Demography*, 43, 511-536.

Snower, Dennis J. (1999) "Causes of Changing Earnings Inequality" IZA Discussion paper No. 29, January 1999.

Texas Education Agency (2001) “Secondary School Completion and Dropouts in Texas Public Schools 1999-00”, August, Austin, Texas.

Tyler, John H., Richard J. Murnane, and John B. Willett (2000) "Estimating the Labor Market Signaling Value of the GED." *Quarterly Journal of Economics* CXV, no. 2, 431-468

Valenzuela, Angela (2006) “The Disappearance of High School English Language Learners from Texas High Schools” *Williams Review*.

Table 1. Sample Means, Class of 1999, by Race/Ethnicity.

	White	Hispanic	African-American	Asian
Dropout Rate	0.260	0.403	0.385	0.234
Student Characteristics				
Male	0.515	0.519	0.509	0.506
Immigrant	0.009	0.116	0.013	0.161
Economically Disadvantaged	0.172	0.702	0.558	0.289
Limited English Proficiency	0.004	0.234	0.005	0.176
English as a Second Language	0.003	0.183	0.004	0.145
Gifted Program Participant	0.140	0.070	0.093	0.242
Special Education	0.132	0.124	0.165	0.030
Held Back a Grade or More	0.273	0.486	0.428	0.264
School Characteristics				
Pupil Teacher Ratio	14.4	15.2	15.5	16.0
Total Enrolment	1,561	1,685	1,636	2,318
Enrolment Less than 700	0.241	0.149	0.144	0.047
Enrolment 700 - 1,400	0.198	0.195	0.264	0.086
Enrolment 1,400 - 1,900	0.176	0.229	0.229	0.158
Enrolment 1,900 - 2,500	0.200	0.283	0.228	0.330
Enrolment Greater than 2,500	0.184	0.144	0.135	0.379
Percent Free Lunch	17.8	40.8	29.4	17.7
Percent White	66.9	26.7	34.3	50.6
Percent Hispanic	19.7	61.8	23.2	23.3
Percent African-American	10.2	9.5	39.1	16.4
Percent Asian	2.9	1.8	3.2	9.5
School Location				
Central City, Large City (250K+)	0.164	0.388	0.461	0.351
Central City, Mid-size City (<250K)	0.139	0.202	0.173	0.104
Urban fringe, Large city	0.327	0.151	0.195	0.441
Urban fringe, Mid-Size city	0.052	0.057	0.021	0.018
Large town	0.010	0.011	0.012	0.002
Small town	0.112	0.100	0.065	0.026
Rural	0.196	0.090	0.074	0.057
School District Characteristics				
Expenditure per Pupil	6,526	6,326	6,170	6,463
Title 1 Revenue per Pupil	103	216	157	89
Local Labor Market Conditions				
Employment Growth (%)	1.8	1.9	1.8	1.9
Unemployment Rate (%)	4.4	6.2	4.4	4.3
Average Weekly Earnings ($)	639	576	655	680
Earnings Growth (%)	4.6	4.3	4.7	4.7
Sample Size	145,365	103,809	40,887	7,830

Table 2. Number and Percent of Dropouts Who Passed and Attempted the GED, Class of 1999, by Race/Ethnicity.

	Total Number of Students	Total Number of Dropouts	Number of Dropouts Who Passed GED	Percent of Dropouts Who Passed GED	Number of Dropouts Who Attempted GED	Percent of Dropouts Who Attempted GED
White	145,365	37,785	10,806	28.6	13,830	36.6
Hispanic	103,809	41,886	6,230	14.9	10,223	24.4
African-American	40,887	15,751	2,196	13.9	3,882	24.6
Asian	7,830	1,834	260	14.2	348	19.0
Native American	690	293	60	20.5	82	28.0
All Students	298,581	97,549	19,552	20.0	28,365	29.1

Table 3. Linear Probability Models of the Probability of Dropping Out of School, Class of 1999.

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Hispanic	**0.131**	**0.069**	**0.044**	**0.043**	**0.039**	**0.039**	**0.036**	**0.031**	**0.028**
	(16.78)	**(9.03)**	**(5.99)**	**(6.20)**	**(8.05)**	**(8.25)**	**(7.62)**	**(6.62)**	**(11.14)**
African American	**0.125**	**0.079**	**0.083**	**0.070**	**-0.015**	**-0.014**	**-0.009**	**-0.009**	**-0.005**
	(12.99)	**(8.29)**	**(8.73)**	**(7.05)**	**(2.82)**	**(2.52)**	**(1.92)**	**(1.79)**	**(1.55)**
Asian	-0.042	-0.054	-0.071	-0.021	-0.053	-0.052	-0.053	**-0.051**	**-0.054**
	(4.53)	(6.39)	(8.47)	(2.27)	(7.08)	(6.92)	(7.80)	**(7.48)**	**(11.14)**
Native-American	0.162	0.142	0.142	0.126	0.111	0.105	0.087	**0.086**	**0.104**
	(7.75)	(7.02)	(7.00)	(6.41)	(5.75)	(5.45)	(4.64)	**(4.77)**	**(5.98)**
Student Characteristics									
Immigrant	0.111	0.095	0.011	0.026	0.013	0.017	0.041	0.037	-0.006
	(6.38)	(5.77)	(0.83)	(2.31)	(1.08)	(1.69)	(2.58)	(2.35)	(1.33)
Male	0.071	0.070	0.068	0.056	0.055	0.056	0.052	0.048	0.046
	(29.40)	(28.75)	(28.66)	(26.77)	(26.57)	(26.66)	(23.53)	(21.91)	(28.76)
Economically Disadvantaged		0.121	0.109	0.082	0.076	0.076	0.071	0.066	0.077
		(20.22)	(17.82)	(14.24)	(15.04)	(15.15)	(13.32)	(12.46)	(37.08)
Limited English Proficiency			0.073	0.056	0.053	0.046	0.048	0.042	0.046
			(3.83)	(3.62)	(3.70)	(2.70)	(1.95)	(1.79)	(7.04)
English as a Second Language			0.130	0.137	0.136	0.142	0.128	0.120	0.130
			(5.90)	(7.44)	(8.19)	(7.55)	(4.78)	(4.68)	(17.90)
Gifted Program Participant				-0.227	-0.240	-0.232	-0.181	-0.175	-0.234
				(30.99)	(25.77)	(24.61)	(19.51)	(19.60)	(117.60)
Special Education				0.074	0.076	0.078	0.074	0.062	0.060
				(14.54)	(15.76)	(16.20)	(11.54)	(9.76)	(22.59)
School Characteristics									
Pupil Teacher Ratio				0.008	0.006	0.006	0.005	0.005	
				(3.12)	(2.85)	(2.77)	(2.43)	(2.46)	
Total Enrolment				-0.005	-0.033	-0.044	-0.052	-0.047	
				(0.21)	(1.52)	(2.05)	(2.53)	(2.30)	
Enrolment 700 - 1,400				0.024	-0.007	-0.023	-0.001	-0.003	
				(1.23)	(0.41)	(1.21)	(0.07)	(0.18)	

Table 3 (Continued)	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**	**Model 9**
Enrolment 1,400 - 1,900				-0.013	-0.059	-0.060	-0.022	-0.015	
				(0.40)	(1.87)	(1.92)	(0.65)	(0.45)	
Enrolment 1,900 - 2,500				-0.073	-0.103	-0.100	-0.073	-0.062	
				(1.77)	(2.62)	(2.56)	(1.85)	(1.56)	
Enrolment Greater than 2,500				-0.070	-0.084	-0.081	-0.025	-0.016	
				(1.15)	(1.46)	(1.42)	(0.41)	(0.27)	
Percent Free Lunch					0.002	0.002	0.001	0.001	
					(3.40)	(4.31)	(1.58)	(0.83)	
Percent Hispanic					-0.0012	-0.001	-0.002	-0.002	
					(3.09)	(1.45)	(3.04)	(2.74)	
Percent African-American					0.001	0.001	0.003	0.003	
					(2.74)	(2.95)	(5.55)	(5.81)	
Percent Asian					0.003	0.003	0.008	0.007	
					(2.00)	(2.03)	(4.12)	(3.90)	
School Location									
Central City, Large City (250K+)					0.191	0.162	0.150	0.136	
					(10.93)	(8.94)	(5.18)	(4.76)	
Central City, Mid-size City (<250K)					0.155	0.129	0.015	0.003	
					(10.34)	(7.65)	(0.38)	(0.07)	
Urban fringe, Large city					0.140	0.117	0.140	0.130	
					(10.20)	(8.33)	(6.61)	(6.28)	
Urban fringe, Mid-Size city					0.032	0.021	-0.173	-0.173	
					(1.78)	(1.05)	(4.25)	(4.36)	
Large town					0.080	0.044	0.045	0.040	
					(2.01)	(1.01)	(0.64)	(0.56)	
Small town					0.033	0.025	0.033	0.032	
					(3.05)	(1.95)	(1.94)	(1.94)	

Table 3 (Continued)	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**	**Model 9**
School District Characteristics									
Expenditure per Pupil						-0.037	-0.033	-0.032	
						(2.96)	(2.90)	(2.89)	
Title 1 Revenue per Pupil						-0.342	-0.485	-0.447	
						(4.39)	(4.24)	(3.96)	
Local Labor Market Conditions									
Unemployment Rate							0.024	0.023	
							(5.27)	(5.32)	
Weekly Earnings							-0.002	-0.002	
							(13.55)	(13.88)	
Employment Growth							0.064	0.064	
							(8.48)	(8.63)	
Weekly Earnings Growth							0.064	0.064	
							(8.48)	(8.63)	
Grade Retention									
Held Back a Grade at Age 15								0.349	0.302
								(37.66)	(61.04)
Number of Students					298,581				
R-squared	0.031	0.043	0.052	0.091	0.113	0.128	0.259	0.275	0.279

Figure 1

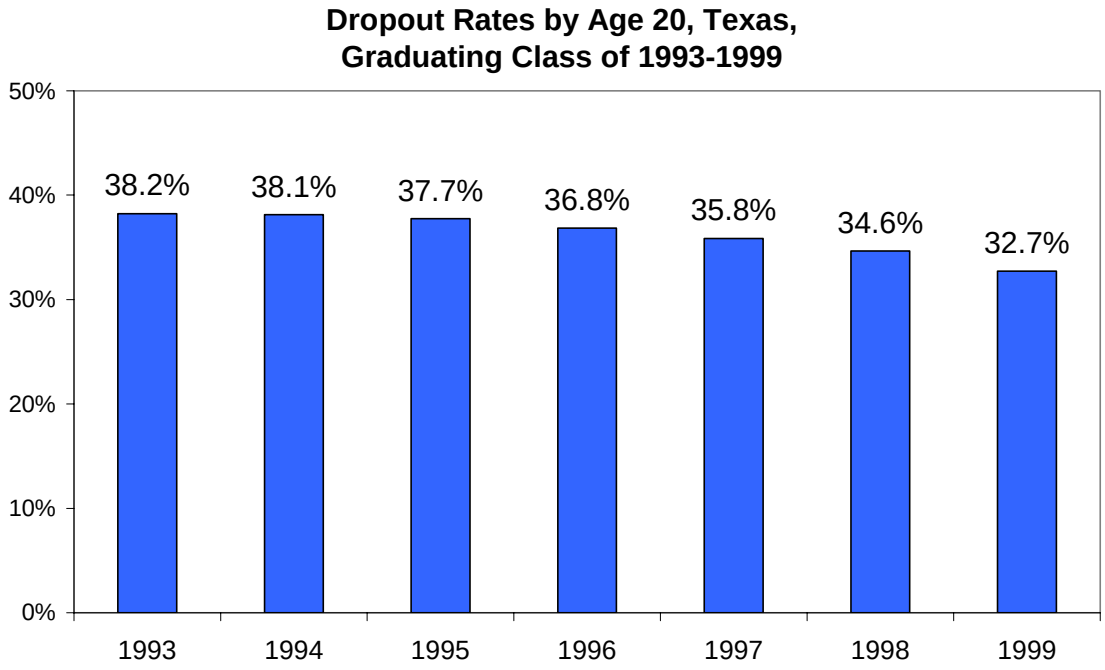

Figure 2

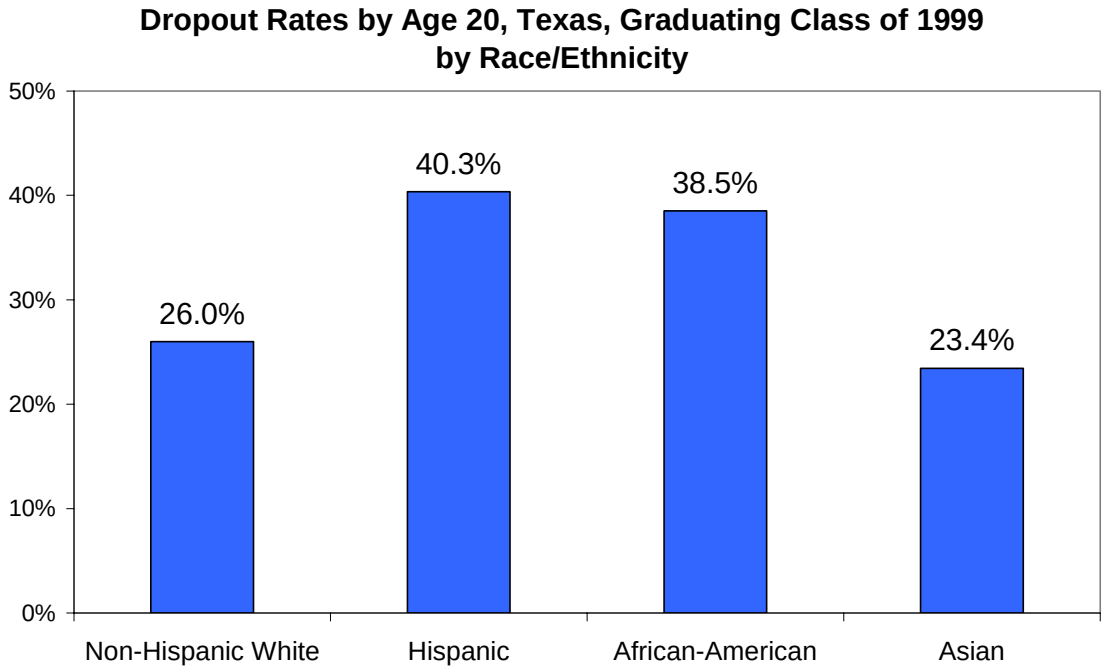

Figure 3

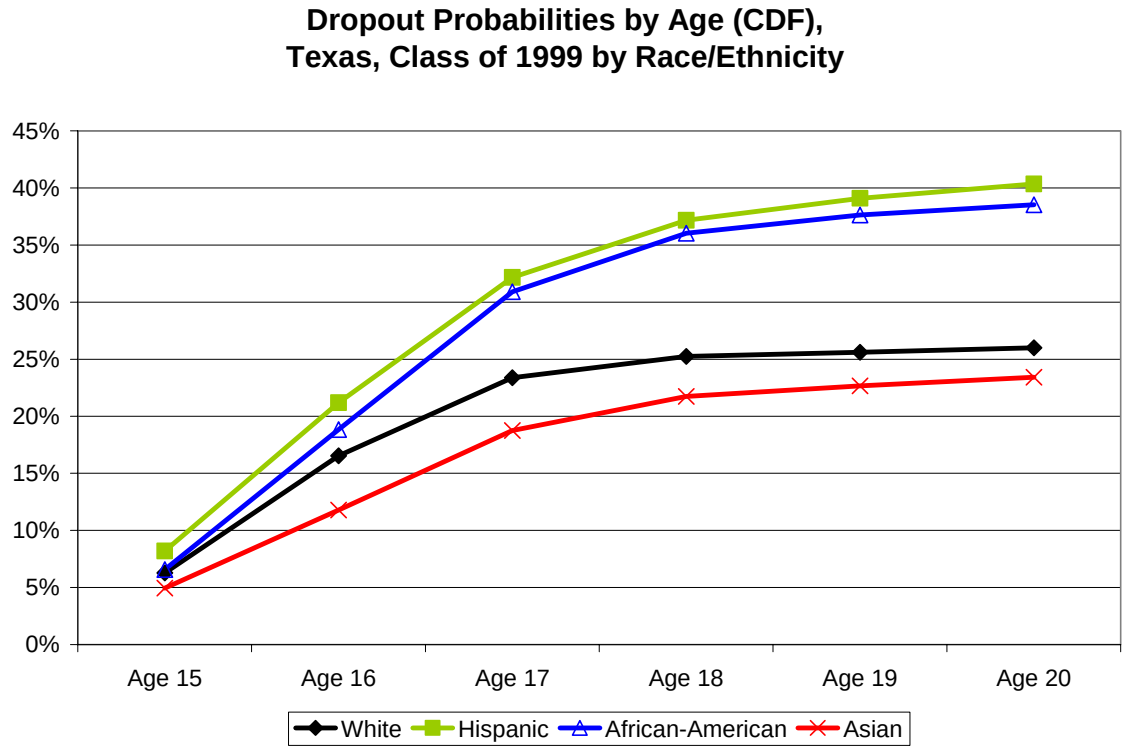

Figure 4

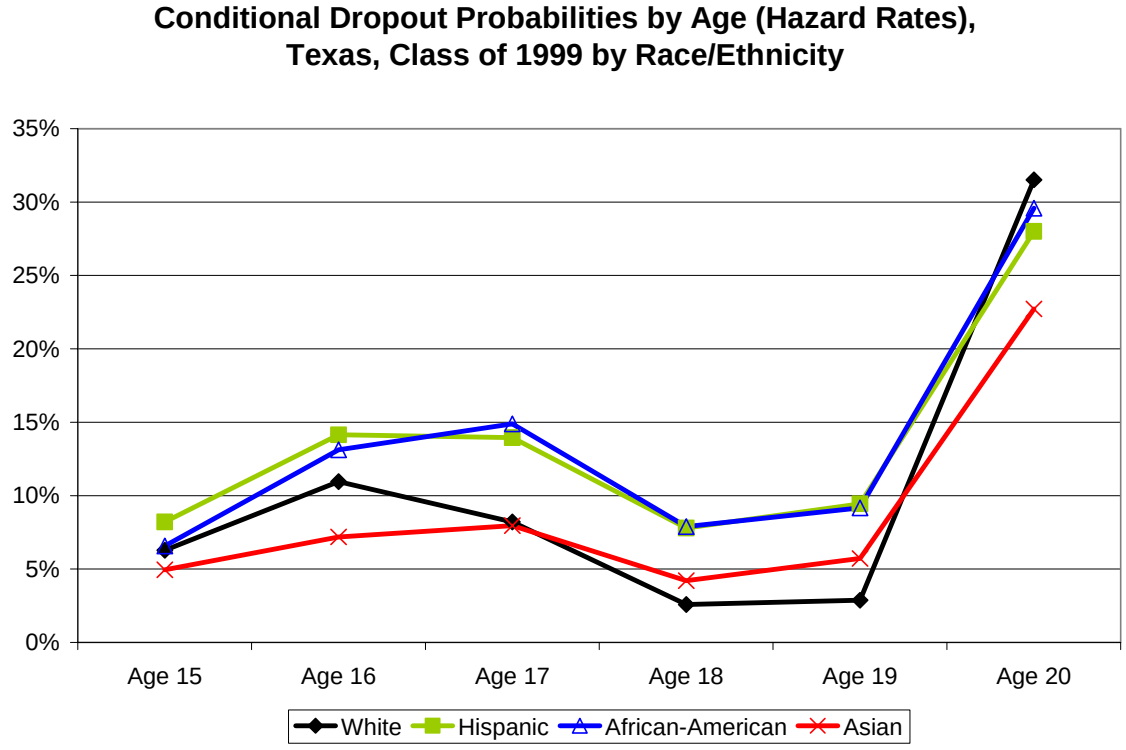

Figure 5

Conditional Dropout Probabilities by Grade (Hazard Rates), Texas, Class of 1999 by Race/Ethnicity

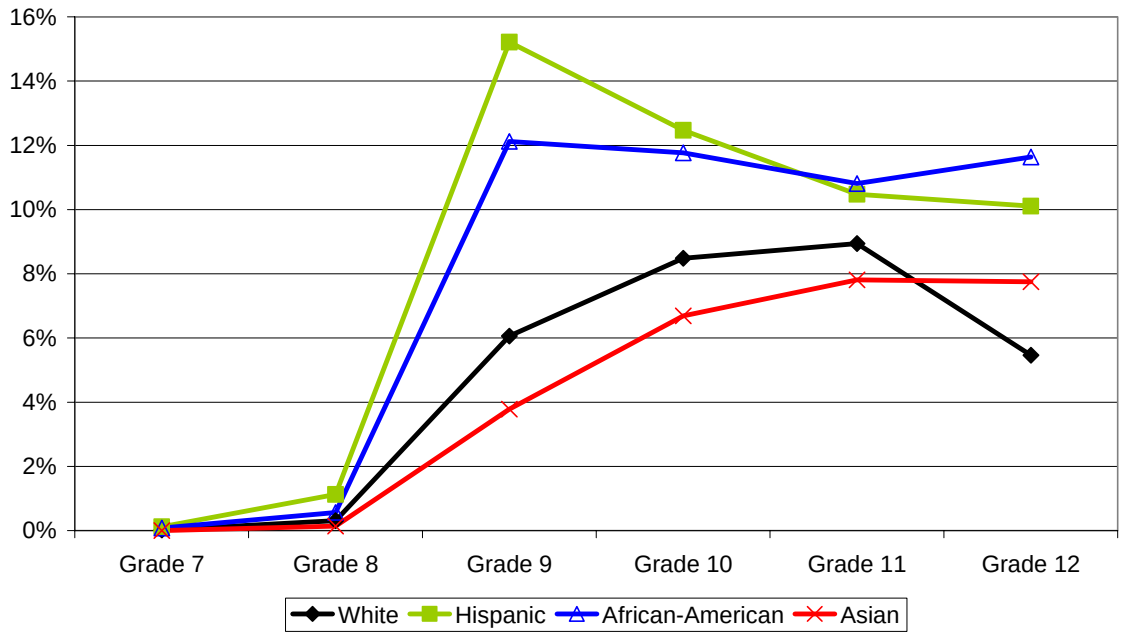

Figure 6

Conditional Dropout Probabilities by Grade (Hazard Rates), Texas, Class of 1999 by Race/Ethnicity Normal Grade Advancement at Age 15