Data Sources, Methods, and Limitations

Data Sources and Methods

The primary data used in the analyses came from the TEA and THECB. The data provided detailed information on college enrollment and completion for students attending public and private postsecondary institutions in the State of Texas as well as measures of students’ demographic and socioeconomic background. Because the datasets contained records only for students who enrolled in institutions in the State of Texas, the statistics presented are likely underestimates of students’ overall participation in postsecondary education.

Aside from presenting figures on college enrollment and completion, we also used data from the TWC, which allowed us to observe whether students were earning wages from employment in the State of Texas. In our annual estimates of student activities, a student was classified as “working” if they were found in the TWC data but not enrolled in a postsecondary institution. Data from the TEA, THECB, and TWC were accessed and analyzed at the UH ERC in Houston, Texas.

Background Characteristics, Fall 1998: For Texas, Harris County, and the 20 districts in Harris County, we provide estimates of the demographic and socioeconomic composition of students who began 8th grade in 1998. For gender, we show the share of male and female students. In terms of race and ethnicity, we present statistics for non-Hispanic white, non-Hispanic black, Hispanic (of any race), non-Hispanic Asian, and non-Hispanic Native American students. (Please note that we do not show any outcome data for Native American students due to their small sample size.) The measure of socioeconomic status we use is developed by the Texas Education Agency (TEA) and shows whether a student is “economically disadvantaged”—receiving free or reduced-price meals through the National School Lunch and Child Nutrition Program.

Student Activities, Fall 2003 – Fall 2014: These annual, non-cumulative estimates show the percentage of students attending a Texas postsecondary institution, working in the Texas labor force, or engaged in another activity in the fall of a given year. We divide postsecondary attendance into two categories: 1) attending a two-year, technical, or vocational institution offering certificates, diplomas, licenses, or associate’s degrees and 2) attending a four-year institution offering bachelor’s or more advanced degrees. Students who are neither attending a Texas postsecondary institution nor working in the Texas labor force are part of the “other
In three select years—2003, 2008, and 2013—we estimate breakdowns of the “other activities” category. These breakdowns include the following:

- Enlisted in the military
- Unemployed
- Not in the labor force
- Incarcerated
- Deceased
- Living out-of-state: attending a postsecondary institution
- Living out-of-state: engaged in other activities

While the administrative data were useful, data on these activities were unavailable at the UH ERC. For example, there was no link between data from K-12 public education to the armed forces. Using data from the following external sources, we estimated the share of students in these other categories in three selected years: 2003, 2008, and 2013.

- Current Population Survey (CPS), October Supplement (School Enrollment), 2001–2016: The CPS is a nationally-representative, monthly survey administered by the Bureau of Labor Statistics. The primary focus of the CPS is employment, but it also asks respondents questions about their demographic background and educational attainment. The October Supplement contains additional items on respondents’ education, specifically whether they are currently attending a college or university. Using the CPS’s robust information on labor force participation and schooling, we calculated five-year average estimates of two categories—unemployment and not in the labor force—for three age ranges: 18-22 (years: 2001-2005), 23-27 (years: 2006-2010), and 29-33 (years: 2012-2016). We broke down these estimates by geography (Texas, Houston metropolitan area), gender, and race/ethnicity, applying the survey weights. The CPS data were made available through the Integrated Public Use Microdata Series’ online database housed at the University of Minnesota.

---

1 Students who were attending neither a postsecondary institution nor working might have pursued a high school diploma, General Equivalency Diploma (GED), or Certificate of High School Equivalency. Due to limitations with the high school graduation data and the unavailability of data on the GED and related certificates, we did not compute estimates of this category. These students are included in the “not in the labor force” category. Based on Current Population Survey data, the share of students pursuing a high school education is very small.

2 The “not in the labor force” category included individuals who opted out of the labor market for reasons other than schooling (e.g., stay-at-home parenting, illness) and students pursuing a high school education.

• *Education Longitudinal Study of 2002 (ELS)*: The ELS was a nationally-representative, longitudinal survey administered by the National Center for Education Statistics. The study first surveyed high school sophomores in spring 2002 and had follow-up surveys in 2004, 2006, and 2012. The data contained detailed information on students’ background characteristics and educational attainment. Since ELS tracked students over time, regardless of whether they moved out-of-state, we could calculate inter-state migration rates and distinguish between students who left Texas to attend a college or university and students who left Texas for other reasons. We used the data to determine the share of students who moved out-of-state between 2004 and 2006, breaking it down by gender and race/ethnicity and applying the survey weights.\(^4\) We applied the one-half migration (0.5) scenario where it assumes rates of net migration one-half of those of the post-2000 decade. The reason for applying this scenario is that many counties in Texas, including Harris County, may not continue to experience the overall levels of relative extensive growth of the 2000 to 2010 decade. This scenario projects rates of population growth that are slower than 2000-2010 changes, but with steady growth. Also the 0.5 scenario is considered as the most appropriate scenario for use in long-term planning, according to the Texas Demographic Center.

• *U.S. Decennial Census, 2000 & 2010*: Census data were used to calculate the number of individuals living in Texas and Harris County by age, gender, and race/ethnicity and served as the denominator in estimates of military enlistment, incarceration, and death rates (described below). The Census data were made available through the National Historical Geographic Information System’s online database housed at the University of Minnesota.\(^5\)

• *Harris County Military Enlistment Records*: Through a recruiter in the Houston metropolitan area, we obtained aggregate military enlistment records for Harris County from 2015-2018. The records showed the number of individuals who enlisted broken down by age, gender, and race/ethnicity. Using the 2010 Census data as the denominator, we calculated enlistment rates for three age ranges: 18-22, 23-27, and 29-33. These rates were used for the 2003, 2008, and 2013 years, respectively.

• *Texas Department of Criminal Justice (TDCJ)*: The TDCJ data contained information on individuals in state correctional facilities by age, gender, and race/ethnicity. Given limitations, we restricted the sample to persons born between 1981 and 1985 who entered prison between 2010 and 2015 and were still in prison in 2015.\(^6\) Using the 2010 Census

---

\(^4\) Due to data limitations, we assumed the same migration rate in 2003, 2008, and 2013.


\(^6\) We were unable to observe individuals who entered and left the prison system in earlier years.
data as the denominator, we calculated incarceration rates in Texas and Harris County for three age ranges: 18-22, 23-27, and 29-33. These rates were used for the 2003, 2008, and 2013 years, respectively.


Additional details on these datasets and the data processing are available upon request.

**Educational Attainment, Spring 2004 – Spring 2015:** These cumulative estimates show the percentage of students who earned a postsecondary credential from a Texas college or university by the spring of a given year. The categories include the following:

- No postsecondary credential
- Certificate, diploma, or license
- Associate’s degree
- Bachelor’s degree
- Master’s degree
- Doctorate or professional degree (JD, MD)

In addition to presenting statistics on student activities and educational attainment for the sample at large, we have broken down these numbers by gender, race and ethnicity, and socioeconomic status (SES). In terms of race and ethnicity, we present statistics for non-Hispanic white, non-Hispanic black, Hispanic (of any race), and non-Hispanic Asian students; we do not show statistics for Native American students due to their small sample size. The measure of SES we use is developed by the Texas Education Agency (TEA) and shows whether a student is “economically disadvantaged”—receiving free or reduced-price meals through the National School Lunch and Child Nutrition Program.

**Limitations**

While the data may be useful in informing stakeholders working to improve postsecondary opportunities and, ultimately, the economic well-being of our communities, they do have some

---

7 This category includes individuals without a high school diploma, GED, or related certificate.
drawbacks.

First, we decided to exclude approximately 10 percent of eighth grade students from the analysis due to data issues. Most of these students were missing identifiers to link to the THECB or TWC datasets. A small share was excluded because they appeared to be enrolled in multiple school districts at the same time. Given our focus on long-term trends to and through college and the workforce at the state, county, and district levels, we dropped these observations from the data and subsequent analysis. Because of this decision, our estimates may not align perfectly with estimates computed by other research organizations. Our estimates of educational attainment, for instance, may be biased slightly upward since it is likely that the students excluded from the analysis are somehow different (i.e., demonstrating lower levels of academic ability) than the students included in the analysis.

Second, as mentioned earlier, we only observed postsecondary education and labor force outcomes for individuals attending Texas institutions or working in-state. Although we used other data, non-linked sources to examine a variety of other activities, including living out-of-state, the numbers we provided should be interpreted as estimates prone to statistical bias and error. For example, none of the external datasets provided information at the school district level, while others provided information at state, metropolitan, and/or county levels only. We encourage policymakers, practitioners, and the general public to interpret these figures with caution.

In compliance with the UH ERC restricted-use data agreement, cells with few cases were required to be masked in order to protect students’ identities. In the data available for download, these cells were marked with an asterisk (*). In the bar charts presented on this website, they were rounded down to zero percent.