<u>Roots of STEM Success Quantitative Dataset Description</u>

The North Carolina Roots Dataset is a unique longitudinal dataset containing highly detailed information about North Carolina students' educational trajectories and experiences from middle school through college. The data take the format of a panel study, tracking the entire population of students from one particular cohort, the high school graduating class of 2004. Data collection began in 1999, when students were in the seventh grade, continuing through high school graduation in 2004. Most of the cohort entered one of the sixteen campuses of the University of North Carolina system in the fall of 2004, while others first attended community college and then transferred into a UNC system school.

Information was gathered through 2011 (7 years after their high school graduation) in order to provide ample time for students to finish their undergraduate degree, as some students were in college for longer than the standard four-year period, or did not proceed directly from high school to college and therefore were not able to finish their undergraduate studies by 2008. Because the data were collected from *all* students who attended middle school, high school, and/or college at a public institution in North Carolina, the Roots dataset provides information for the entire population of the state (rather than a representative sample). However, attrition occurred when students transferred out of state or to a private institution for middle school, high school, or college, or when they stopped attending school (i.e. dropped out of high school or did not pursue an undergraduate degree after high school). Because the Roots dataset was originally conceived to explain postsecondary outcomes through prior educational experiences, it was important to assess the attrition occurring when NC public school students left the state for college or chose to attend a private university. We found that this loss did not significantly alter the characteristics in the student population in terms of race and gender (see table below). Indeed, this is not entirely unexpected due to the extremely wide range of campuses available in North Carolina's public university system in terms of selectivity, area of specialty, size, location, status as a Historically Black College/University (HBCU) and Tribal College/University and the fact that approximately only 3% of high school graduates in NC stated they planned to attend an out-of-state institution.

The dataset was compiled from three separate sources, each of which contributed distinct information for a specific period in a student's educational career. The North Carolina Department of Public Instruction (NCDPI) oversees North Carolina's K-12 public school system and provides the data that were gathered while students were in middle and high school. NCDPI supplies information at the student level, teacher level, classroom level, school level and district level. Student-level information includes measures of achievement, such as standardized test scores (End-of-Grade [EOG] and End-of-Course [EOC] tests) and track placement; demographic information, such as age, gender, race/ethnicity, free/reduced-price lunch status, parents' education level, learning disability or gifted status, limited English proficiency (LEP) status, and transfer status; and students' responses to survey questions, such as expected level of educational attainment, amount of TV watched, amount of reading done for leisure, and participation in afterschool activities. Teacher-level information includes instructor's demographic background (gender and race/ethnicity) as well as educational background (level and area of degree attained, license and certifications, and tenure in the profession), and their perceptions of the school's working conditions (level of empowerment, quality of facilities and resources, leadership, professional

development, and time constraints). Importantly, our data does not allow for the possibility of matching each individual student with each teacher, therefore teacher-level information available in the ROOTS dataset is utilized at the high school level. Classroom-level information includes the composition of students within a specific classroom (by both gender and race/ethnicity). School-level information includes size, course offerings, school's Title 1 eligibility, urbanicity, school-wide achievement and behavioral measures (such as percentage scoring at or above grade level on EOGs/EOCs, and percentage suspended and expelled), composition measures (percentage receiving free or reduced lunch, percentage of each race/ethnicity and gender), and charter/magnet status. District-level information includes per-pupil local, state, and federal expenditures and what these expenditures were spent on, overall demographic information about principals (gender and race/ethnicity), as well as the district-wide measures of achievement and behavior that are available at the school level. The middle school data contains information for approximately 199,000 students in about 600 middle schools, while the high school data contains information for approximately 230,000 students in about 500 high schools.

The College Board is the nonprofit organization responsible for administering the SAT test, which is the standardized entry examination for most colleges in the United States, and is a required part of the application at most public universities in North Carolina. The College Board not only supplied SAT scores (a composite score of student's performance eon the critical reading and mathematics sections), but also provided student responses to survey questions concerning the number of courses taken in each subject area and whether those courses were regular level, honors or Advanced Placement (AP), extracurricular and sports participation, and college preferences (in terms of size, region, urbanicity, type of institution as well as intended major). This dataset includes information for approximately 43,000 students from about 360 high schools.

University of North Carolina General Administration (UNCGA) oversees the sixteen-campus public university system for the state. The system includes Appalachian State University, East Carolina University, Elizabeth State University, Fayetteville State University, NC Agricultural and Technical State University, North Carolina Central University, NC State University, UNC Asheville, UNC-Chapel Hill, UNC-Charlotte, UNC-Greensboro, UNC-Pembroke, UNC-Wilmington, UNC School of the Arts, Western Carolina University, Winston-Salem State University, and North Carolina School of Science and Mathematics (a UNCGA-administered high school which does not supply college-level data for the Roots dataset). UNCGA supplied two sets of information for the dataset: information that students provided while applying to any one of these colleges (this supplemented the information from NCDPI about secondary school performance) and educational information from students' university-level experiences. These former data include overall high school GPA and class rank. The latter includes declared and completed major, college courses taken and grades earned, and financial aid received. In addition, measures captured the demographic breakdowns of each campus and major by both race/ethnicity and gender. This dataset includes information for approximately 29,000 students, of which about 19,000 have College Board data, and for whom about 21,700 have NCDPI (middle and/or high school) data.

These three datasets were combined and anonymized by the North Carolina Education Research Data Center (NCERDC), which is located at Duke University's Center for Child and Family Policy. NCERDC

manages a large amount of educational data collected in the state of North Carolina, extending beyond what makes up the Roots dataset. Data managed by NCERDC have been and are currently utilized for a wide variety of studies investigating education and educational policies. For more information about NCERDC, see http://www.childandfamilypolicy.duke.edu/project_detail.php?id=35.

Demographics of NC class of 2004 compared to the group enrolling in UNC system colleges.

	Race						Sex	
	White	Black	Hispanic	Asian	AI	Other	Female	Male
NC public school 2004 graduates	62%	31%	3%	2%	1%	1%	50.5%	49.5%
NC public school 2004 graduates who entered UNC system	67%	25%	2%	4%	1%	2%	55%	45%