Merit Aid Scholarships and Human Capital Production in STEM: Evidence from New Mexico

February 19, 2018

Christopher Erwin[1]

ABSTRACT

The New Mexico Legislative Lottery Scholarship is a broad, "low-bar," state lottery-funded scholarship designed to increase access to higher education on behalf of New Mexico residents. The natural experiment of a state lottery scholarship is used to measure the effect of generous financial aid on major choice at New Mexico's flagship public university. A potential unintended consequence of state merit aid scholarships is to discourage the production of human capital in science, technology, engineering, and mathematics (STEM) fields. This may occur if students avoid more rigorous majors in order to increase the likelihood of scholarship retention. I find no evidence that the scholarship decreased the overall likelihood that a student first declares a STEM major or earns a STEM degree. There are significant effects when disaggregating by academic preparation: less-academically prepared entering freshmen are 6.8 percentage points (40 percent) less likely to initially declare a STEM major, while more-academically prepared entering freshmen are 12.1 percentage points (44.3 percent) more likely to initially declare a STEM major. No significant effects are found when examining whether a STEM degree was earned. Evidence suggests these effects are at least in-part due to compositional changes in the student body before and after the advent of the lottery scholarship.

JEL Classification: I22, I23, I24, H75

Keywords: merit aid, major choice, STEM

---

[1]Erwin: PhD Candidate and corresponding author, Department of Economics, University of New Mexico, MSC 05 3060, 1 University of New Mexico, Albuquerque, NM 87131. Tel: +1 5052775304. The author is solely responsible for any errors in this analysis.

1. INTRODUCTION

The introduction of broad, merit-based college scholarships in the 1990s created a natural experiment for measuring relationships between college costs and academic outcomes. State merit-based scholarships generally fund most if not all tuition for qualified resident students. State legislation establishing merit-based scholarships share several common goals: retaining talent in-state, increasing access to higher education by reducing financial burdens, and promoting timely completion. There is considerable variation in initial and continuing eligibility requirements across states. Researchers have cataloged how such programs affect enrollment and course taking behavior, and, more recently, degree completion. I analyze the effect of the New Mexico Legislative Lottery Scholarship (NMLLS), a uniquely "low-bar" merit-based scholarship, on student major choice. Specifically, this paper is interested in two related research questions. First, do generous, low-bar merit scholarships discourage students from choosing majors in science, technology, engineering, and mathematics (STEM)? Second, do such scholarships affect the number of STEM degrees produced?

The major focus of this paper is on the first research question. Since merit-based scholarships require students to maintain a set level of academic achievement in order to continue to receive aid, there are potential unintended consequences that may occur, including dissuading students from studying more difficult subjects, including those categorized as STEM. The consequences of this outcome may be significant to economic interests at both the state and national levels, as STEM occupations are often seen as major drivers of innovation, and well as key to economic growth.

Since 1993, 27 states have implemented merit-based scholarships, the first and most studied being Georgia's Helping Outstanding Pupils Educationally, or HOPE, scholarship program.[2] HOPE marked the beginning of what has been a major restructuring of the financial aid landscape in America. According to the College Board, from 1993 to 2013, the percentage of total undergraduate state grant aid for which students' financial circumstances were considered decreased from 90 percent to 76 percent. In the 2013-2014 academic year, New Mexico was one of 13 states where this percentage was below 40 percent.[3]

We know more about the relationship between financial aid, enrollment, and degree completion than financial aid and major choice. Different types of financial aid have varying effects on college enrollment. Loans tend to have little to no effect, while grants have a positive and significant effect on student enrollment (Linsenmeier *et al.* 2006). Students from low-income families and students of color seem to be most responsive to such aid. Van der Klaauw (2002) demonstrates that students' choice of college are sensitive to financial aid offers. Several studies show a significant and positive relationship between grant aid and student enrollment (Seftor and Turner 2002; Kane 2003; Heller 2009) and a negative relationship between net cost and enrollment (McPherson and Schapiro 1991). The effects of merit-based aid on enrollment have also been well documented. In an experimental setting, Monks (2009) finds large, positive effects of merit aid on enrollment. Studying HOPE, Dynarski (2000) finds that a $1,000 award increased student enrollment by approximately four percent. Also studying HOPE, Cornwell *et al.* (2006) find the program increased student enrollment by 6 percent. In New Mexico, Binder and Ganderton (2002, 2004) find that while the NMLLS boosted enrollment at four-year colleges in New Mexico, the effect appears to be driven by additional enrollment of students that

---

[2] See Sjoquist and Winters (2015a) for a complete list.
[3] The College Board, Trends in Student Aid 2015, Figure 28A and Figure 28B. Retrieved January 29, 2016 from http://trends.collegeboard.org/sites/default/files/trends-student-aid-web-final-508-2.pdf.

otherwise would have attended college out-of-state. The effect of merit aid on college completion has also been studied.

Analyzing statewide educational attainment data, Sjoquist and Winters (2012, 2015b) found no difference in college attainment for those exposed to lottery scholarship programs. Using a similar methodology, Jia (2017) found that program features matter: lower initial scholarship eligibility requirements increased two-year degree attainment, and funding generosity increased the completion of a bachelor's degree. Scott-Clayton (2011) found completion effects of 9.4 percentage points (59 percent) for students just above an ACT cut-off for West Virginia's lottery-funded PROMISE scholarship program, compared to students just below. Using similar strategies, Bruce and Carruthers (2014) and Welch (2014) found no program effect for Tennessee's lottery scholarship. Erwin and Binder (see Chapter 2) found no overall effect of generous, low-bar merit aid on college completion. Divergent effects appeared when disaggregating the sample by academic preparation. Less-academically prepared students appeared to exhibit lower completion rates as a result of the scholarship while more-academically prepared students exhibited higher completion rates, two significant effects similar in magnitude but opposite in sign. The authors argue that changes in student composition are potentially driving results.

I examine how the NMLLS affects STEM engagement at the University of New Mexico (UNM) by exploring changes in 1) the likelihood of initially declaring a STEM major and 2) the likelihood of earning a baccalaureate degree in a STEM field before and after the implementation of the scholarship for eligible resident students and a matched sample of nonresident (and therefore ineligible) students. Estimates reveal no significant overall effect of the NMLLS on declaring a STEM major or earning a STEM degree. However, there are large and statistically significant completion effects after disaggregating by academic preparation. Academically less-prepared eligible freshmen are 6.8 percentage points (40 percent) less likely to first declare a STEM major, while academically more-prepared freshmen are 12.1 percentage points (44.3 percent) more likely to first declare a STEM major, compared to ineligible peers with similar high school GPAs. In addition, there is evidence that some program effects may be a result of the NMLLS inducing compositional changes in the student body.

The paper proceeds as follows: Section 2 discusses existing literature regarding merit-aid and major choice, and introduces the NMLLS; Section 3 presents a theoretical model of major choice; Section 4 describes the data; Section 5 summarizes the empirical approach; Section 6 discusses main findings and robustness checks; Section 7 discusses other explanations for patterns found in the results; and Section 8 concludes.

2. MERIT-AID AND MAJOR CHOICE

The natural experiment of lottery-financed merit-based aid programs provides a promising avenue for determining the relationship between aid and major choice. Several studies have analyzed how students sort into different majors. An early study of this behavior can be found in Berger (1988). Berger uses a life cycle approach that assumes students choose majors based on the expected discounted stream of future earnings rather than beginning wages following graduation. The author provides evidence to support this approach using data from the National Longitudinal Survey of Young Men. Montmarquette *et al.* (2002) relax two assumptions common in previous literature, including Berger (1988): uniform probabilities of success across majors and constant earnings streams across majors. Using the National Longitudinal Survey of Youth, the authors estimate the probability of success across different majors for all students in the sample. These data are combined with estimates of predicted future earnings in all majors from Rumberger and Thomas's (1993) analysis of the 1987 Survey of Recent College Graduates, which allows the construction of a multinomial logit model of major choice. Results suggest that one's expected earnings stream is the most significant factor influencing major choice, yet the probability of success is an important factor as well. Arcidiacono *et al.* (2012) argues that both expected earnings and students' perceived abilities across majors are important determinants of major choice.

Arcidiacono (2004) estimates a dynamic model of college major choice, finding that even after controlling for selection, large earnings premiums and ability differences still exist for some majors. Differences in monetary returns explain little of the ability sorting across majors. Instead, Arcidiacono (2004) provides evidence that virtually all ability sorting is due to differences in preferences for taking particular majors in college and workplace preferences for jobs likely to be obtained after graduation, the former being more influential than the latter. Similarly, Beffy *et al.* (2010) find a small, but statistically significant, positive earnings elasticity of major choice, suggesting that nonpecuniary factors are a large part of major choice (e.g., preferences for workload, workplace conditions, opportunities field research, *et cetera*.).

Focusing on STEM fields, Wang (2013) finds that choosing a STEM major is positively related to high school performance, as well as initial college performance/experiences. Similarly, Griffith (2010) finds that differences in academic preparation and educational experiences drive differences in persistence rates in STEM majors. Wiswall and Zafar (2015) find that while expected earnings and perceived ability play a major role in choosing STEM, unobserved tastes are the largest factor in major choice. Henry and Rubenstein (2002) argue that merit aid may result in greater effort on behalf of high school students, thus better preparing students for difficult majors such as those included in STEM.

Four studies directly examine the relationship between merit aid and major choice. Analyzing Georgia's HOPE Scholarship, Cornwell *et al.* (2006) use administrative data to compare qualified residents and nonqualified nonresident students in a difference-in-differences framework. The authors find that HOPE resulted in a small 1.2 percentage point increase in the likelihood that residents chose education majors, relative to their nonresident counterparts. Cornwell *et al.* (2006) do not find any meaningful change in the likelihood that students chose STEM majors due to the advent of HOPE. Both Dynarski (2000) and Cornwell *et al.* (2006) find evidence that state merit-based scholarships increase the likelihood that highly-academically prepared students stay in-state for college, and thus affect the type and quality of institutions attended. This implies that crowding out of moderately-academically prepared students may occur as competition increases within more difficult majors.

Using Integrated Postsecondary Education Data System files, Zhang (2011) examines whether Georgia's HOPE Scholarship and Florida's Bright Futures Scholarship affected the likelihood that students embarked on a course of study within STEM fields. Zhang uses differences-in-differences estimation for aggregate state data, and finds a statistically significant 1.6 percentage point (11.4 percent) increase in the proportion of degrees classified as STEM at private institutions in Florida, but no broader effect of merit aid on STEM degree completion in either Florida or Georgia. Two significant problems should be noted with this approach. Since the unveiling of such programs affects how students sort into institutions, it is difficult to distinguish compositional change from real program impact (see Chapter 2, for example). Also, asymptotic refinement should be applied in situations where there are relatively few treated units or policy changes in a difference-in-differences framework (Conley and Taber, 2011).

Stater (2011) uses administrative data from three large public universities to examine the relationship between tuition and financial aid on the first major a student declares. He finds that larger net tuition results in students being more likely to choose professional fields such as architecture, business, or law and less likely to declare majors in humanities and sciences. Merit aid was shown to increase the likelihood of declaring majors in humanities and sciences, while having a negative effect on social sciences. It is difficult to view these estimates as causal, however, since Stater does not address the endogeneity of merit aid: students that receive merit aid are better academically prepared for college. Thus, recipients may be more likely to choose STEM majors for reasons other than merit aid.

A recent paper regarding the relationship between merit aid and major choice comes from Sjoquist and Winters (2015a). Their analysis relies on a difference-in-differences strategy using American Community Survey (ACS) microdata. They assign treatment status to individuals that were 18 years of age in a state where a merit aid program was in place, with all others assigned to the control group. Sjoquist and Winters divide the 27 adopting states into "strong" and "weak" merit aid state categories, based on their judgement of how broad-based programs are and how much funding they provide. New Mexico is defined as a strong merit aid state. Findings suggest that state merit-based scholarships reduce the rates of STEM completion. Overall, strong merit aid programs (from 9 states) were found to reduce the number of male STEM graduates by 8 percent, with no meaningful impact on women in the sample. The overall impact of merit aid on the production of STEM degrees is estimated to be -6.5 percent. The authors argue that men may be more willing to switch majors in order to retain a merit-based scholarship. Weak merit aid programs were not found to have any effect on STEM degree completion. There are several notable weaknesses in Sjoquist and Winters (2015a). The authors are also not able to control for student-level characteristics, which is important as merit aid may result in changes in student composition. Also, as noted in Jia (2017), program features matter, and vary considerably across programs. With this in mind, approaches which treat all state merit-aid programs as homogeneous are problematic.

Literature on the relationship between merit aid and major choice is not in agreement, but the most dependable studies suggest either null or negative effects on STEM degree completion. In this study, I employ a rich administrative data set from New Mexico's flagship university to revisit this question and others. The main contribution to the literature is that I control for, and disaggregate by, student-level characteristics, which allows for more detailed insight into the effects of merit aid on subpopulations. Cornwell *at al.* (2006) control for high school GPA, but do not split the sample as I do, so it's difficult to interpret how academic preparation impacts major choice. I also consider how compositional changes in academic preparation of the student body play an important role in interpreting results.

2.1 NMLLS program details

The NMLLS, established by the New Mexico Legislature in 1996, first became available to students in fall 1997. New Mexico residents qualify for the NMLLS if they earn a high school diploma or general educational development equivalency in New Mexico and enroll at a public postsecondary institution in the first regular fall or spring semester following high school graduation. Most state lottery scholarship programs reward high school achievement and begin with the first semester of college enrollment. In New Mexico, however, students become eligible for full tuition at any of the 16 qualified public two- or four-year colleges after they complete a full-time course load (at least 12 credits) with a 2.5 GPA or higher in their first college semester. To encourage students to try for the scholarship, New Mexico colleges offer students “Bridge to Success” scholarships which completely or mostly offset tuition in their first semester. In the period examined, students could receive the award for up to eight semesters, provided they enroll full-time, continuously, and maintained a cumulative 2.5 GPA. Only 58 percent of first semester students over 1994-1999 met NMLLS requirements, and only 30 percent remained eligible at the end of their second year.

Before the NMLLS, New Mexico nearly exclusively awarded financial aid based on need. According to a 1994 National Association of State Student Grant & Aid Programs report, New Mexico devoted an average of $222 per full-time equivalent (FTE) undergraduate student in financial aid in the 1993-1994 academic year. Of the $222 total per FTE, only $3 (1.4 percent) was merit-based. By contrast, in 2000, New Mexico allocated $687 per undergraduate FTE, with $368 (54 percent) being merit-based. It appears the NMLLS not only supplemented rather than supplanted student aid, but drastically changed the student aid landscape throughout the state.

Compared to states with similar programs, NMLLS eligibility requirements are relatively “low-bar.” For example, Georgia’s HOPE scholarship requires students to graduate high school with a 3.0 cumulative GPA and maintain a 3.0 GPA in college.[4] Eligibility for Tennessee’s HOPE scholarship requires minimum ACT/SAT scores in addition to the 3.0 high school GPA requirement. Renewal requires a 2.75 minimum overall GPA after attempting 24 and 48 credit hours, and requires a 3.0 minimum overall GPA at 72- and 96-credit hour reviews.[5] Florida’s

---

[4]Georgia Student Finance Commission, GACollege411, Georgia Hope Scholarship Program Overview. Retrieved May 29, 2013 from https://secure.gacollege411.org/Financial_Aid_Planning/HOPE_Program/Georgia_s_HOPE_Scholarship_Program_Overview.aspx.

[5]Tennessee Student Assistance Corporation, Tennessee Hope Scholarship. Retrieved May 29, 2013 from http://www.tn.gov/collegepays/mon_college/hope_scholar.htm.

Bright Futures Scholarship has three levels of merit-based awards, each with varying high school GPA, standardized test scores, and community service requirements.[6]

If financial constraints are binding for students, then the NMLLS should have the desired effect of increasing access to higher education and boosting resident enrollment. However, due to low-bar initial and ongoing scholarship qualifications of the NMLLS, much of the increase in resident enrollment may be on behalf of less-academically prepared students who otherwise would have enrolled at a less prestigious university, a two-year program at a community college, or perhaps not have enrolled in college at all. With price signals in the market for higher education removed, some students may choose to embark on a more prestigious, yet riskier, academic path—one that maximizes the "worth" of the scholarship (i.e., that which covers the largest cost).[7] Because the NMLLS is structured so that students lose the scholarship permanently if they fail to meet renewal requirements in any semester, some students may respond to merit aid by choosing easier majors which improve their chances of scholarship retention. In this case, the NMLLS could have the unintended consequence of decreasing the proportion of students choosing and ultimately completing degrees in STEM fields. However, as discussed above, renewal requirements for the NMLLS are relatively low. If students expect their probabilities of success in STEM majors are sufficiently to satisfy eligibility criteria then students may not avoid pursuing majors in STEM. The direction of any estimated program effects relies heavily on the academic preparation of resident students before and after the

[6]Florida Department of Education, Office of Student Financial Aid, Florida Student Scholarship and Grant Programs, Chart of Eligibility and Award Criteria. Retrieved May 29, 2013 from http://www.floridastudentfinancialaid.org/ssfad/PDF/BFEligibilityAwardChart.pdf.

[7]Consider full-time tuition at all 16 participating public institutions in New Mexico as depicted in Table 1. A student better matched at Santa Fe Community college may decide to attend UNM instead simply because the scholarship covers more costs, the degree carries more prestige, and thus the NMLLS is "worth" more at the state's flagship university.

introduction of the NMLLS, and is ultimately an empirical question. Composition of the student body at UNM is discussed in detail below.

3. MERIT AID AND MAJOR CHOICE

Students' choice of college major are modeled using a modified life-cycle approach developed by Montmarquette *et al.* (2002). This approach holds that students choose college majors so as to maximize lifetime utility, which depends on expected earnings and the likelihood of merit scholarship retainment. For simplicity, assume that students who are able to retain merit aid ultimately graduate with a bachelor's degree. Let p_{ij} be the likelihood of scholarship retainment for student i in major j. The expected lifetime utility for student i choosing major j, $E(U_{ij})$, is a function of predicted future earnings so that:

(1) $$E(U_{ij}) = p_{ij}(\boldsymbol{X})e_{ij}(\boldsymbol{Z}) + \left(1 - p_{ij}(\boldsymbol{X})\right) e_{i0}(\boldsymbol{Z}), \;\; i = 1, \dots, N; j = 1, \dots, m,$$

where **X** includes factors influencing the probability of retaining the scholarship, including academic preparation. The vector **Z** includes factors affecting earnings after college. e_{ij} are the discounted value of lifetime earnings after completing a degree in major j and e_{i0} are discounted value of lifetime earnings after losing the scholarship and dropping out of college without a degree. Students will choose major j over major k whenever $E(U_{ij}) \geq E(U_{ik})$ for all $k \neq j$, or whenever,

(2) $$p_{ij}(\boldsymbol{X})\left[e_{ij}(\boldsymbol{Z}) - e_{ik}(\boldsymbol{Z})\right] + \left[p_{ij}(\boldsymbol{X}) - p_{ik}(\boldsymbol{X})\right]\left[e_{ik}(\boldsymbol{Z}) - e_{i0}(\boldsymbol{Z})\right] \geq 0.$$

According to equation (2), if the likelihoods of retaining merit aid differ substantially across majors, and lifetime earnings differences across majors are relatively small, then success probabilities will play a major role in major choice. If likelihoods of scholarship retainment are approximately the same, then expected earnings will be the major driver in the choice of major. Thus, one would expect highly-academically prepared students, whose likelihoods of retaining the merit scholarship are high across all majors, to be more likely to choose majors based on which has the highest expected return (i.e., STEM).[8] For less-academically prepared students, I assume the likelihood of retaining the merit scholarship is lower for some majors relative to others, thus these students choose majors primarily on the basis of success probabilities, and choose majors which are less difficult. Importantly, this simplified model does not account for tastes and preferences of students, which the literature has indicated plays an important role in major choice (Arcidiacono, 2004; Beffy *et al.*, 2010; Wiswall and Zafar, 2015).

In the context of whether broad, low-bar merit scholarships such as the NMLLS affect student major, the theoretical framework above suggests that more-academically prepared students will tend to embark on more difficult, higher-paying majors such as those in STEM fields, while less-academically prepared students will tend to avoid such majors in favor of less-difficult majors, such as those within education and the liberal arts, for example.

4. DATA SET

The analysis uses administrative data for all first-time, full-time entering freshmen at UNM before and after the implementation of the NMLLS to estimate effects on major choice. UNM enrolls over 20,000 students each year in the City of Albuquerque, the largest metropolitan area of the state with over 500,000 residents. UNM is nearly an open-enrollment institution. Data include socio-demographic information (age, race, ethnicity, gender, family income,

---

[8]Sjoquist and Winters (2015a) calculate mean earnings for persons aged 40 – 49 using 2009 – 2011 ACS data, finding that those majoring in STEM fields earned $95,389; those with business degrees earned $78,122; those in social science earned $67,735; those with health-related degrees earned $58,937; those with liberal arts degrees earned $58,823; and those with degrees in education earned $46,169. They choose this age range because 1) these respondents are too old to be affected by state merit aid programs and 2) according to Berger (1988) mid-career earnings are likely to be more relevant than early-career earnings.

declined to state race-ethnicity), high school academic performance (high school GPA, standardized test scores, indication of remedial coursework at UNM), and college academic outcomes by semester (credits earned, declared major, college GPA, date of graduation). Majors are categorized into five areas using ACS definitions: STEM, liberal arts, education, business, social science, and health-related.[9] I also consider alternative definitions provided by UNM as a robustness check. Data are complete with the exception of family income and high school GPA. The data set only contains family income for FAFSA-filers, constituting 51 percent of the sample. For those that did not submit a FAFSA, it is assumed their family income is sufficiently high (i.e., ≥ $40,000) as to not qualify for the Federal Pell Grant Program. This assumption is supported by a 1995-1996 Federal Pell Grant End-of-Year Report showing that less than two percent of Pell recipients had family income in excess of $40,000.[10] This assumption is not perfect. King (2004) estimates that in 2000 over ten percent of all Pell-eligible students did not file a FAFSA.[11] If the analysis in King (2004) holds for our data set, then there would exist systematic measurement bias in the family income variable—some lower income students would be incorrectly placed in the higher income category. High school GPA is missing for home-schooled students, a small portion of matriculating students at UNM. For these students, they are assigned the mean high school GPA of 3.28.

Models concentrate on the years 1994 to 1999, bounding the policy change by three years before and after implementation. These years encompass the largest economic expansion in the

---

[9] Majors are categorized into these bins according to the U.S. Census Bureau, found online at https://www2.census.gov/programs-surveys/acs/tech_docs/code_lists/2016_ACS_Code_Lists.pdf (accessed 19 Feb 2019).

[10] 1995-1996 Federal Pell Grant Program End-of-Year Report, U.S. Department of Education, online at https://www2.ed.gov/finaid/prof/resources/data/pell-historical/pell-eoy-1995-96.pdf (accessed 26 March 2017).

[11] King, Jacqueline E. "Missed Opportunities: Students who do not Apply for Financial Aid," American Council on Education Issue Brief, 2004. Online at http://www.soe.vt.edu/highered/files/Perspectives_PolicyNews/10-04/2004FAFSA.pdf (accessed 1 April 2017).

U.S. since World War II. During this period labor market conditions in New Mexico were gradually tightening but remained relatively stable, so one needs not to worry much that broad economic conditions are driving results. To my knowledge, there were no concurrent policy changes at the high school or postsecondary level in New Mexico over the 1994-1999 period which would have differentially impacted enrollment and/or major choice for residents and nonresidents.

In preferred specifications, recent high school graduates from New Mexico (who are NMLLS eligible) are compared with those from out of state (who are not eligible, but experience the same campus environment), while excluding foreign students.

Table 2 compares summary statistics for resident and nonresident students before and after the implementation of the NMLLS. It appears the composition of these groups changed across pre- and post-treatment periods. In years before the implementation of the NMLLS, resident students had higher high school GPAs and ACT composite scores compared to years following the implementation of the scholarship. Moreover, students matriculating after implementation were more likely to take remedial coursework at UNM. These changes are statistically significant, suggesting that the NMLLS may have induced students with weaker academic preparation to enroll at UNM. Table 2 also shows that residents were less likely to come from lower-income families following implementation of the NMLLS, another indication of a compositional effect. The academic achievement of nonresident students improved following implementation of the scholarship, according to HSGPA and composite ACT scores. Also note the statistically significant decline in resident students initially declaring a STEM major—a decline not seen in the nonresident group. Table 3 presents descriptive statistics for those earning a degree at UNM during the study period. Note there is less evidence of a compositional change in resident students, with only a small decline in high school GPA. For degree earning residents, there is no descriptive evidence of a decline in STEM degree production after the initiation of the NMLLS.

Although several statistically significant differences exist between resident and nonresident students in terms of high school GPA, composite ACT scores, remedial coursework, family income, race, and ethnicity, this does not threaten the validity of our difference-in-differences model of STEM engagement if the common trends assumption holds. The identifying assumption of the difference-in-differences model is that pre-treatment trends in the outcome variable be similar in trajectory across treatment and control groups. As a visual check of this identifying assumption, Figure 1 presents pre-treatment trends in the likelihood of declaring a first major in STEM for residents and nonresidents between 1994 and 1999. Visual inspection supports the validity of a difference-in-differences identification strategy examining six-year graduation rates. Figure 2 presents pre-treatment trends in the likelihood of earning a STEM degree for residents and nonresidents over the same time period. Because completion rates at UNM are relatively low, there are far fewer observations for this group and consequently the graph is quite noisy, especially for nonresidents who are greatly outnumbered by resident students at UNM (by nearly 11 times over). Although Figure 1 seems reasonably comparable before the NMLLS was launched in 1997, Figure 2 does not pass visual inspection. An empirical test of the common trends assumption is conducted following Autor (2003). Autor suggests estimating flexible difference-in-differences models by interacting the resident dummy variable with cohort dummy variables, producing a model allowing for treatment at different time periods. This model can be expressed as

$$Prob(STEM_{ist}) = \gamma_s + \lambda_t + \sum_{j=-m}^{q} \beta_j \, D_{st}(t = k + j) + X_{ist}\delta + \varepsilon_{ist} \qquad (3)$$

where i denotes the student, s denotes residency status, and t denotes cohort year. The variable D_{st} is the binary treatment indicator and k is the year which the treatment started (k = 1997 in this case). X_{ist} contains controls for race, ethnicity, gender, family income, remedial coursework in college, high school GPA, and standardized test scores. Models report robust standard errors. In equation (3), m and q are the number of leads and lags of the treatment effect included. Two leads and three lags are included in the test, defining 1999 as the reference cohort.

Testing the common trends assumption using (3) requires examining whether

$$\beta_j = 0 \; \forall j < 0. \qquad (4)$$

In other words, the common trends assumption holds when the coefficients on all leads of the treatment are zero. This specification can also have the advantage of informing whether estimated treatment effects occur in multiple post-treatment time periods, fade away with time, or remain constant, for example. Tests are conducted for the two STEM outcomes using ordinary least squares and results are presented in Appendix A. Results provide evidence that the common trends assumption holds for all specifications, as estimated coefficients on all leads are not statistically different from zero.

Data include 10,022 resident students, 6,307 of which enrolled during the post-NMLLS period and were eligible for the Bridge to Success Scholarship. Of these, 2,664 met cumulative GPA and credit attainment requirements to begin the NMLLS in their second semester. Table 4 documents the number of students that maintain the scholarship in the second through ninth semester. It is apparent scholarship loss was quite common. Of the 2,664 students that qualified for the NMLLS, approximately 30 percent were still eligible for the NMLLS going into their third year.

5. EMPIRICAL MODEL

Difference-in-differences matching estimation on the propensity score is conducted to mitigate any observable differences between resident and nonresident students. The approach uses kernel matching, a one-to-many matching technique assigning larger weights to control units closer in propensity score. The general form of the matching estimator is given by

$$\Delta^{DDME} = \frac{1}{n_{1t}} \sum_{i \in I_{1t} \cap S_p} \left\{ Y_{1ti} - \sum_{j \in I_{0t} \cap S_p} W(i,j) Y_{0t} \right\} - \frac{1}{n_{1t'}} \sum_{i \in I_{1t'} \cap S_p} \left\{ Y_{1t'i} - \sum_{j \in I_{0t'} \cap S_p} W(i,j) Y_{0t'j} \right\} \qquad (5)$$

where n_{1t}, $n_{1t'}$ are the number of treated cases before and after the inception of the NMLLS, S_p is the common support region, and I_{0t}, $I_{0t'}$, I_{1t}, $I_{1t'}$ are the resident and nonresident groups before and after the NMLLS. Major choice outcomes for resident and nonresident students are given by Y_{1t}, Y_{0t}, $Y_{1t'}$, $Y_{0t'}$. The function $w(i,j)$ denotes the weight given to jth case, where $\sum_j w(i,j) = 1$ and $0 < w(i,j) < 1$. The weighting function $w(i,j)$ is given by

$$w(i,j) = \frac{K[\hat{l}(x_j) - \hat{l}(x_i)]}{\sum_{j \in I_{0t} \cap S_p} K[\hat{l}(x_j) - \hat{l}(x_i)]} \qquad (6)$$

where K is the Epanechnikov kernel function and $\hat{l}(\cdot) \equiv \ln\left(\frac{\hat{p}(\cdot)}{1-\hat{p}(\cdot)}\right)$ is the fitted linearized propensity score from a logistic regression model estimated by maximum likelihood. Linearized propensity scores are used as they are more likely to have a distribution that is approximately normal. Treatment effects, Δ^{DDME}, are calculated using kernel-weighted least squares according to equation (6). Robust standard errors are reported. The propensity score model includes all covariates in levels, as well as several quadratic terms.[12] Results of the propensity score model

---

[12]A sequential search for quadratic terms to include in the propensity score model was conducted. The first step involved estimating logistic models including all terms in levels and one of all possible quadratic terms. I then calculate the likelihood ratio statistic for the null hypothesis that the most recently added quadratic term has a coefficient of zero. The quadratic term with the highest test statistic over 2.71, corresponding to a z-statistic of 1.645, is selected for inclusion. This covariate is then added to the “baseline” model and the process repeated until all remaining likelihood ratio statistics are below the threshold of 2.71.

are presented in Table 5. It is important to note that while the propensity score model may seem awkward in that it predicts the immutable condition of being a New Mexico resident, it is not essential that the propensity score model have a meaningful interpretation. Instead, the validity of the propensity score model rests on how well it balances covariates across treatment and control groups (Imbens and Rubin, 2015; Imbens, 2015).

Having a small group of nonresident students relative to resident students has implications for the estimates. In order to increase the precision of estimated treatment effects, and to avoid imposing functional form where possible, kernel density matching is chosen.[13] This method has the advantage of lower variance since more information is used. On the other hand, it may result in an increase in bias due to the potential for considering "bad" matches. Although the further the observations are in terms of propensity score, the less weight is given to the potential bad match, this makes adequate overlap a necessary condition for the validity of this method.

In our analysis, matching is limited to those individuals whose propensity scores lie in the common support region, which is over 99.5 percent of the original sample. No observations are trimmed from the analysis. As a sensitivity analysis, effects are estimated using various fixed bandwidths, h, in the kernel function. Importantly, the choice of bandwidth also involves a bias-variance trade-off. Smaller bandwidths consider a smaller portion of the pool of control observations, and thus use less information, which tends to reduce bias (from being less likely to consider poor matches) while increasing sampling variance. In order to assess the effectiveness

[13] There are 9,979 resident students and only 1,233 nonresident students in the sample. One-to-many matching allows us to proceed without a significant loss in information. For example, if I was to conduct a simple nearest neighbor matching procedure, estimates would (at most) be based on 1,233 matches, or 2,466 observations, which constitutes approximately 22 percent of the sample.

of the matching procedure, several tests are conducted following Imbens and Rubin (2015), although they are modified for difference-in-differences matching with repeated cross sections. An explanation of these tests and their results are presented in Appendix B.

In addition to estimating the overall effect of the NMLLS, I am also interested whether program effects differ depending on academic preparation. This is explored by estimating separate models on students above and below the mean high school GPA.[14] Robustness checks using various STEM definitions, cohorts, and smoothing parameters are discussed in Section 6.1.

While difference-in-differences models hinge on the comparability of pre-treatment trends in outcomes across residents and nonresidents, combining difference-in-differences methods with propensity score matching controls for compositional changes in groups over time (Stuart *et al.* 2014). It is also worth noting that regressions control for high school achievement and standardized test scores, the main indication of compositional change. Also, because UNM is a *de facto* open enrollment institution, changes in selectivity are not likely to confound the analysis (Binder and Ganderton, 2004). It is clear that compositional change in the student body occurred, yet this does not diminish the validity of treatment effects estimated.

6. RESULTS

Means and normalized differences after kernel matching are presented in Table 6. Comparing means before and after the NMLLS, it appears that the matching algorithm performed well in balancing covariates. Normalized differences for pre- and post-NMLLS periods are near zero, with the largest normalized difference (-.122) far below one-quarter of a standard deviation unit in absolute value. These statistics are produced by academic preparation

---

[14] Results are similar when the sample is split around the median high school GPA.

as well, revealing a similar pattern, although differences were slightly higher when considering students more than one standard deviation above the mean high school GPA. Overall, normalized differences suggest excellent balance in covariates following kernel matching.

Table 7 presents results of the difference-in-differences kernel matching estimation. Results provide no evidence of an overall effect on either first declaring a STEM major or earning a STEM degree. Furthermore, there is no evidence suggesting the NMLLS had an impact on earned STEM degrees when the sample is disaggregated by academic preparation. Considering students' decisions to first declare a major in STEM, there appears to be a divergent effect: students with below average academic preparation are 6.8 percentage points (40 percent) less likely to declare their first major to be in a STEM field, while those with above-average academic preparation are 12.1 percentage points (44.3 percent) more likely to declare a first major in STEM. Effects are significant at ten and one percent-levels, respectively. These divergent effects mask any overall program effect of the NMLLS on declaring a first major in STEM.

In summary, results reveal no meaningful impact on first declaring a STEM major or earning a STEM degree in the aggregate. In terms of declaring an initial major in STEM, I find that less-academically prepared students are averse to doing so. Conversely, I find that more-academically prepared students declare initial majors in STEM at higher rates compared to their nonresident counterparts as a result of the scholarship.

6.1 Alternative STEM definitions, smoothing parameters, and freshmen cohorts

Robustness checks are conducted to examine the sensitivity of results to various assumptions. Appendix C offers three different definitions of STEM based on the student's major. Table C1 presents STEM majors from the ACS, our preferred categorization scheme.

We prefer this set of STEM majors as it was developed by the U.S. Census Bureau, is sufficiently narrow in scope, and is the most comprehensive list that can be found. Further, it is employed by previous literature which we are keen to compare our results to (Sjoquist and Winters, 2015a). Tables C2 and C3 present alternative lists of STEM majors compiled by the STEM Collaborative Center (SCC) at UNM. Table C2 presents the “broad” list of STEM majors compiled by SCC while C3 presents the “narrow” version. The broad list is problematic because it includes many majors which one may not agree qualify as being designated as STEM, including anthropology, economics, geography, and nursing. The narrow list should be a subset of the broad list put out by SCC, yet it is not. For example, the narrow list includes statistics while the broad list does not. I nonetheless run models of STEM major declaration and STEM degree completion using broad and narrow lists from SCC. Appendix D displays results of these regressions.

Table D1 presents estimates using the narrow STEM definition provided by the SCC. Aggregate results and those disaggregated by academic preparation are shown. In general, point estimates are similar to our preferred results using the ACS definition, but are attenuated in both magnitude and statistical significance. Using the narrow definition, the point estimate for first majoring in STEM for less-academically prepared students remains negative, but is no longer precisely estimated. The point estimate for first majoring in STEM for more-academically prepared students is still positive, yet the magnitude is smaller and it achieves statistical significance at a lower level. Table D3 is structured just as other results tables, but employs the broad list from SCC. One would expect the broader scope of this definition to result in further attenuation in terms of magnitude and statistical significance, which it does with one exception. Results using the broad definition estimate a large statistically significant decline in STEM major declaration for the most-academically prepared entering freshmen, although the point estimate is significant only at the ten percent-level.

In addition to examining the sensitivity of results to various definitions of STEM, it is also imperative to examine whether results are sensitive to the choice of smoothing parameter used in the kernel matching procedure. Appendix E presents such sensitivity tests. According to test performed in Appendix B, the matching procedure performed remarkably well. This is further evidenced by Table E1 where one notes that only a few additional observations are included when increasing the bandwidth from 0.1 to 0.3. Point estimates using bandwidths of $h = \{0.1, 0.2, 0.3\}$ are remarkably close in magnitude and statistical significance. There are no sign changes when varying the bandwidth across these values. This provides evidence that bandwidth choice is not a significant driver of our main results presented in Table 7.

Appendix F presents results using different sets of freshmen cohorts. Although a bit noisier than robustness checks using alternative bandwidths, we see a similar pattern of completion rates emerge as compared to our preferred specification. Some coefficients become imprecisely estimated when including either the 1993 cohort, the 2000 cohort, or both.

7. COMPOSITIONAL EFFECTS

Key results from Section 6 are not entirely in agreement with the most thorough treatment of this subject to date. Estimates in this paper reveal no meaningful effect of the NMLLS on the likelihood that students earn degrees in STEM fields, in contrast to Sjoquist and Winters (2015a). However, results also provide no evidence that merit aid decreases students' likelihoods of majoring in STEM, in agreement with Cornwell *et al.* (2006). Further, estimates provide evidence of negative STEM degree effects for men, with no statistically meaningful effects for women in the sample, in-line with findings in Sjoquist and Winters (2015a).[15] It is valuable to entertain compositional effects as an alternative hypothesis for the results obtained.

The NMLLS was designed to increase access to higher education for resident students, which it certainly did. According to Table 2, the post-NMLLS period of the sample showed a resident population increase of 70 percent (with a much smaller 11 percent increase in nonresidents). After the NMLLS was introduced, however, resident high school GPAs and standardized test scores fell significantly, and resident students were required to take more remedial courses at UNM. This apparent change in student composition is likely key to interpreting much of the results found in Table 7. According to the theoretical model presented in Section 3, academically marginally prepared students are likely to respond to merit aid by choosing majors for which their probability of success is higher. This may explain why results show that less-academically prepared students majored in STEM significantly less in response to the NMLLS. On the other hand, theory predicts that more-academically prepared students have high probabilities of success in all majors, and so are likely to choose majors with higher expected lifetime earnings, such as STEM.

8. CONCLUSIONS

I examine the effect of an exceptionally generous and low-bar merit-based scholarship on intitially declaring a major in STEM and ultimately earning a degree in STEM. Variants of the difference-in-differences model are estimated using qualified resident students as the treatment group and a matched sample of ineligible nonresident students as the control group. The

[15]Although we estimate regressions splitting the sample by academic preparation *and* gender, we do not report these as the number of control units is problematically small when disaggregating the sample in this way.

common trends assumption is supported empirically. The sample is stratified by academic preparation and gender to see which, if any, subgroups are driving completion effects. Kernel matching is conducted and its success is examined through rigorous statistical testing. A flexible difference-in-differences model is estimated to verify that program effects are limited to treatment years. Sensitivity to cohorts included as well as the smoothing parameter used in the matching alogirthm are reported. Additionally, I use alternative definitions of STEM, finding similar patterns in results that are attenauted in magnitude and significance-level.

Results reveal find no meaningful program effects in terms of declaring a STEM major or earning a STEM degree in the aggregate. As per declaring an initial major in STEM, less-academically prepared students are more likely to declare a non-STEM major, an effect that appears to be driven by women. Conversely, I find that more-academically prepared students declare initial majors in STEM at higher rates compared to their nonresident counterparts as a result of the scholarship, an effect that is again driven by women at UNM. These effects are similar in magnitude but opposite in sign, masking any program effect in the aggregate.

In motivating the paper, two main research questions were proposed. First, do generous, low-bar merit scholarships discourage students from choosing majors in STEM? Results suggest the answer is "no" in the aggregate, but "yes" on behalf of less-academically prepared students. Moreover, such programs may actually increase interest in STEM majors on behalf of well-academically prepared students. Second, do scholarships such as the NMLLS affect the number of STEM degrees produced? The answer is a resounding "no" according to my results.

The main conclusion we can draw from the analysis is that low-bar merit-based scholarships neither increase nor decrease the production of STEM degrees. I find little evidence that merit aid eligibility requirements result in students pursuing easier, non-STEM course of study. Although overall production of STEM degrees is not affected by such scholarships, they may alter the composition of who majors in STEM and who eventually completes a STEM degree. To my knowledge, no other studies have looked at merit aid and STEM degree production by high- and low-achieving students. We find a divergent effect of the NMLLS on major choice, in accordance with the theoretical model posed by Montmarquette *et al.* (2002): more-academically prepared students are more likely to declare a major in STEM, while less-academically prepared students are less likely to do so.

Since its inception in 1997, the NMLLS has seen significant changes. Starting in the 2014-2015 academic year, the scholarship was capped at seven semesters (plus the initial bridging semester) and initial and renewal credit requirements were increased from 12 to 15 credits earned per semester. A statewide budget crisis in 2017 resulted in the legislature making major cuts to the NMLLS—whereas the scholarship paid 100 percent of tuition over our study period, the program only covers approximately 60 percent of tuition as of the 2017-2018 academic year. The 2017 Regular Session saw the passage of SB 420, which allows students to take a "gap" year after high school and still remain eligible for the NMLLS. It is not clear how recent program changes will affect student choice of major at UNM.

REFERENCES

Arcidiacono, P. (2004). Ability sorting and the returns to college major. Journal of Econometrics, 121(1-2), 343-375.

Arcidiacono, P., Hotz, V. J., & Kang, S. (2012). Modeling college major choices using elicited measures of expectations and counterfactuals. Journal of Econometrics, 166(1), 3-16.

Autor, D. H. (2003). Outsourcing at will: the contribution of unjust dismissal doctrine to the growth of employment outsourcing. Journal of Labor Economics, 21, 1-42.

Beffy, M., Fougere, D., & Maurel, A. (2012). Choosing the field of study in postsecondary education: Do expected earnings matter?. Review of Economics and Statistics, 94(1), 334-347.

Berger, M. C. (1988). Predicted Future Earnings and Choice of College Major. Industrial and Labor Relations Review, 41(3), 418.

Binder, M., & Ganderton. P. T. (2002). Incentive effects of New Mexico's merit-based state scholarship program: Who responds and how? In D. E. Heller & P. Marin. (Eds), Who should we help? The negative social consequences of merit scholarships (pp. 41-56). Cambridge, MA: The Civil Rights Project, Harvard University.

________. (2004). The New Mexico lottery scholarship: Does it help minority and low-income students? In D. E. Heller & P. Marin. (Eds), State merit scholarships and racial inequality (pp. 101-122). Cambridge, MA: The Civil Rights Project, Harvard University.

Bruce, D. J., and Carruthers, C. K. (2014). Jackpot? The impact of lottery scholarships on enrollment in tennessee. Journal of Urban Economics, 81, 30-44.

Cattaneo, M. D., Jansson, M., and X. Ma. (2017). Simple local polynomial density estimators. Working paper, University of Michigan.

Cornwell, C. M., Lee, K. H., & David, B. (2006). Mustard. The Effects of State-Sponsored Merit Scholarships on Course Selection and Major Choice in College. Institute for the Study of Labor (IZA) Discussion Paper Series, (IZA DP No.1953).

Cornwell, C., Mustard, D. B., & Sridhar, D. J. (2006). The enrollment effects of merit-based financial aid: Evidence from Georgia's HOPE program. Journal of Labor Economics, 24(4), 761-786.

Dillon, E. W., & Smith, J. A. (2013). The determinants of mismatch between students and colleges (No. w19286). National Bureau of Economic Research.

Dynarski, S. (2000). Hope for whom? Financial aid for the middle class and its impact on college attendance. National Tax Journal, 53(3), pt. 2: 629-61.

Heller, D. E. (1999). The effects of tuition and state financial aid on public college enrollment. The Review of Higher Education, 23(1), 65-89.

Henry, G. T., & Rubenstein, R. (2002). Paying for grades: Impact of merit-based financial aid on educational quality. Journal of Policy Analysis and Management, 21(1), 93-109.

Imbens, G.W., and D. B. Rubin. (2015). Causal inference for statistics, social, and biomedical sciences: An introduction. New York, NY: Cambridge University Press, 2015.

Imbens, G. W. (2015). Matching methods in practice: Three examples. Journal of Human Resources, 50, 373-419.

Jia, Ning. (2017). Heterogeneous effects of state merit scholarship programs: Do program features matter? Working Paper. China Center for Human Capital and Labor Market Research, Central University of Finance and Economics.

Kane, T. J. (2003). A quasi-experimental estimate of the impact of financial aid on college-going (No. w9703). National Bureau of Economic Research.

Linsenmeier, D. M., Rosen, H. S., & Rouse, C. E. (2006). Financial aid packages and college enrollment decisions: An econometric case study. The Review of Economics and Statistics, 88(1), 126-145.

McPherson, M. S., & Schapiro, M. O. (1991). Does student aid affect college enrollment? New evidence on a persistent controversy. The American Economic Review, 81(1), 309-318.

Monks, J. (2009). The impact of merit-based financial aid on college enrollment: A field experiment. Economics of Education Review, 28(1), 99-106.

Montmarquette, C., Cannings, K., & Mahseredjian, S. (2002). How do young people choose college majors? Economics of Education Review, 21(6), 543-556.

Rumberger, R. W., & Thomas, S. L. (1993). The economic returns to college major, quality and performance: A multilevel analysis of recent graduates. Economics of Education Review, 12(1), 1-19.

Scott-Clayton, J. (2011). On money and motivation a quasi-experimental analysis of financial incentives for college achievement. Journal of Human Resources, 46, 614-646.

Seftor, N. S., & Turner, S. E. (2002). Back to school: Federal student aid policy and adult college enrollment. Journal of Human Resources, 336-352.

Sjoquist, D. L., and Winters, J. V. (2012). Building the stock of college-educated labor revisited. Journal of Human Resources, 47, 270-285.

________. (2015a). State merit aid programs and college major: A focus on STEM. Journal of Labor Economics, 33(4), 973-1006.

________. (2015b). State Merit-Based Financial Aid Programs and College Attainment. Journal of Regional Science, 55, 364-390.

Stater, M. (2011). Financial aid, student background, and the choice of first-year college major. Eastern Economic Journal, 37(3), 321-343.

Stuart, E. A, Huskamp, H. A., Duckworth, K., Simmons, J., Song, Z., Chernew, M. E., and C.L. Barry. (2014), Using propensity scores in difference-in-differences models to estimate the effects of a policy change. Health Services and Outcomes Research Methodology, 14, 166-182.

Van der Klaauw, W. (2002). Estimating the effect of financial aid offers on college enrollment: A regression–discontinuity approach. International Economic Review, 43(4), 1249-1287.

Wang, X. (2013). Why students choose STEM majors: Motivation, high school learning, and postsecondary context of support. American Educational Research Journal, 50(5), 1081-1121.

Welch, J. G. (2014). HOPE for community college students: The impact of merit aid on persistence, graduation, and earnings. Economics of Education Review, 43, 1-20.

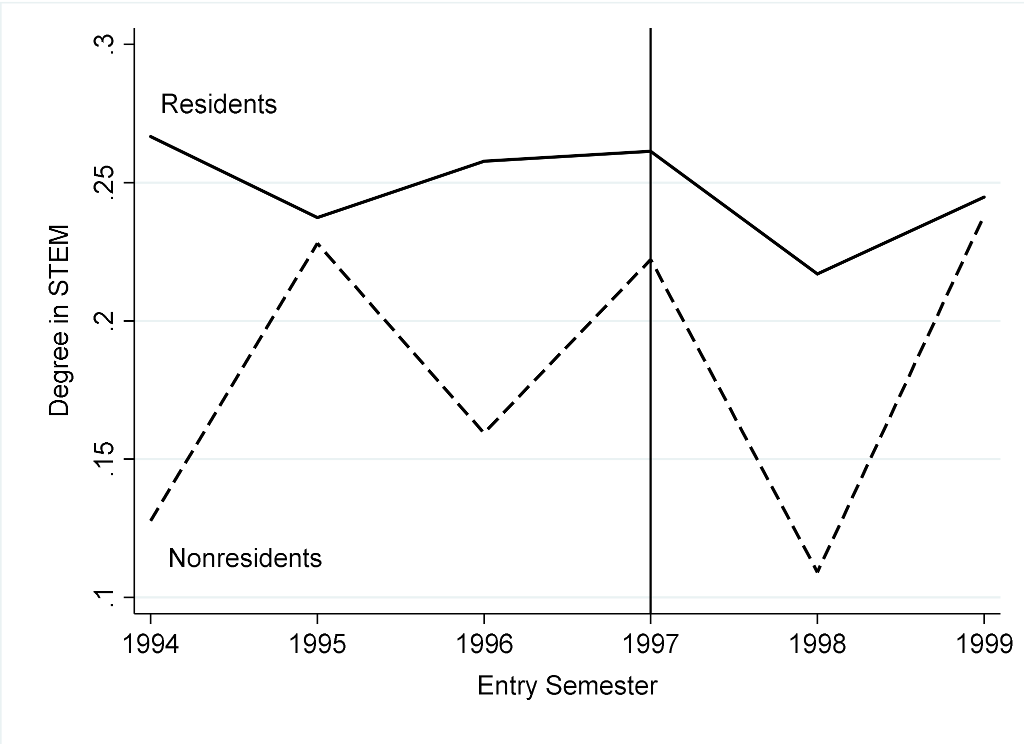
Wiswall, M., & Zafar, B. (2014). Determinants of college major choice: Identification using an information experiment. The Review of Economic Studies, 82(2), 791-824.

Xue, Y., & Larson, R. (2015). "STEM crisis or STEM surplus? Yes and yes," Monthly Labor Review, U.S. Bureau of Labor Statistics, https://doi.org/10.21916/mlr.2015.14.

Zhang, L. (2011). Does merit-based aid affect degree production in STEM fields? Evidence from Georgia and Florida. The Journal of Higher Education, 82(4), 389-415.

Note: The plot above show the likelihood of declaring the first major in STEM for incoming cohorts over the period 1994 to 1999. Solid lines represent resident students while dashed lines represent nonresident students. The vertical bars at 1997 mark the implementation of the New Mexico Legislative Lottery Scholarship.

Figure 1. Pre-Post Trends in the Likelihood of Declaring First Major in STEM, by Residency



Note: The plot above show the likelihood of declaring the first major in STEM for incoming cohorts over the period 1994 to 1999. Solid lines represent resident students while dashed lines represent nonresident students. The vertical bars at 1997 mark the implementation of the New Mexico Legislative Lottery Scholarship.

Figure 2. Pre-Post Trends in the Likelihood of Earning a Degree in STEM, by Residency

Table 1. Full-time resident tuition at all NMLLS-eligible institutions

Institution	Program Length (years)	Tuition and Fees
New Mexico Institute of Mining and Technology	4	7,000
University of New Mexico	4	6,950
New Mexico State University	4	6,729
Western New Mexico University	4	6,644
Eastern New Mexico University	4	5,630
New Mexico Highlands University	4	5,550
New Mexico Military Institute	2	5,179
Northern New Mexico College	4	5,112
Mesalands Community College	2	1,990
San Juan College	2	1,773
Central New Mexico Community College	2	1,340
Clovis Community College	2	1,324
Santa Fe Community College	2	1,196
New Mexico Junior College	2	1,158
Luna Community College	2	968
Southwestern Indian Polytechnic Institute	2	730

Source: Institution financial aid department websites. Accessed 28 March 2017. Figures present tuition and fees for one academic year taking fifteen credit hours per semester. For two-year schools it is assumed the student is within the community college district, where applicable.

Table 2. Student characteristics before and after initiation of the NMLLS program, First Major Declared, ACS Major Codes

	Residents			Nonresidents		
Variable	Before	After	Diff.	Before	After	Diff.
First Major Declared:						
STEM	.236	.221	-.015*	.179	.151	-.026
Liberal Arts	.158	.184	.026***	.206	.237	.031
Education	.074	.101	.027***	.061	.088	.027*
Business	.075	.094	.019***	.065	.071	.006
Social Science	.110	.101	-.009	.112	.122	.010
Health-Related	.133	.114	-.019***	.121	.083	-.038**
Never Declared	.214	.183	-.031***	.256	.248	-.008
HSGPA	3.312	3.273	-.038***	3.233	3.300	.067**
	(.502)	(.471)		(.532)	(.503)	
ACT	22.530	22.176	-.354***	22.317	22.861	.544**
	(3.834)	(3.887)		(4.109)	(4.096)	
Remedial	.264	.290	.026***	.164	.227	.063***
Income < $40K	.230	.205	-.025***	.155	.162	.007
Female	.571	.565	-.006	.526	.545	.019
Hispanic	.386	.375	-.011	.147	.166	.020
Native	.043	.045	.002	.041	.051	.010
Asian	.047	.037	-.010**	.034	.026	-.008
Black	.021	.022	.002	.082	.080	-.002
Observations	3,715	6,307		587	649	

Source: Freshmen Tracking System, Office of Institutional Analytics, UNM. ***, **, and * represent statistical significance at the 1, 5, and 10 percent-levels, respectively. Standard deviations are in parentheses.

Table 3. Student characteristics before and after initiation of the NMLLS program, Degree Type Earned, ACS Major Codes

	Residents			Nonresidents		
Variable	Before	After	Diff.	Before	After	Diff.
Degree Type Earned:						
STEM	.253	.240	-.013	.173	.195	.022
Liberal Arts	.262	.249	-.013	.341	.326	-.015
Education	.116	.081	-.035***	.121	.026	-.095***
Business	.157	.190	.033***	.185	.163	-.022
Social Science	.167	.177	.010	.145	.237	.092**
Health-Related	.044	.064	.020***	.035	.053	.018
HSGPA	3.479	3.455	-.024*	3.483	3.473	-.010
	(.467)	(.439)		(.460)	(.442)	
ACT	23.268	23.085	-.183	23.526	23.807	.281
	(3.761)	(3.784)		(3.865)	(3.888)	
Remedial	.192	.196	.004	.138	.150	.012
Income < $40K	.194	.173	-.021*	.128	.140	.012
Female	.616	.616	.000	.622	.609	-.013
Hispanic	.366	.358	-.008	.097	.159	.062*
Native	.022	.023	.001	.010	.039	.029*
Asian	.052	.041	-.011*	.041	.019	-.021
Black	.016	.018	.002	.102	.058	-.044
Observations	1,547	2,543		173	190	

Source: Freshmen Tracking System, Office of Institutional Analytics, UNM. ***, **, and * represent statistical significance at the 1, 5, and 10 percent-levels, respectively. Standard deviations are in parentheses.

Table 4: NMLLS Student Attrition, 1994-1999

Semester	Residents Eligible	Percent Remaining
2	2,664	100.0%
3	2,249	84.4%
4	2,017	75.7%
5	1,863	69.9%
6	1,734	65.1%
7	1,629	61.1%
8	1,568	58.9%
9	1,510	56.7%

Source: Office of Institutional Analytics, University of New Mexico. We consider the sample of resident students that met cumulative GPA and credit requirements in their first semester to qualify for the NMLLS.

Table 5: Estimated parameters for propensity score model of NMLLS data, 1994-1999

Variable	Estimate	Std. Error
HSGPA	1.729**	.724
ACT	.498***	.090
Remedial	.891***	.118
Income < 20K	.268*	.158
Income < 40K	.160	.108
Female	1.670***	.367
Hispanic	1.865***	.550
Native American	1.884**	.923
Asian	.032	.207
Black	-5.729***	1.155
Declined to state race-ethnicity	-.108	.282
ACT^2	-.013***	.002
ACT*Black	.141***	.045
Female*White	-.571***	.146
$HSGPA^2$	-.461***	.116
ACT*Female	-.053***	.016
ACT*HSGPA	.059***	.020
Remedial*Asian	1.147**	.505
GPA*Black	.546	.339
ACT*Native	-.082**	.041
Female*Native	-.608*	.317
HSGPA*Hispanic	-.312*	.165
Constant	-7.711***	1.600
Observations		11,258

Standard errors are in parentheses. ***, **, and * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively. Propensity scores are estimated using a logistic model. Forty-nine observations were dropped following estimation of the propensity score to ensure overlap, leaving 11,209 observations. The variable *Declined to state race-ethnicity* is equal to one if the student declined to state their race-ethnicity, and zero otherwise.

Table 6: Means and normalized differences after kernel matching, full sample, 1994-1999

	Pre-NMLLS			Post-NMLLS		
Variable	Res.	Nonres.	ND	Res.	Nonres.	ND
HS GPA	3.31	3.27	.088	3.27	3.33	-.122
Composite ACT	22.56	22.37	.047	22.19	22.58	-.099
Remedial	0.26	0.24	.032	0.29	0.28	.012
Income < $40,000	0.22	0.21	.040	0.20	0.21	-.032
Female	0.57	0.58	-.009	0.56	0.59	-.063
Hispanic	0.39	0.39	-.019	0.37	0.36	.025
Native	0.04	0.04	.001	0.05	0.05	-.030
Asian	0.04	0.03	.058	0.04	0.03	.019
Black	0.02	0.02	-.023	0.02	0.02	.018

Means are from Epanechnikov kernel matching performed with a bandwidth of h = .2. Normalized differences (ND) are calculated by taking the difference average covariate values by residency status and dividing by a measure of standard deviation.

Table 7: NMLLS and major choice by academic preparation, American Community Survey definition, 1994-1999

Group	Obs.	First Declared STEM	Obs.	Majored in STEM
Full Sample	11,209	.026	4,438	-.012
		(.030)		(.057)
$\bar{Y}$		.221		.240
HSGPA ≤ 3.28	5,473	-.068*	1,507	.147
		(.040)		(.093)
$\bar{Y}$		.170		.145
HSGPA > 3.28	5,734	.121***	2,930	-.051
		(.046)		(.073)
$\bar{Y}$		.273		.291
HSGPA > 3.78	2,105	-.063	1,271	-.061
		(.067)		(.119)
$\bar{Y}$		.334		.386

Robust standard errors are reported in parentheses. *, **, and *** denote statistical significance at the 10, 5, and 1 percent-level, respectively. Estimates are from difference-in-differences kernel matching performed with a bandwidth of h = .2 using the Epanechnikov kernel function. We report estimates for students with below average or average high school GPAs (≤ 3.28), above average high school GPAs (> 3.28), and high school GPAs greater than one standard deviation above the mean (> 3.78). $\bar{Y}$ denotes the baseline rate of STEM major choice by academic preparation.

APPENDIX A. FLEXIBLE DIFFERENCE-IN-DIFFERENCES RESULTS

Table A1: Common trends assumption test, American Community Survey definition, 1994-1999

Leads/Lags	First Declared	Degree Earned
NMLLS $_{t-3}$	.026	.064
	(.037)	(.075)
NMLLS $_{t-2}$	-.052	-.018
	(.040)	(.081)
NMLLS $_{t-1}$	-.058	.050
	(.042)	(.073)
NMLLS $_{t0}$	.026	-.003
	(.037)	(.074)
NMLLS $_{t+1}$	.009	.070
	(.036)	(.070)
R^2	.0609	.1249
Prob > F	.464	.666
Observations	11,258	4,453

Robust standard errors are reported in parentheses. Ordinary least squares estimates for all students entering UNM between 1994 – 1999 given. Reported coefficients are on interactions between cohort years and a resident dummy variable. Models include resident and cohort dummies as well as controls for race, ethnicity, standardized test scores, high school GPA, gender, and family income. The period t_0 is 1997, the year the NMLLS was implemented. *, **, and *** denote statistical significance at the 10, 5, and 1 percent-level, respectively. NMLLS$_{t+2}$ (1999) serves as the base year. Prob > F displays the p-value of the null hypothesis that estimated coefficients on leading periods are jointly different from zero.

APPENDIX B. ASSESSING PROPERTIES OF THE PROPENSITY SCORE

In order to examine the effectiveness of our matching procedure, we first assess overlap in the propensity score both before and after the NMLLS is in place. Note that tests conducted in this section use only information concerning covariates and residency classification, and do not consider completion rates, therefore cannot intentionally introduce bias in subsequent analyses. For a thorough treatment of these tests, see Imbens and Rubin (2015).

Figure B1 presents histogram estimates of the distribution of linearized propensity scores before and after the implementation of the NMLLS, by residency. First inspection reveals substantial overlap in the linearized propensity score across residents and nonresidents, both before and after the NMLLS was launched. As a more formal check, we calculate the percent of observations where there exists an observation of the opposite treatment status with a difference in linearized propensity score less than 10 percent. These measures are presented in Table B1. For residents, approximately 99 percent of students had at least one closely matching nonresident student in terms of linearized propensity score both before and after the launch of the NMLLS. For nonresidents, this percentage was approximately 97 percent. This suggests we should be able to credibly estimate causal effects of the NMLLS on student graduation under the assumption of unconfoundedness.

We next perform two tests assessing the balancing property of the propensity score, which asserts that conditional on the propensity score, treatment assignment and student characteristics are independent of one another. We perform these tests both before and after the NMLLS is launched. If results of these tests are favorable, this constitutes evidence supporting the assumption of unconfoundedness, although it cannot be directly tested. The balancing property can be formally represented as:

$$W_i \perp\!\!\!\perp X_i \mid l(X_i) \qquad (B1)$$

where W_i is a binary treatment indicator equal to one if student i is a New Mexico resident, and zero otherwise, X_i is a vector of covariates, and $l(X_i)$ is the true linearized propensity score. Because we do not know the true linearized propensity score, we approximate this test by instead using its estimated counterpart, $\hat{l}(X_i)$. Our strategy is to stratify the sample into J blocks, $B_i(1), \dots, B_i(J)$, so there will be no significant difference between linearized propensity scores within each block. This way, (B1) becomes

$$W_i \perp\!\!\!\perp X_i \mid B_i(1), \dots, B_i(J). \qquad (B2)$$

Equation (B2) can be examined by testing whether residency classification and covariates are uncorrelated within each of the J blocks, so that

$$E[X_i|W_i = 1, B_i(j) = 1] = E[X_i|W_i = 0, B_i(j) = 1] \qquad (B3)$$

for all blocks, $j = 1, \dots, J$. Tables B2 and B3 present the results of this stratification procedure. For the pre-NMLLS period, we split the sample into 11 blocks using the linearized propensity score. Near the upper end of the propensity score distribution, we were not able to further split blocks 10 and 11 due to a small number of nonresident students relative to the number of resident students.[16] We also encountered this issue when stratifying the sample in the post-NMLLS period, although at the opposite end of the propensity score distribution. We nonetheless consider the stratification successful, as only two of the 25 blocks created were left with propensity scores that were significantly different across resident and nonresident groups at the five percent level.

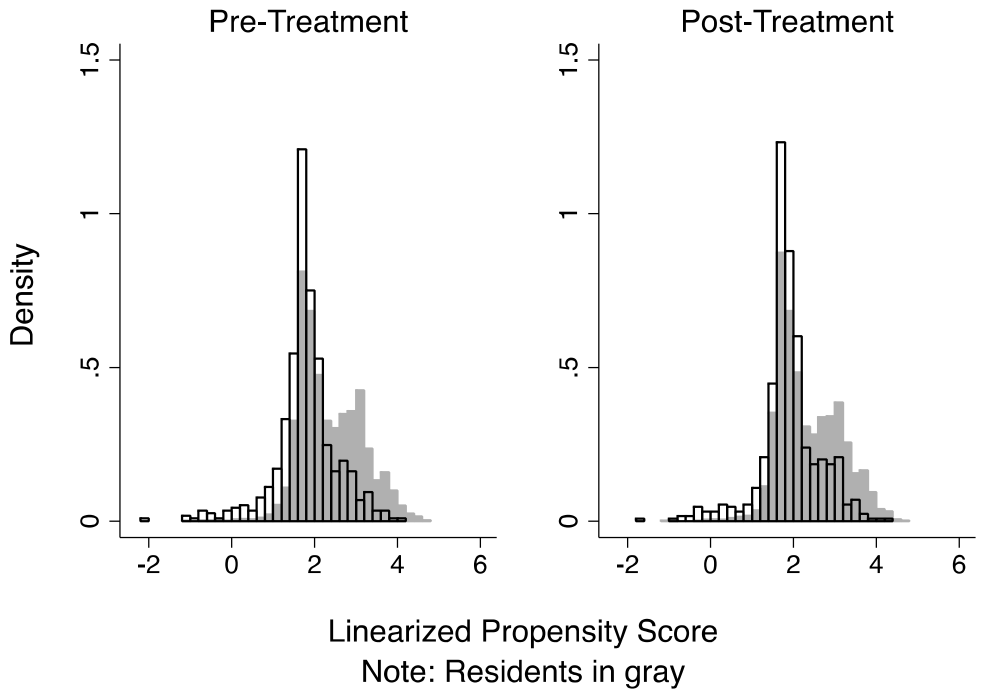
[16] In order to perform subsequent hypothesis testing, we are not able to further split blocks where new blocks would have fewer members than $K + 2$, where K is the number of covariates.

With pre- and post-NMLLS samples stratified, we then assess covariate balance within blocks. These tests can be thought of "pseudo treatment effects" as they examine the effect of treatment on pre-treatment covariates, where the effects are a priori known to be zero. Confirmation that pseudo treatment effects are zero constitutes evidence that equation (B3) holds, supporting the assumption of unconfoundedness. We conduct two different tests. First, we test separately, by each covariate, whether within-block differences between residents and nonresidents are equal to zero. Second, for each covariate we test whether the weighted average of within-block differences between residents and nonresidents are equal to zero. Results of these tests for pre- and post-NMLLS cohorts are reported in Tables B4 and B5, respectively.

We analyze the results of these tests as if data arose from a stratified random experiment. The first approach for assessing covariate balance focuses on one covariate-block dyad at a time. We calculate z-statistics testing the null hypothesis that the difference between residents and nonresidents in the dyad is equal to zero. These tests produce a large amount of information, however they are not very informative when examined individually. Of the 113 pre-NMLLS tests, only seven (six percent) had z-statistics above two. Similarly, of the 148 post-NMLLS tests, only twelve (eight percent) exceeded two in absolute value. It is informative to present these statistics in Q-Q plots, where z-values are compared against their expected values under independent draws from a standard normal distribution. If the distributions of z-values closely follow the 45 degree lines in these plots, it is evidence that the propensity score was effective in balancing covariates as if treatment was randomly assigned within blocks. Q-Q plots are presented in Figures B2 and B3. Both appear to follow the normal distribution reasonably well, although they are slightly skewed to the right (especially for pre-treatment cohorts). One major outlier deserves attention in Figure B2—it is due to the incomparability of black resident and nonresident students at UNM at a particular region of the propensity score distribution.[17]

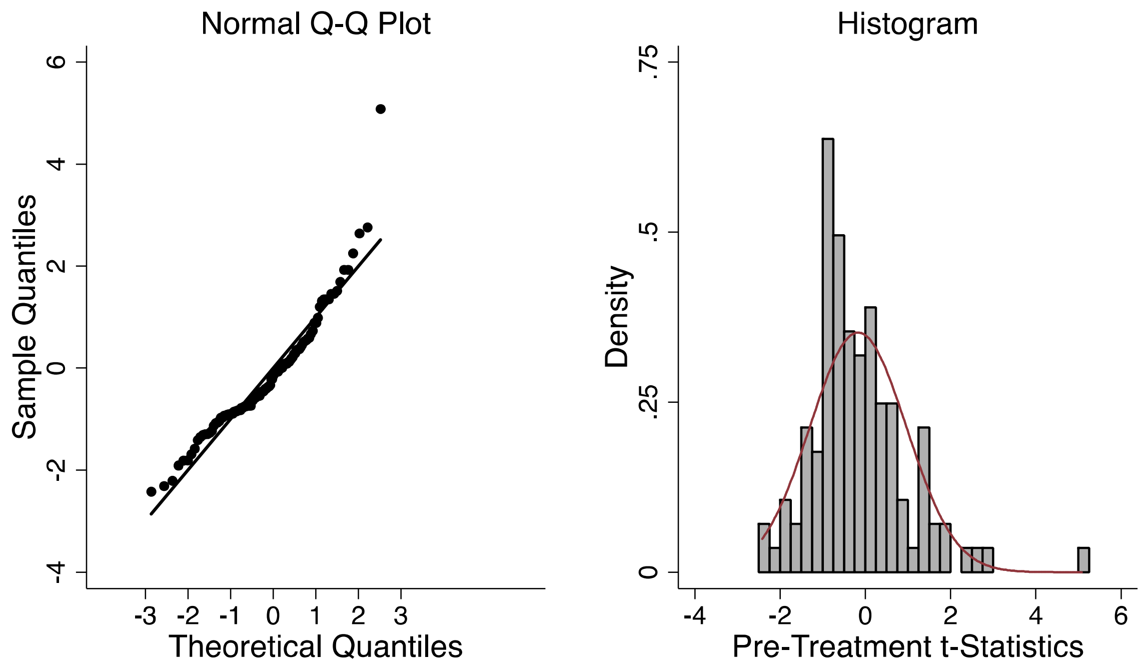
The column labeled as the overall t-statistic tests the null hypothesis that the block-adjusted weighted average of within block differences is equal to zero. Finding z-values larger in absolute value than we would expect if they were drawn independently from a standard normal distribution is evidence that the stratification does not adequately balance covariates, suggesting that the propensity score model is not satisfactory and may need to be refined. According to Table B4, there do not appear to be any significant balance issues for pre-NMLLS cohorts. For these cohorts, the largest t-statistic we find is 1.73, suggesting excellent balance. Table B5 reveals that there may exist some imbalance in the high school GPA and ACT composite score covariates for post-NMLLS cohorts. The z-statistics for these covariates are 2.17 and 2.14 in absolute value, respectively, indicating that we can reject the null hypothesis that the weighted averages of within-block differences are zero for both of these variables at the five percent level. Analyzing normalized differences between residents and nonresidents for these covariates after matching is performed provides an additional check as to whether this imbalance requires estimating a more flexible propensity score or perhaps trimming the sample. Although the propensity score model did not perform as well as would randomization into treatment within blocks, overall we feel it worked adequately in balancing covariates across resident and nonresident students.

[17]Although this outlier is visually striking, it is driven by the relatively small number of black students at UNM (less than 3 percent of the sample).



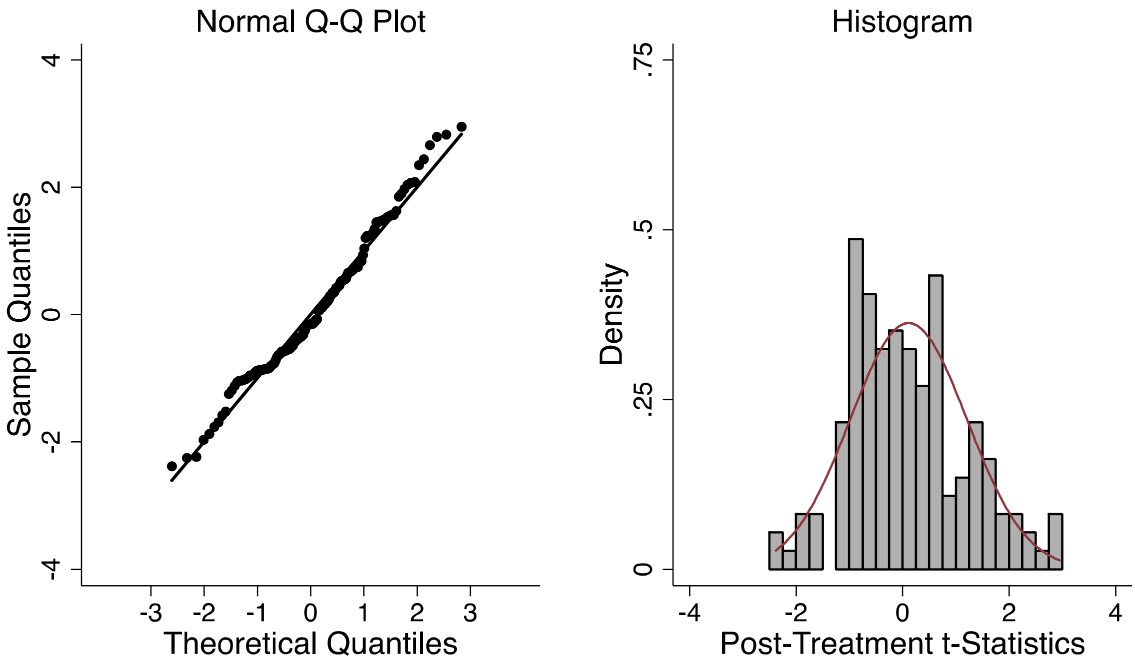
Note: The left and right panels overlap linearized propensity scores for residents and nonresidents before and after the implementation of the NMLLS, respectively, allowing for visual inspection of sufficient overlap, a critical requirement for successful propensity score matching. Both figures indicate there is sufficient overlap of residents and nonresidents.

Figure B1. Linearized Propensity Scores, by Residency and NMLLS Implementation



Note: The left panel presents a normal Q-Q plot of t-statistics from within-block tests before implementation of the NMLLS. Normal Q-Q plots graph actual percentiles against theoretical percentages from a normal distribution with the same mean and standard deviation. Normality is evidenced by a straight line of plotted values. Above right is a histogram of the same t-statistics with a fitted normal curve. Both plots provide visual evidence of slight positive skew.

Figure B2. Visual check of normality of within-block differences across resident status, pre-NMLLS



Note: The left panel presents a normal Q-Q plot of t-statistics from within-block tests after implementation of the NMLLS. Normal Q-Q plots graph actual percentiles against theoretical percentages from a normal distribution with the same mean and standard deviation. Normality is evidenced by a straight line of plotted values. Above right is a histogram of the same t-statistics with a fitted normal curve. Again, both plots provide visual evidence of slight positive skew.

Figure B3. Visual check of normality of within-block differences across resident status, post-NMLLS

Table B1. Proportion of units with match discrepancy in terms of linearized propensity score less than 10 percent

Measure	Pre-NMLLS	Post-NMLLS
$q_{nonresident}$	.971	.968
$q_{resident}$	.989	.995

Table B2. Propensity score blocks and their boundaries, pre-NMLLS

Block	Lower	Upper	Width	Nonresidents	Residents	t-Statistic
1	.238	.688	.449	37	29	-.854
2	.688	.757	.070	24	43	-.796
3	.757	.800	.043	47	86	-.598
4	.800	.830	.030	57	209	-.057
5	.830	.843	.013	47	219	.054
6	.843	.851	.009	62	205	.154
7	.851	.888	.036	167	899	.783
8	.888	.920	.032	62	471	.304
9	.920	.945	.025	38	496	.065
10	.945	.961	.015	23	510	-3.274
11	.961	.985	.024	21	513	-1.519

Above presents results of an attempt to stratify the sample on the linearized propensity score. t-statistics are for the null hypothesis of equality in linearized propensity scores between resident and nonresident students. Blocks 10 and 11 could not be split further as there would be an insufficient number of members in new groups for subsequent hypothesis testing.

Table B3. Propensity score blocks and their boundaries, post-NMLLS

Block	Lower	Upper	Width	Nonresidents	Residents	t-Statistic
1	.249	.717	.468	40	68	-3.393
2	.717	.780	.063	25	83	-1.314
3	.780	.813	.033	28	188	-1.493
4	.813	.832	.019	49	383	-.835
5	.832	.852	.020	110	758	.419
6	.852	.869	.017	109	755	-.516
7	.869	.888	.020	89	779	-.557
8	.889	.896	.007	36	180	-.292
9	.896	.904	.008	18	199	-.877
10	.904	.922	.017	26	407	-.971
11	.922	.937	.015	31	402	.328
12	.937	.946	.010	22	411	-1.339
13	.946	.962	.015	45	821	-.569
14	.962	.987	.025	20	846	-.390

Above presents results of an attempt to stratify the sample on the linearized propensity score. t-statistics are for the null hypothesis of equality in linearized propensity scores between resident and nonresident students. Block 1 could not be split further as there would be an insufficient number of members in new groups for subsequent hypothesis testing.

Table B4: Tests for balance conditional on propensity score, pre-NMLLS

Covariate	Within Blocks											Overall
	1	2	3	4	5	6	7	8	9	10	11	t-Statistic
High School GPA	0.37	-0.37	-1.81	**-2.21**	1.51	-0.15	-0.74	-1.27	-0.78	-0.03	-0.22	1.73
Composite ACT	-0.94	**-2.31**	-1.58	-1.82	-0.75	-0.60	0.09	-1.35	-0.56	**2.25**	0.09	0.85
Remedial	**-2.43**	0.22	-0.85	0.67	1.46	-0.40	0.01	0.45	-0.97	-0.92	-0.83	0.83
Income < 20K	0.27	-0.76	-0.07	0.99	1.20	0.59	-0.83	-1.04	-0.36	0.40	0.35	0.59
Income < 40K	0.19	-0.85	-0.85	-0.90	-0.45	1.46	-0.47	0.88	-0.97	-1.08	0.29	1.17
Female	-0.83	-1.07	0.05	-0.42	0.58	-0.90	0.50	-1.91	1.69	-0.74	0.16	0.27
Hispanic	-1.13	1.35	-0.74	1.92	-	-0.78	-0.08	0.73	-0.58	0.12	1.31	-0.76
Native	-	-0.74	-1.69	-0.54	**2.76**	-0.07	-0.39	-0.33	0.01	0.53	-	-0.10
Asian	0.88	1.35	-0.07	1.93	-0.93	0.08	-0.02	-1.23	**2.64**	-0.64	-1.31	-0.03
Black	-1.41	-0.63	-1.29	-0.90	0.36	-0.55	0.54	**5.08**	-0.55	-	-	-1.08
Declined	-	1.35	-0.74	-0.91	-0.46	-	-1.30	0.11	-0.48	-0.21	-	1.22

z-statistics test the null hypothesis of equality of means within blocks for resident and nonresident students. Overall t-statistics test the null hypothesis that the weighted average of differences across blocks is equal to zero. *Declined* is equal to zero if the student declined to state their race-ethnicity, and zero otherwise.

Table B5: Tests for balance conditional on propensity score, post-NMLLS

Covariate	Within Blocks														Overall
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	t-Statistic
High School GPA	-1.88	**-2.38**	-1.76	1.49	1.56	**2.07**	1.20	-0.19	0.67	0.94	1.50	0.82	0.51	0.06	**-2.17**
Composite ACT	**-2.24**	-0.63	0.42	**2.35**	1.90	-0.88	**2.44**	0.34	1.85	0.53	-0.56	0.57	-0.15	0.54	**-2.14**
Remedial	**2.79**	1.45	-0.43	0.25	0.11	1.23	-1.97	0.74	**-2.25**	-0.99	1.63	-1.03	1.46	-0.80	0.28
Income < 20K	-0.11	-0.38	-0.87	1.45	0.13	1.24	-0.54	0.17	0.66	-0.11	-1.06	0.22	-0.61	0.66	-0.33
Income < 40K	1.28	-0.16	-1.25	0.35	-0.85	0.62	-0.14	-0.91	**2.95**	-0.87	-0.87	-0.15	-1.69	**2.09**	-0.69
Female	-0.96	-0.82	-0.08	1.34	0.06	1.97	-0.07	-0.27	-1.02	0.75	0.84	1.25	-0.58	0.10	-1.14
Hispanic	-0.77	-0.55	-0.67	-0.36	0.74	-0.85	0.45	-1.52	-0.44	0.53	-1.04	-0.64	-1.19	0.38	0.86
Native	-	-0.17	-0.49	-0.50	-0.35	0.69	-1.12	**2.04**	-0.16	0.16	0.83	0.46	0.19	-0.15	-0.45
Asian	-	-0.55	1.57	1.04	-0.85	-1.58	-0.33	-0.25	0.20	-0.32	0.74	1.53	1.55	-0.36	-0.06
Black	1.47	0.55	-0.58	-0.95	0.70	0.30	0.35	-0.78	-0.52	-0.72	-0.39	-0.57	-	-	0.38
Declined	-	**2.66**	-0.96	-0.88	0.29	0.43	1.24	-1.01	0.10	**2.83**	-0.84	-0.57	-0.41	-	-1.23

z-statistics test the null hypothesis of equality of means within blocks for resident and nonresident students. Overall t-statistics test the null hypothesis that the weighted average of differences across blocks is equal to zero. Declined is equal to zero if the student declined to state their race-ethnicity, and zero otherwise.

APPENDIX C. CATEGORIZING OF STEM MAJORS

Table C1: Majors classified as STEM according to the American Community Survey

ACS Code	ACS Code Description
2402	Biological engineering
2403	Architectural engineering
2404	Biomedical engineering
2405	Chemical engineering
2406	Civil engineering
2407	Computer engineering
2408	Electrical engineering
2409	Engineering mechanics, physics, and science
2410	Environmental engineering
2411	Geological and geophysical engineering
2412	Industrial and manufacturing engineering
2413	Materials engineering and materials science
2414	Mechanical engineering
2415	Metallurgical engineering
2416	Mining and mineral engineering
2417	Naval architecture and marine engineering
2418	Nuclear engineering
2419	Petroleum engineering
2499	Miscellaneous engineering
2500	Engineering technologies
2501	Engineering and industrial management
2502	Electrical engineering technology
2503	Industrial production technologies
2504	Mechanical engineering related technologies
2599	Miscellaneous engineering technologies
3600	Biology
3601	Biochemical sciences
3602	Botany
3603	Molecular biology
3604	Ecology
3605	Genetics
3606	Microbiology
3607	Pharmacology
3608	Physiology
3609	Zoology
3611	Neuroscience
3699	Miscellaneous biology

Table C1 (Continued)

ACS Code	ACS Code Description
3700	Mathematics
3701	Applied mathematics
3702	Statistics and decision science
3801	Military technologies
4002	Nutrition sciences
4003	Neuroscience
4005	Mathematics and computer science
4006	Cognitive science and biopsychology
5000	Physical sciences
5001	Astronomy and astrophysics
5002	Atmospheric sciences and meteorology
5003	Chemistry
5004	Geology and earth science
5005	Geosciences
5006	Oceanography
5007	Physics
5008	Materials science
5098	Multi-disciplinary or general science
5102	Nuclear, industrial radiology, and biological technologies
5901	Transportation sciences and technologies
6106	Health and medical preparatory programs
6108	Pharmacy, pharmaceutical sciences, and administration
6202	Actuarial science
6212	Miscellaneous information systems and statistics

The code list from the American Community Survey was referenced 22 Jan 2018 and can be found online at https://www2.census.gov/programssurveys/acs/tech_docs/code_lists/2016_ACS_Code_Lists.pdf. See Sjoquist and Winters (2015a) for a more exhaustive list that categorizes majors into other categories including liberal arts, health-related, social sciences, education, and business.

Table C2: Majors classified as STEM according to the STEM Collaborative Center at the University of New Mexico, Broad Definition

Major Code(s)	Major Description
5, ANTH	Anthropology
6, ARCH	Architecture
249, BIOC	Biochemistry
12, BIOL	Biology
15, CHE	Chemical Engineering
16, CHEM	Chemistry
17, CE	Civil Engineering
171, CPE	Computer Engineering
109, 168, ACS, CS	Computer Science
262, CONE	Construction Engineering
263, 474, CMGT, CONM	Construction Management
22, 23, DEHY, DHYG	Dental Hygiene
340, EPS	Earth and Planetary Sciences
27, ECON	Economics
173, EE	Electrical Engineering
379, EMS	Emergency Medical Services
438, ENSC	Environmental Science
371, GENG	General Engineering
46, GEOG	Geography
481, HMHV	Health, Medicine and Human Values
INGV	Integrative Studies
110	Management Information Systems
64, MATH	Mathematics
65, ME	Mechanical Engineering
353, MEDL	Medical Laboratory Sciences
76, NE	Nuclear Engineering
77, 456, NUR, NURS	Nursing
24, NDIT	Nutrition/Dietetics
81, PHYC	Physics
405, PAP	Physics and Astrophysics
FANT	Pre Anthropology
FBIC	Pre Biochemistry
FBIO	Pre Biology
FCH	Pre Chemical Engineering
FCHM	Pre Chemistry
FCE	Pre Civil Engineering
FCP	Pre Computer Engineering
FCS	Pre Computer Science
FEPS	Pre Earth and Planetary Sciences
FECO	Pre Economics

Table C2 (Continued)

Major Code(s)	Major Description
FEE	Pre Electrical Engineering
FESC	Pre Environmental Science

The code list was provided by the Office of Institutional Analytics at the University of New Mexico. STEM-designated majors are according to the University of New Mexico STEM Collaborative Center and can be found online at https://stem.unm.edu/tools-for-faculty-and-staff/5517-broad-data.pdf (accessed 24 Jan 2018). This is considered the "broad" list of STEM majors at the University of New Mexico.

Table C3: Majors classified as STEM according to the STEM Collaborative Center at the University of New Mexico, Narrow Definition

Major Code(s)	Major Description
249, BIOC	Biochemistry
12, BIOL	Biology
15, CHE	Chemical Engineering
16, CHEM	Chemistry
17, CE	Civil Engineering
171, CPE	Computer Engineering
109, 168, ACS, CS	Computer Science
262, CONE	Construction Engineering
263, 474, CMGT, CONM	Construction Management
340, EPS	Earth and Planetary Sciences
173, EE	Electrical Engineering
438, ENSC	Environmental Science
371, GENG	General Engineering
64, MATH	Mathematics
65, ME	Mechanical Engineering
76, NE	Nuclear Engineering
81, PHYC	Physics
405, PAP	Physics and Astrophysics
FANT	Pre Anthropology
FBIC	Pre Biochemistry
FBIO	Pre Biology
FCH	Pre Chemical Engineering
FCHM	Pre Chemistry
FCE	Pre Civil Engineering
FCP	Pre Computer Engineering
FCS	Pre Computer Science
FCOE	Pre Construction Engineering
FCON	Pre Construction Management
FEPS	Pre Earth and Planetary Science
FEE	Pre Electrical Engineering
FESC	Pre Environmental Science
FMAT	Pre Mathematics
FME	Pre Mechanical Engineering
FNE	Pre Nuclear Engineering
FPHY	Pre Physics
FSTA	Pre Statistics
STAT	Statistics

The code list was provided by the University of New Mexico. STEM-designated majors are according to the University of New Mexico STEM Collaborative Center and can be found online at https://stem.unm.edu/common/pdfs/17-benchmarking-narrow.pdf (accessed 5 Feb 2018). This is considered the "narrow" list of STEM majors at the University of New Mexico.

APPENDIX D. ALTERNATIVE STEM DEFINITIONS

Table D1: NMLLS and major choice by academic preparation, UNM narrow STEM definition, 1994-1999

Group	Obs.	First Declared STEM	Obs.	Majored in STEM
Full Sample	11,209	.022	4,692	.009
		(.029)		(.052)
$\bar{Y}$		.197		.194
HSGPA ≤ 3.28	5,473	-.022	1,574	.121
		(.031)		(.086)
$\bar{Y}$		.153		.110
HSGPA > 3.28	5,734	.084*	3,117	-.011
		(.045)		(.068)
$\bar{Y}$		.312		.238
HSGPA > 3.78	2,105	-.099	1,357	-.094
		(.063)		(.099)
$\bar{Y}$		.299		.332

Robust standard errors are reported in parentheses. *, **, and *** denote statistical significance at the 10, 5, and 1 percent-level, respectively. Estimates are from difference-in-differences kernel matching performed with a bandwidth of $h = .2$ using the Epanechnikov kernel function. We report estimates for students with below average or average high school GPAs (≤ 3.28), above average high school GPAs (> 3.28), and high school GPAs greater than one standard deviation above the mean (> 3.78). $\bar{Y}$ denotes the baseline rate of STEM major choice by academic preparation.

Table D2: NMLLS and major choice by academic preparation, UNM broad STEM definition, 1994-1999

Group	Obs.	First Declared STEM	Obs.	Majored in STEM
Full Sample	11,209	.025 (.033)	4,692	-.033 (.057)
$\bar{Y}$		.264		.276
HSGPA ≤ 3.28	5,473	-.009 (.041)	1,574	.065 (.096)
$\bar{Y}$		.217		.187
HSGPA > 3.28	5,734	.079* (.047)	3,117	-.054 (.072)
$\bar{Y}$		.312		.323
HSGPA > 3.78	2,105	-.126* (.067)	1,357	-.112 (.103)
$\bar{Y}$		.364		.400

Robust standard errors are reported in parentheses. *, **, and *** denote statistical significance at the 10, 5, and 1 percent-level, respectively. Estimates are from difference-in-differences kernel matching performed with a bandwidth of h = .2 using the Epanechnikov kernel function. We report estimates for students with below average or average high school GPAs (≤ 3.28), above average high school GPAs (> 3.28), and high school GPAs greater than one standard deviation above the mean (> 3.78). $\bar{Y}$ denotes the baseline rate of STEM major choice by academic preparation.

APPENDIX E. ALTERNATIVE BANDWIDTHS

Table E1. NMLLS kernel matching estimates with bandwidths of 0.1, 0.2, and 0.3; 1994-1999

	STEM First Declared	
Group	Obs.	Estimate
Full Sample		
$h = 0.1$	11,207	.025
$h = 0.2$	11,209	.026
$h = 0.3$	11,210	.027
GPA ≤ 3.28		
$h = 0.1$	5,470	-.063*
$h = 0.2$	5,473	-.068*
$h = 0.3$	5,474	-.057
GPA > 3.28		
$h = 0.1$	5,732	.115**
$h = 0.2$	5,734	.121***
$h = 0.3$	5,735	.119***
GPA > 3.78		
$h = 0.1$	2,103	-.088
$h = 0.2$	2,105	-.063
$h = 0.3$	2,105	-.049

Robust standard errors are reported in parentheses. *, **, and *** denote statistical significance at the 10, 5, and 1 percent-level, respectively. Estimates are from difference-in-differences kernel matching using an Epanechnikov kernel function with various bandwidth parameters, *h*. We report estimates for students with below average or average high school GPAs (≤ 3.28), above average high school GPAs (> 3.28), and high school GPAs greater than one standard deviation above the mean (> 3.78).

Table E2. NMLLS kernel matching estimates with bandwidths of 0.1, 0.2, and 0.3; 1994-1999

	STEM Degree Earned	
Group	Obs.	Estimate
Full Sample		
h = 0.1	4,437	-.024
h = 0.2	4,438	-.012
h = 0.3	4,439	-.016
GPA ≤ 3.28		
h = 0.1	1,506	.156*
h = 0.2	1,507	.147
h = 0.3	1,508	.134
GPA > 3.28		
h = 0.1	2,929	-.045
h = 0.2	2,930	-.051
h = 0.3	2,930	-.065
GPA > 3.78		
h = 0.1	1,271	-.028
h = 0.2	1,271	-.061
h = 0.3	1,271	-.118

Robust standard errors are reported in parentheses. *, **, and *** denote statistical significance at the 10, 5, and 1 percent-level, respectively. Estimates are from difference-in-differences kernel matching using an Epanechnikov kernel function with various bandwidth parameters, *h*. We report estimates for students with below average or average high school GPAs (≤ 3.28), above average high school GPAs (> 3.28), and high school GPAs greater than one standard deviation above the mean (> 3.78).

APPENDIX F. ALTERNATIVE COHORTS

Table F1. NMLLS kernel matching estimates with alternative cohort sets

Group	First Declared STEM Major	
	Obs.	Estimate
Full Sample		
1993 – 1999	12,788	.027
1993 – 2000	15,308	.013
1994 – 1999	11,209	.026
1994 – 2000	13,756	.012
GPA ≤ 3.28		
1993 – 1999	6,335	-.036
1993 – 2000	7,564	-.072
1994 – 1999	5,473	-.068*
1994 – 2000	6,725	-.097**
GPA > 3.28		
1993 – 1999	6,451	.091**
1993 – 2000	7,742	.079**
1994 – 1999	5,734	.121***
1994 – 2000	7,029	.112**
GPA > 3.78		
1993 – 1999	2,364	-.025
1993 – 2000	2,860	-.018
1994 – 1999	2,105	-.063
1994 – 2000	2,601	-.054

Robust standard errors are reported in parentheses. *, **, and *** denote statistical significance at the 10, 5, and 1 percent-level, respectively. Estimates are from difference-in-differences kernel matching using an Epanechnikov kernel function with various freshmen cohorts included. We report estimates for students with below average or average high school GPAs (≤ 3.28), above average high school GPAs (> 3.28), and high school GPAs greater than one standard deviation above the mean (> 3.78).

Table F2. NMLLS kernel matching estimates with alternative cohort sets

	Earned STEM Degree	
Group	Obs.	Estimate
Full Sample		
1993 – 1999	4,932	-.024
1993 – 2000	5,953	.001
1994 – 1999	4,438	-.012
1994 – 2000	5,466	.002
GPA ≤ 3.28		
1993 – 1999	1,712	.165*
1993 – 2000	2,037	.165**
1994 – 1999	1,507	.147
1994 – 2000	1,836	.155*
GPA > 3.28		
1993 – 1999	3,219	-.068
1993 – 2000	3,915	-.040
1994 – 1999	2,930	-.051
1994 – 2000	3,629	-.034
GPA > 3.78		
1993 – 1999	1,400	-.104
1993 – 2000	1,711	-.083
1994 – 1999	1,271	-.061
1994 – 2000	1,581	-.062

Robust standard errors are reported in parentheses. *, **, and *** denote statistical significance at the 10, 5, and 1 percent-level, respectively. Estimates are from difference-in-differences kernel matching using an Epanechnikov kernel function with various freshmen cohorts included. We report estimates for students with below average or average high school GPAs (≤ 3.28), above average high school GPAs (> 3.28), and high school GPAs greater than one standard deviation above the mean (> 3.78).