School of Economics

Working Paper Series

Low-performing student responses to state merit scholarships

Christopher Erwin

2019/02

Low-performing student responses to state merit scholarships

February 8, 2019

Christopher Patrick Erwin[1]

ABSTRACT

State merit scholarships may affect academic outcomes for low-performing college students, yet low-performers are generally overlooked in existing literature. New Mexico's lottery scholarship provides tuition-free college to residents meeting a uniquely "low-bar" eligibility criteria. Using administrative data, a discontinuity in eligibility rules identifies local average treatment effects on degree completion and course taking behavior for students with below-average college grades. Results suggest a reduction in time-to-degree corresponding to the scholarship's funding cap, with no overall change in degree completion. Despite modest eligibility requirements related to credit completion, the scholarship increased credit completion among low-achieving students. Some students appear to manipulate scholarship eligibility by taking fewer courses or strategically dropping courses during a qualifying semester in order to secure aid. A bounding exercise suggests partial manipulation of eligibility rules results in selection bias which underestimates the true effect of the scholarship on time to degree and credit completion.

JEL Classification: C14, D04, I22, I23, H75

Keywords: higher education, state merit scholarships, time to degree, college completion

[1]Erwin: Postdoctoral Fellow, New Zealand Work Research Institute, Faculty of Business, Economics, and Law, Auckland University of Technology. Mail: Private Bag 92006, Auckland, New Zealand 1142. Tel: +64 027 545 7774. Email: cpe@unm.edu. Thanks to Benjamin Hansen, Harold Cuffe, Gail Pacheco, and Peer Skov for helpful comments. The author is solely responsible for any errors.

INTRODUCTION

Since the early 1990s, over half of U.S. states appealed to merit-based scholarships to reduce college costs for qualified resident students.[1] State merit scholarships cover a significant proportion of tuition and fees for residents attending in-state public institutions meeting eligibility requirements typically tied to high school GPA, standardized test scores, class rank, or some combination thereof. The shift toward merit scholarships has changed the landscape of financial aid in the United States. Merit scholarships have not merely replaced existing need-based funding, but instead increased the overall amount of financial aid available to students (Bell, Wehde, and Stucky, 2018). In the 1990-1991 academic year, prior to the launch of the first state merit scholarship in Arkansas, states awarded $427.6 million in merit-based financial aid. The proportion of state awards based solely on academic merit was approximately 12 percent. By the 2015-2016 academic year, this amount increased to $2.6 billion, representing 24 percent of all state awards.[2] Given substantial merit-based investments aiming to increase access and achievement in higher education, it remains central whether such programs meet their objectives.

---

[1]Levitz, Jennifer and Scott Thurn, "Shift to Merit Scholarships Stirs Debate," Wall Street Journal, 19 December 2012, https://www.wsj.com/articles/SB10001424127887324481204578175631182640920 (retrieved 10 October 2018).

[2]Figures are in 2016 dollars. From National Association of State Student Grant and Aid Programs, 47th Annual Survey Report on State-Sponsored Student Financial Aid, 2015-2016 Academic Year, https://www.nassgapsurvey.com/survey_reports/2015-2016-47th.pdf (retrieved 10 October 2018).

Stated objectives of state merit scholarships include increasing access to higher education for financially constrained students, improving academic performance, combatting brain drain, and incentivizing top-performing high school students to stay in-state (who would otherwise potentially attend private or out-of-state institutions). Arguments in favor of state merit scholarships typically claim significant positive externalities. Proponents contend that higher education is made more accessible for low-income households, where educational attainment is an effective way to lift families out of poverty. Further, top talent is retained in-state, ultimately benefitting the local economy and positively affecting the learning environment at public institutions. In New Mexico, the impetus for launching the New Mexico Legislative Lottery Scholarship (NMLLS) was to encourage high school completion (Ness, 2008).

Those opposing state merit scholarships argue their lack of a need-based component results in routine funding of college for students from relatively wealthy families who would attend college in absence of the scholarship. In states where programs are funded through lottery revenues, such as New Mexico, some argue merit scholarships serve as a regressive tax: People playing the lottery are often low-income, and their spending subsidizes college for students from high-income families.[3] Despite their controversial nature, state merit scholarships are a testing ground for broader proposals to make college tuition-free in the United States. According to the Education Commission of the States, 36 states considered a total of 131 bills related to free

[3]Binder and Ganderton (2004) find for every low-income student receiving the scholarship at UNM, three students from families with relatively high incomes are awarded the scholarship. Cornwell and Mustard (2002) provide evidence on the regressivity of Georgia's Helping Out Pupils Educationally (HOPE) scholarship.

college from 2014 to 2018.[4] Several recent proposals to make higher education tuition-free at the federal-level in the United States have been entertained.[5]

This study examines the effect of the NMLLS on college completion and course taking behavior at the state's flagship university, the University of New Mexico (UNM). A qualifying semester GPA eligibility cutoff is leveraged to identify treatment effects in a quasi-experimental setting. Because the NMLLS eligibility requires modest academic achievement, local average treatment effects (LATE) are estimated for students with below-average academic performance in college—a group generally overlooked in the literature. Robustness of results to various falsification tests is examined, as are nonrandom heaping in the running variable and the external validity of LATE. Because of the NMLLS's unique design, findings inform how students in the lower end of the ability distribution respond to generous financial aid tied to modest academic requirements.

The NMLLS is arguably the most broadly-available, low-bar, state merit scholarship to be launched in the United States. Over the study period, the state paid 100 percent of residents' tuition for up to four years provided students earned a high school equivalency in New Mexico, enrolled at a public institution in the next regular semester, and earned 12 credits with a minimum C+ average in the first semester of college (which was also tuition-free through the state's Bridge to Success scholarship). Since inception in 1997, the NMLLS has funded 116,531

---

[4]Education Commission of the States, "Free College and Adult Student Populations," https://www.ecs.org/free-college-and-adult-student-populations/ (retrieved 15 August 2018).
[5]In 2015, the Obama administration proposed making the first two years of college tuition-free, and multiple 2016 presidential candidates proposed tuition-free college for students below a household income cap.

recipients with over $785 million in tuition costs. Of those funded, over 55 percent earned at least an associate's degree at one of 15 participating public institutions in New Mexico.[6]

In contrast to other state merit scholarships, eligibility for New Mexico's program is not based on high school, but rather college, performance. For example, Georgia's Helping Out Pupils Educationally (HOPE) scholarship requires a minimum high school GPA of 3.0. West Virginia's Providing Real Opportunities to Maximize In-state Student Excellence (PROMISE) scholarship requires a 3.0 high school GPA and a 22 composite ACT score. Florida's Bright Futures scholarship requires a minimum 3.0 high school GPA, 26 composite ACT score, and 75 hours of community service. Massachusetts's Adams Scholarship requires high school students be in the top quartile of a state standardized test.

This paper examines degree completion and course taking behavior using administrative data on all first-time, full-time, resident freshmen at UNM. To preview results, low-achieving students respond positively to some merit scholarship program features, including semester-based funding caps and credit completion-contingent scholarship eligibility. NMLLS scholars are more likely to complete a degree within four or four-and-a-half years, in line with the program's funding cap. Degree completion at later semesters is unchanged, suggesting a reduction in time to degree without any meaningful increase in overall completion rates. Low-achieving scholarship recipients earn approximately 16 percent more credits compared to nonrecipients. However, because initial eligibility requires completing only 12 credit hours in the

[6]New Mexico Lottery, https://www.nmlottery.com/scholarships.aspx (retrieved 11 October 2018).

qualifying semester, students respond by registering for fewer classes, earning fewer credits, and withdrawing from courses more often in order to secure initial scholarship funding. This manipulation appears partial in nature and results in underestimating true completion effects of the NMLLS.

LITERATURE

Many studies examine state merit scholarships, yet few focus on college outcomes beyond enrollment.[7] Ideally, in addition to improving access to higher education, state merit scholarships increase overall degree completion for attendees, thereby raising the level of educational attainment within the state. As state merit scholarships vary substantially in program design, it is not surprising the literature offers mixed evidence. Several studies employ state-level data from the U.S. Census Bureau to estimate treatment effects. Examining Georgia's HOPE scholarship and Arkansas's Academic Challenge Scholarship, Dynarski (2008) finds a 2.98 percentage point (15.7 percent) increase in the stock of bachelor's degrees (or higher) for the population exposed to such programs. In contrast, Sjoquist and Winters (2012) find no evidence of increased degree completion using a larger sample and adjusting standard errors according to Conley and Taber (2011). In a follow-up study incorporating American Community Survey estimates for years 2001 to 2010, Sjoquist and Winters (2015a) again find no evidence of positive completion effects for the 25 states adopting merit scholarships between 1991 and 2004. More recently, this line of research has been extended by accounting for

[7]Previous work finds the NMLLS significantly increased college enrollment in New Mexico (Binder, Ganderton, and Hutchins, 2002; Binder and Ganderton, 2004).

heterogeneity in program features. Jia (2019) finds that scholarships covering a larger fraction of net price (i.e., tuition less aid received) are associated with increased bachelor's degree attainment. Results suggest analyses assuming homogeneity in state merit scholarships may be misleading. Papers focusing on a single university or state system generally find markedly different results compared to papers assuming homogeneity in eligibility rules.

The earliest study of state merit scholarships and college completion using administrative data examines West Virginia's PROMISE scholarship. Limiting the sample to residents meeting a 3.0 high school GPA requirement, Scott-Clayton (2011) estimates LATE by comparing students in the neighborhood of a minimum 22 ACT composite score cutoff. Robustness of results is assessed by estimating difference-in-differences models using variation in policy timing and residency classification. Results suggest PROMISE increased four-year completion rates by 9.4 percentage points (58.8 percent), with five-year completion rates increasing by 4.5 percentage points (12.2 percent) for students just above the ACT cutoff. Six-year completion rates are not reported, but given a substantial drop in LATE from the fourth- to fifth-year, it is possible that PROMISE may have decreased time to degree in West Virginia without changing overall degree completion at 150 percent of normal time. A bounding exercise is conducted to account for selection bias due to student manipulation of the ACT cutoff score. Also appealing to FRD, Welch (2014) examines completion for community college students funded by Tennessee's HOPE scholarship. Treatment effects are identified for students just above an ACT composite score threshold but below a 3.0 high school GPA threshold (as eligibility is determined by meeting at least one of the cutoffs). The author does not find any evidence of HOPE affecting community college students in terms of completing an associate's or bachelor's degree in Tennessee.

Erwin and Binder (2018) use variation in policy timing and residency classification to estimate completion effects of the NMLLS using difference-in-differences propensity score matching. The authors do not find any evidence of completion effects in the aggregate, but do find divergent effects when disaggregating by student ability and family income. Estimates show less-academically prepared students 11.6 percentage points (38.8 percent) less likely to graduate within six years while more-academically prepared students are 10 percentage points (17.9 percent) more likely to graduate within the same timeframe, relative to nonrecipients. Changes in resident student composition before and after the launch of the NMLLS are central to interpreting results—it appears the low-bar nature of the scholarship resulted in additional marginally-prepared residents attending the state's flagship, possibly overmatching and completing college at lower rates. Note there is at least one other study finding a negative relationship between merit scholarships and degree completion. Cohodes and Goodman (2014) focus on Massachusetts's Adams scholarship, concluding that award eligibility results in a 2.5 percentage point (four percent) decrease in six-year completion rates for students just above a standardized test percentile rank cutoff—a result attributed to students' apparent willingness to trade college costs for college quality. Scholarship recipients tended to forego matriculation at elite private institutions for lower-quality, subsidized public institutions. Compositional change is a key component to understanding the relationship between broad-based merit scholarships and college completion.

The literature on state merit scholarships also offers clues on how students respond to aid in terms of course taking behavior. Cornwell, Lee, and Mustard (2005) provide an early examination of the relationship between state merit scholarship receipt and course taking behavior. The authors use administrative data to study Georgia's HOPE scholarship, finding that students respond to grade-based retention rules by taking fewer courses and withdrawing from courses more often, presumably to maximize the likelihood of scholarship retention. In contrast, Scott-Clayton (2011) finds that West Virginia's PROMISE resulted in more full-load course taking. Welch (2014) finds that community college students completed approximately three (nine percent) more credits by the end of the second year.

Disagreement in the literature is likely driven by heterogeneous program design. Georgia's HOPE scholarship caps funding based on the amount of credits typically required to earn a bachelor's degree, but does not impose any time limit on the number of semesters students can receive funding. In this case, students likely do not have strong incentives to complete college within four or five years. In contrast, students in West Virginia are limited to four years of funding, and appear to respond to state merit aid by completing degrees at significantly higher rates within four years. Tennessee's HOPE scholarship terminates after five years or receipt of a bachelor's degree, whichever occurs first. The NMLLS is capped at four years after the qualifying semester, for a total of four-and-a-half years. As semester funding caps move further away from matriculation, students have increasingly weaker incentives to graduate in a timely manner. This finding is borne out in previous studies examining credit completion and state merit scholarships, and is supported by results herein.

This paper contributes to the literature in three principal ways. First, it estimates a novel and policy-relevant LATE seldom approached in empirical work—the response of below-average-ability students to generous financial aid. Stricter eligibility requirements in other states beget estimates of treatment effects for students with relatively strong academic backgrounds.[8] These students are no doubt better prepared to succeed in college relative to marginally-eligible scholarship recipients in New Mexico. Here, LATE is estimated for compliers in the neighborhood of a 2.5 first semester college GPA—C+ students with below-average academic performance. These are below-average students, many of which may not have attended college without the scholarship. Students with weaker academic preparation likely respond differently to broad-based merit scholarships compared to those with stronger preparation. Second, previous studies of the NMLLS rely on comparability between resident and nonresident students. Because the University of New Mexico is not a popular destination for out-of-state students, it is likely that residents and nonresidents vary considerably on non-observable characteristics. Out-of-state students are more likely to be on athletic scholarship, for example. The identification strategy in this paper does not rely on comparability between residents and nonresidents. Third, the unique structure of New Mexico's lottery scholarship allows direct examination of potential cutoff manipulation strategies. Because students are only eligible after a qualifying semester, outcomes in this semester are used to assess whether students nonrandomly sort around the cutoff. This is key to interpretation—because GPAs are subject to partial manipulation on behalf of students, assessing the possible strategies students employ, and their effectiveness, provides additional context to results.

[8]For example, resulting from the multidimensional nature of West Virginia's eligibility rules, LATE in Scott-Clayton's (2011) sample is estimated for students meeting a 3.0 high school GPA and a composite ACT cutoff score of 22. Estimates of Georgia's HOPE scholarship refer to students meeting a minimum 3.0 high school GPA.

DATA

The sample includes the first three NMLLS-eligible cohorts of first-time, full-time, resident freshmen at UNM, 1997 to 1999, approximately half of which received scholarship funding in their second semester. Included are information on socio-demographics (age, race-ethnicity, gender, family income), academic performance in high school (GPA, composite ACT scores, indication of remedial coursework required in college), and college academic performance by semester (credits earned, college GPA, date of graduation, enrollment by semester, and declared major). Race and ethnicity are referred to collectively as "race-ethnicity" since, at the time, UNM recorded Hispanic identification as a race rather than an ethnicity. For example, students in the sample could identify as black or Hispanic, but not both. Descriptive statistics are shown in Table 1. The sample consists of 3,495 resident students meeting all eligibility requirements for the NMLLS except for the 2.5 qualifying semester GPA cutoff—some surpass the cutoff while others do not. These students are New Mexico residents, recent high school graduates, having completed 12 credit hours during the qualifying semester. Degree completion rates are approximately 16 percent within four years and 56 percent within six years. Binary completion outcomes are cumulative in nature. To illustrate, every student graduating within four years, by definition, graduates within four-and-a-half, five, and six years as well. Average high school GPA for the sample is 3.4. Residents are 55 percent female, with nearly seven percent of all residents required to take remedial mathematics or English at UNM (based on standardized test scores). The university is designated as a Hispanic-Serving Institution with nearly one-third of students identifying as such. Beside family income, student characteristics are complete. Family income is populated only for students completing a Free Application for Federal Student Aid (FAFSA)—approximately half of the sample. It is assumed that non-FAFSA-filers have family incomes more than $40,000, rendering them ineligible for Federal Pell

Grant Program aid. The 1995-1996 Federal Pell Grant End-of-Year Report shows less than 2.5 percent of Pell recipients had family income in excess of $40,000.[9]

In 2000, UNM launched a program meant to increase persistence for first-time, full-time freshmen. Known as the First-Year Learning Communities program, it was primarily a response to low retention rates. Because the program aims to increase retention by providing additional training on adjusting to life in college, it could potentially confound the analysis. It may be difficult to parse out effects of the First-Year Learning Communities program from effects of the NMLLS. Erwin and Binder (2018) entertain the idea of an FRD approach using 1997 to 2008 cohorts, but data fail to pass McCrary's (2008) test for running variable manipulation. Upon reexamination, strong nonrandom sorting around the eligibility cutoff coincides with the launch of the First-Year Learning Communities program, decidedly limiting the sample from 1997 to 1999. To the author's knowledge there are no other concurrent policy changes at the secondary or postsecondary level in New Mexico potentially confounding the analysis.

EMPIRICAL MODEL

Treatment effects are estimated by FRD for resident students earning at least 12 credits in the qualifying semester. Outcomes for complying students just above the 2.5 qualifying semester GPA cutoff are compared to students just below. FRD is an appropriate strategy given the probability of receiving NMLLS funding increases significantly at the qualifying semester GPA cutoff of 2.5, but not from zero to one. There are three explanations as to why FRD, and

---

[9] U.S. Department of Education, Office of Postsecondary Education, 1996-97 Federal Pell Grant Program End-of-Year Report, Table 10-A, https://www2.ed.gov/finaid/prof/resources/data/pell-historical/pell-eoy-1996-97.pdf (retrieved 18 January 2019).

not sharp regression discontinuity, is appropriate. First, university policies allow exceptions for students unable to immediately enroll after receiving a high school credential because of military service or an approved medical condition. Second, in rare cases some students with less than a 2.5 qualifying semester GPA are approved for an exemption through providing evidence of "special circumstances." Third, the NMLLS prioritizes other scholarship monies above those from the Lottery Scholarship Fund (i.e., the NMLLS is structured as a last-dollar scholarship). The probability of treatment is thus not determined entirely by the qualifying semester GPA cutoff, c. Instead,

(1) $$P(w = 1|x) \equiv F(x)$$

where w equals one under NMLLS scholarship receipt and zero otherwise, x is the qualifying semester GPA, and $F(x)$ is a discontinuous function at $x = c$. LATE is found by dividing the discontinuity in the outcome variable by the discontinuity in the probability of treatment at the cutoff.[10] Hahn, Todd, and van der Klaauw (2001) show the treatment effect is numerically identical to the local instrumental variables estimator where an indication of passing the qualifying semester GPA cutoff serves as an instrument for receiving NMLLS funding. The effect of passing the GPA eligibility requirement on NMLLS receipt is identified in the first stage equation

(2) $$NMLLS_i = \alpha_0 + \alpha_1 \mathbf{1}[x_i \geq c] + \alpha_2(x_i - c) + \alpha_3(x_i - c) * \mathbf{1}[x_i \geq c] + \boldsymbol{X\theta} + v_i$$

where $NMLLS_i$ is equal to one if student i receives lottery scholarship funding in the first eligible semester, and zero otherwise. $\boldsymbol{X}$ contains a vector of control variables, including race-ethnicity,

[10]See Imbens and Lemieux (2008) for an introduction to FRD design.

gender, high school GPA, ACT composite scores, family income, and an indication of whether the student was required to take remedial coursework in college. The second stage takes the form

(3) $$Y_i = \beta_0 + \tau_{FRD}\widehat{NMLLS_i} + \beta_1(x_i - c) + \beta_2(x_i - c) * \mathbf{1}[x_i \geq c] + \boldsymbol{X\Gamma} + \varepsilon_i$$

where Y_i is an outcome and τ_{FRD} is the LATE of the NMLLS. Outcomes include degree completion, credits earned, and course withdrawals at different points in time. One common mean squared error (MSE)-optimal bandwidths, h, are found following Calonico *et al.* (2017, 2018). Data-driven approaches avoid selecting arbitrary bandwidths which may be subject to bias. Robustness to alternative data-driven bandwidths is reported in the online appendix.[11] Local linear regression is estimated in the neighborhood of the GPA cutoff, $[c - h < x_i < c + h]$, as is the recommended practice in empirical applications (Cattaneo, Titiunik, and Vazquez-Bare, 2017; Gelman and Imbens, 2018). The use of higher-order polynomials is useful for visual and empirical checks of running variable manipulation, but may be biased in local regression resulting from erratic behavior at the endpoints of fitted curves. There is a bias-variance trade-off at work: Because higher-order polynomials are more flexible and incorporate more data points, variance decreases at the cost of increased bias from over-fitting the curve. Robust bias-corrected confidence intervals are reported following Calonico, Cattaneo, and Titiunik (2014). Standard errors are clustered at the cohort-level.

Nonrandom Heaping in the Qualifying Semester GPA

[11] All appendices are available at the end of the article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

Two identifying assumptions of FRD warrant discussion. First, identification requires qualifying semester GPAs vary smoothly around the 2.5 cutoff in the absence of the NMLLS. This assumption deserves scrutiny as GPAs are discrete, with certain GPAs more common than others. Heaping in college grades may occur for various reasons—institution policies, the curving or centering of grades, instructor distaste for assigning plus or minus grades (e.g., A+, B-), or the proclivity of instructors to help out persistent students just below letter-grade cutoffs with generous, sometimes creative, rounding policies. Barreca, Lindo, and Waddell (2016) show that LATE is biased in the presence of heaping, even when heaping occurs far away from the cutoff. Figure 1 presents a visual check for nonrandom heaping in the running variable. The histogram of qualifying semester GPA (centered at zero) shows a major spike at one-half, with several other spikes occurring at multiples of one-third and one-fourth, dampening toward tails of the distribution. The distribution is not dominated by the zero area, although a spike is observed there. Instead, the distribution shows the most significant mass in the one-half region.

Barreca, Lindo, and Waddell (2016) recommend two strategies to address nonrandom heaping, each having their own advantages and limitations. The first method pools heaped and non-heaped students, flexibly controlling for heaping by allowing different intercepts and slopes for each group. Suppose H_i equals one when the qualifying semester GPA is a multiple of one-third or one-fourth, and zero otherwise. Adding separate intercepts and slopes for heaped students augments equations (2) and (3) such that the first stage becomes

(4) $$NMLLS_i = \gamma_0 + \gamma_1 H_i + \gamma_2 \mathbf{1}[x_i \geq c] + \gamma_3 (x_i - c) + \gamma_4 H_i * (x_i - c) + \gamma_5 (x_i - c) * \mathbf{1}[x_i \geq c] + \boldsymbol{X\Psi} + \varphi_i$$

and the second stage becomes

(5) $$Y_i = \delta_0 + \delta_1 H_i + \tau_{FRD} \widehat{NMLLS_i} + \delta_2(x_i - c) + \delta_3 H_i * (x_i - c) + \delta_4(x_i - c) * \mathbf{1}[x_i \geq c] + \boldsymbol{X\Phi} + \omega_i.$$

The advantage of flexibly controlling for nonrandom heaping is that no observations are lost, and results are potentially less-biased. However, this strategy is limited in the sense that it does not remove all bias from nonrandom heaping. The second strategy estimates treatment effects separately for heaped and non-heaped students. The advantage of this strategy is that it provides unbiased estimates for each group. Unfortunately, splitting the sample in this way often leads to an insufficient number of observations within MSE-optimal bandwidths of the cutoff. Heaped students in the UNM data comprise 842 out of 3,495 (24 percent) of observations, so this strategy often yields estimates of LATE only for non-heaped students. Results from both strategies are reported. Because it provides unbiased estimates, removing heaped students from the sample is the preferred approach.

Manipulation of the Qualifying Semester GPA Cutoff

The second identifying assumption requires qualifying semester GPAs not respond to treatment. In other words, students cannot "game" the GPA cutoff to secure favorable treatment. Since eligibility rules are known to students, and differences in direct college costs vary considerably on either side of the cutoff, this assumption is difficult to maintain. Knowledge of the cutoff may be correlated with performance in college, which would bias results in an unclear direction. Consider low-performing students which would have earned first semester GPAs under 2.5 in the absence of the NMLLS. Assume students prefer to receive NMLLS funding, and there are two ways of securing it: either by increasing effort or by taking a lighter course load. Students that increase effort may be more immersed in the college experience and more motivated than others. If these motivated students surpass the qualifying semester GPA cutoff, then degree completion rates above the cutoff will be higher than they would in the absence of the scholarship, and estimates are biased upward. Students choosing to instead take a lighter course load may be less immersed in the college experience and less motivated compared to others. If these less motivated students surpass the qualifying semester GPA cutoff, then degree completion rates above the cutoff will be lower than they would in the absence of the scholarship, and estimates are biased downward.[12]

McCrary (2008) offers a formal yet intuitive test of running variable manipulation. If students are able ensure a favorable outcome by manipulating the qualifying semester GPA around the known cutoff of 2.5, then there should exist an excess mass of students just above the cutoff. McCrary's test is informative only when manipulation is monotonic—that is, occurring only in one direction. Monotonic and partial manipulation can be expected, as qualifying semester GPAs are only conceivably manipulated upwards in order to secure funding, and students do not have full control over their grades. Lee (2008) demonstrates that partial manipulation generally does not threaten identification, and in such cases LATE are recovered without difficulty.[13] Importantly, if students are only able to partially manipulate qualifying semester GPAs, so there remains a substantial stochastic component determining the qualifying semester GPA, then McCrary's test will not find a failure of identification. In the case of the

[12]Carruthers and Özek (2016) make the same argument in the context of Tennessee's HOPE scholarship.
[13]See McCrary (2008) for additional discussion of treatment effects under partial manipulation of the running variable.

NMLLS, low-performing students near the cutoff may have little control over their grades, even if taking a lighter course load or easier courses.

Figure 2 presents a visual representation of McCrary's (2008) test using local polynomial density estimation methods developed by Cattaneo, Jansson, and Ma (2018)—methods shown to have improved size and power compared to other tests of density continuity. Quadratic curves are fit on each side of the 2.5 qualifying semester cutoff with ninety-five percent confidence intervals shown in gray. One is not able to reject the null hypothesis of continuity in the running variable around the 2.5 cutoff, evidence that qualifying semester GPAs are not systematically manipulated by students (or are, but only partially so). Appendix A, available online, presents empirical results of manipulation tests under various conditions for 1997 to 1999 cohorts.[14]

Another standard practice in testing for nonrandom sorting is to use predetermined covariates as outcomes. Because it is known *a priori* there cannot be any effect of the policy change on such variables, tests are considered a check for placebo treatment effects. The presence of discontinuities in predetermined covariates may be informative about student sorting around the cutoff. In our case, the unique structure of the NMLLS affords the researcher to also examine qualifying semester outcomes, thereby directly testing whether students adopt certain common strategies to manipulate GPAs, such as reducing course loads or increasing course withdrawal in the qualifying semester. As mentioned above, the literature on state merit aid provides several examples of students modifying their behavior to increase the likelihood of aid

[14]All appendices are available at the end of the article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

receipt.[15] Table 2 presents results of FRD models using qualifying semester course taking outcomes. Results show that students just above the GPA cutoff register for .69 (4.8 percent) fewer credits, earn 1.15 (8.4 percent) fewer credits, and withdraw .23 (40.5 percent) more credits in the qualifying semester relative to students just below the cutoff. Estimates are presented graphically in Appendix B.[16] These findings may be explained by two different strategies to manipulate the cutoff. The first strategy is simply registering for a smaller course load (e.g., 12 credits) as it is easier to manage, and more time may be devoted to each course. Another strategy is enrolling in a full course load (e.g., 15 credits) and withdrawing from courses with expected grades which put one in jeopardy of not meeting the 2.5 GPA eligibility cutoff. Students likely participated in both of these strategies. Because only low-ability students need to game a relatively low qualifying semester GPA requirement, estimates of LATE are likely biased downward. Accordingly, main findings are cautiously presented as lower bounds of the true treatment effects of the NMLLS.

Another falsification test inspects for treatment effects around false cutoffs in the qualifying semester GPAs of 2.3 and 2.7. This provides a check for discontinuities in other

[15]Cornwell, Lee, and Mustard (2005) find evidence that students in Georgia respond to the HOPE scholarship with lighter course loads and increased withdrawals. In a follow-up paper, Cornwell, Lee, and Mustard (2008) find that HOPE induced more students to major in education. Sjoquist and Winters (2015b) find that state merit scholarships result in lower production of degrees in science, technology, engineering, and mathematics. Of course, students may also manipulate the GPA cutoff by taking easier courses or pursuing easier majors, but data do not contain information needed to test for these responses.

[16]All appendices are available at the end of the article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

regions of the running variable distribution. For proper identification, there should be no discontinuities in outcomes other than those at the cutoff value. Results of models using false cutoffs are not statistically significant, and are presented in Appendix C.[17]

RESULTS

Figure 3 plots the first stage of the FRD model; the likelihood of receiving NMLLS funding given a qualifying semester GPA. Linear fitted lines are shown on each side of the 2.5 qualifying semester GPA cutoff, which is centered at zero. Points depict within-bin means of NMLLS receipt, which are evenly spaced, where the number of bins is chosen to approximate the unknown regression function (see Calonico, Cattaneo, and Titiunik, 2015 for details). Other regression discontinuity plots use similar conventions for comparability. Visually, the likelihood of receiving NMLLS jumps by approximately 70 percentage points at the cutoff. Note the fuzzy nature of the discontinuity: Some students slightly below the cutoff are awarded funding while many above the cutoff are not. Figure 4 to Figure 6 present regression discontinuity plots. Graphs hint heuristically to potentially meaningful LATE of the NMLLS. Figure 4 shows jumps in graduation rates around the cutoff at four and four-and-a-half years. There is little graphical support for a completion effect at the five- or six-year mark. Plots in Figure 5 and Figure 6 are quite noisy.

Empirical estimates of degree completion are presented in Table 3. The first column offers estimates pooling heaped and non-heaped students, which for reasons mentioned in the

[17]All appendices are available at the end of the article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

previous section, are likely biased. The second column again pools heaped and non-heaped students, but allows for different slopes and intercepts for each group. These estimates are strikingly similar to those not controlling for heaped observations. The preferred specification is presented in column (3), where local linear regression is used to construct point estimates for non-heaped students with qualifying semester GPAs not multiples of one-third or one-fourth. The number of effective observations below the cutoff is denoted as N_w^-, while effective observations above the cutoff is denoted as N_w^+. One common MSE-optimal bandwidths are denoted as h. The triangular kernel function is used to construct polynomial estimators, as it is shown to be optimal for estimating conditional means at boundary points (Fan and Gijbels, 1996). First-stage results suggest that exceeding the 2.5 qualifying semester GPA increases the likelihood of receiving NMLLS funding by approximately 68 percent. In the second stage, the NMLLS is found to increase degree completion within four years for non-heaped students by 10.8 percentage points (65.9 percent). The NMLLS is shown to increase degree completion within four-and-a-half years by 14.1 percentage points (49 percent) for non-heaped students just above the qualifying semester GPA cutoff. Because the NMLLS funds a maximum of four years after the student's qualifying semester, for a total of four-and-a-half years, it is not surprising that completion effects are lumped at the four and four-and-a-half year marks. Completion effects are noisily estimated at later periods, evidence the lottery scholarship shortened time to degree while not having any overall effect on six-year graduation rates.

Course taking behavior is examined to better understand the mechanism behind college completion. If scholarship recipients are more likely to graduate within four-and-a-half years, then we may observe a higher level of credits attained within that timeframe relative to students that did not receive NMLLS funding. Yet, this may not be evident for various reasons. College credits may be classified into two categories: excess credits, which do not count towards the student's degree, and efficient credits, which count toward the degree. Many students accumulate excess credits through academic exploration or poor academic advising. Increased graduation at four- and four-and-a-half years, by definition, means more students accumulating 120 efficient credit hours within four and four-and-a-half years. While the sample does not distinguish between excess and efficient credits, pure substitution of excess credits for efficient credits would be marked by no change in overall credit completion coupled with a decrease in time to degree. Table 4 offers evidence of this response. The preferred specification in column (3) offers no evidence of any effect on credit accumulation over time, supporting the notion that scholarship recipients took more efficient paths to degree completion. No change in overall credit accumulation may also be explained by a different response: students responding to 12-credit per semester eligibility rule by registering for more classes at the start of semesters, then subsequently dropping courses where a poor grade may jeopardize scholarship renewal. To test whether students respond in this manner, the total number of credits withdrawn is examined as another secondary outcome. Results of credit withdrawal models are presented in Table 5. None of the specifications provide evidence of students responding to the NMLLS by withdrawing from courses more frequently in order to maintain scholarship eligibility. Instead, Table 3 through Table 5 support students responding to the lottery scholarship by decreasing time to degree and substituting away from excess credits in favor of efficient credits, with little discernable impact on course withdrawal.

Appendix D presents estimates using three other data-driven bandwidths as discussed in Calonico *et al.* (2017).[18] The first two columns present MSE-optimal bandwidths, while the last two columns present coverage error rate (CER)-optimal bandwidths. Estimates are shown for one common bandwidth as well as for different bandwidths on either side of the cutoff. Results are not sensitive to the choice of optimal bandwidth.

EXTERNAL VALIDITY

Recent methodological developments in regression discontinuity focus on the external validity of LATE. Dong and Lewbel (2015) and Cerruli *et al.* (2017) present treatment effect derivatives (TED) and complier probability derivatives (CPD) as measures used to assess the stability of estimates. A TED significantly different from zero suggests treatment effects are likely different given small changes in the cutoff value. In the case of FRD, a nonzero CPD also provides evidence of instability, as small changes in the cutoff value result in significant changes in the population of compliers. Both TED and CPD can also be used to extrapolate treatment effects to other (nearby) areas of the running variable distribution under the assumption of local policy invariance. These measures provide insight into how LATE may change given a marginal increase or decrease in the eligibility cutoff value even though no such change is observed. Cerulli *et al.* (2017) offer guidelines for interpreting relative TED and relative CPD: External validity of LATE is threated when TED is nonzero and relative TED (LATE divided by the product of the bandwidth and the TED) is less than one or two. Similarly, instability in the

[18]All appendices are available at the end of the article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

FRD context is indicated by a nonzero CPD and relative CPD (first stage jump in treatment probability divided by the product of the bandwidth and CPD) less than one.

Appendix E, available online, offers detailed results of tests for external validity. TED and CPD are small and indistinguishable from zero, suggesting those with qualifying semester GPAs near 2.5 can be expected to experience treatment effects similar in magnitude. Extrapolation to other areas of the running variable distribution is possible when TED equals the marginal threshold treatment effect, which is true only when local policy invariance holds. Local policy invariance requires that one's expected outcomes do not change if the cutoff marginally changed for all other compliers. Because of general equilibrium effects, the assumption is not likely to hold for the NMLLS. For example, if the qualifying semester GPA cutoff was lowered, enrollment would likely increase, resulting in less individual attention from instructors or academic advisors, and decreased completion rates. Because TED and CPD are not statistically different from zero in our case, extrapolating LATE away from the cutoff is not insightful—LATE is expected to be similar in magnitude for other cutoff values of the qualifying semester GPA.

SELECTION BIAS FROM CUTOFF MANIPULATION

As discussed in the empirical model section, falsification tests find that some low-achieving students successfully manipulate the qualifying semester GPA cutoff by taking fewer credits or strategically dropping courses. Because such students are likely lower-ability, or perhaps less motivated or engaged in college, successful manipulators have the potential to lower average academic achievement for scholarship recipients, resulting in underestimating the true effect of the NMLLS on college and credit completion. This argument motivates conducting a bounding exercise removing suspected manipulators from the analysis. Due to the unique structure of the NMLLS, where data allow the researcher to partially observe potential manipulation strategies, a very simple method which excludes suspected manipulators is adopted. I make the extreme assumption that every scholarship recipient with a qualifying semester GPA above, but near, the 2.5 cutoff who either registers for the minimum number of credit hours for initial eligibility, or drops a course after registering for 15 credit hours in the qualifying semester, is a manipulator. I define manipulators over two different ranges of qualifying semester GPA: those meeting the 2.5 cutoff but not above 2.75, and those meeting the 2.5 cutoff but not above three. Depending on the definition, removal of suspected manipulators further trims the sample from 2,653 to 2,578 or 2,481, decreases of approximately three and six percent, respectively. This strategy has the advantage of being far simpler than previous studies trimming the sample to address nonrandom selection in regression discontinuity designs (Lee, 2009; Scott-Clayton, 2011; Gerard, Rokkanen, Roth, 2018). A notable limitation of this strategy is that only one source of potential manipulation is investigated—other strategies, including taking easier courses or pursuing easier majors, are not observed.

Table 6 presents results of the bounding exercise for completion outcomes. As predicted, removing lower-achieving, less motivated manipulators significantly increases estimated treatment effects of the NMLLS. The estimated four-year completion effect increases appreciably from 10.8 percentage points (65.9 percent) in column (1) to 17.2 percentage points (97.7 percent) in column (2). The four-and-a-half-year completion effect is nearly unchanged, falling slightly from 14.2 percentage points (49.3 percent) to 13.6 percentage points (45.2 percent). Completion effects at later points remain not statistically different from zero, again supporting the notion that the NMLLS reduces time to degree while having no overall impact on degree completion. When using the broader definition of manipulation in column (3), the completion effect at four years is 15.3 percentage points (84.5 percent) and completion at four-and-a-half years is not statistically different from zero. Table 7 presents results of the bounding exercise for credit completion outcomes. Estimates at the three-year mark and later are now all statistically significant using either definition of manipulation. Low-achieving scholarship recipients earn 16.2 more credits (22.7 percent) compared to nonrecipients after three years at UNM. Estimates at later periods are similar in magnitude and statistically significant. Table 8 shows that, as with main results in column (1), low-achieving students did not withdraw from courses more often than nonrecipients.

CONCLUSIONS

Leveraging a discontinuity in eligibility rules, this paper examines how low-achieving students respond to low-bar state merit scholarships. Tests for pseudo treatment effects reveal that scholarship recipients registered for fewer credits, earned fewer credits, and withdrew from more classes in the qualifying semester, presumably to increase the likelihood of scholarship receipt. Formal tests of manipulation in the qualifying semester GPA cutoff are favorable, suggesting that low-achieving college students are only able to partially manipulate grades. Heaping in the qualifying semester GPA is addressed by removing students with GPAs that are multiples of one-third or one-fourth. First stage results show that qualifying semester GPA is highly predictive of receiving NMLLS funding. Second stage results find that overall completion rates are not affected, though students respond to scholarship eligibility caps by reducing time to degree. Abounding exercise is conducted which aims to purge the sample of students that observably manipulated the qualifying semester GPA cutoff. In such models, there is evidence that the NMLLS results in increased credit attainment beginning at the end of the third year. An increase in credits earned paired with shorter time to degree implies students respond to merit scholarship eligibility requirements by embarking on more efficient paths to degree completion. Tests for the external validity of LATE find that treatment effects for those *near* the discontinuity are similar to those *at* the discontinuity.

Results largely agree with previous work using administrative data to study the relationship between state merit aid and college completion. Existing literature finds that state merit scholarships affect *how* students graduate, not necessarily *if* they graduate. Program features determine how students respond, and programs across states are very different. Like Scott-Clayton (2011) and Welch (2014), I find NMLLS recipients earned significantly more credits than nonrecipients. Like Cornwell, Lee, and Mustard (2005), I find that students take fewer courses and withdraw from courses more often in order to ensure scholarship eligibility, although this behavior only occurs during the qualifying semester. Results are also in-line with Garibaldi *et al.* (2012), who show that students reduce time to degree when faced with increasing costs of persistence beyond normal time. Since the NMLLS provides funding for a maximum of four-and-a-half years, students face substantially higher costs starting in the fifth year, which prompts students to finish earlier. Overall, results show that even marginally prepared students respond favorably to generous financial aid.

The NMLLS has seen significant changes since its launch in 1997. In the 2014-2015 academic year the scholarship was capped at eight semesters in total (including the qualifying semester). Eligibility requirements were increased to 15 credits earned per semester. More recently, a 2017 budget crisis reduced funding to approximately 60 percent of tuition and it is unclear whether full funding will be restored in the future. The 2017 Regular Session resulted in the passage of SB 420. Now signed into law, legislation allows students to take a "gap" year after high school and still remain eligible for the NMLLS. In other words, New Mexico's program does not at all operate as it did for 1997 to 1999 cohorts. It is unclear how recent changes will impact student achievement, but changes are likely to increase program costs. Consequently, raising the qualifying semester GPA cutoff is worth considering, as it would limit the program scope and improve the financial health of the Lottery Scholarship Fund. Tests for external validity suggest a small increase in the eligibility cut off could be instituted without sacrificing the benefits of shorter time to degree and more efficient course taking.

REFERENCES

Barreca, A. I., Lindo, J. M., & Waddell, G. R. (2016). Heaping-Induced bias in regression-discontinuity designs. Economic Inquiry, 54, 268—293.

Bell, E., Wehde, W., & Stucky, M. (2018). Supplement or supplant? Estimating the impact of state lottery earmarks on higher education funding. Education Finance and Policy, 1—64.

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

Binder, M., & Ganderton, P.T. 2004. The New Mexico lottery scholarship: Does it help minority and low-income students? In State merit scholarships and racial inequality, edited by Donald E. Heller and Patricia Marin, pp. 101—122. Cambridge, MA: The Civil Rights Project, Harvard University.

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2017). rdrobust: software for regression discontinuity designs. Stata Journal, 17, 372—404.

_______. (2018). Regression discontinuity designs using covariates. Review of Economics and Statistics, 1—30.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. Econometrica, 82, 2295—2326.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2015). Optimal data-driven regression discontinuity plots. Journal of the American Statistical Association, 110, 1753—1769.

Carruthers, C. K., & Özek, U. (2016). Losing HOPE: Financial aid and the line between college and work. Economics of Education Review, 53, 1—15.

Cattaneo, M. D., Jansson, M., & Ma, X. (2018). Manipulation testing based on density discontinuity. The Stata Journal, 18, 234—261.

Cattaneo, M. D., Titiunik, R., & Vazquez-Bare, G. (2017). Comparing inference approaches for RD designs: A reexamination of the effect of head start on child mortality. Journal of Policy Analysis and Management, 36, 643—681.

Cerulli, G., Dong, Y., Lewbel, A., & Poulsen, A. (2017). Testing stability of regression discontinuity models. In Regression Discontinuity Designs: Theory and Applications (pp. 317—339). Bingley, United Kingdom: Emerald Publishing Limited.

Cohodes, S. R., & Goodman, J. S. (2014). Merit aid, college quality, and college completion: Massachusetts' Adams scholarship as an in-kind subsidy. American Economic Journal: Applied Economics, 6, 251—285.

Conley, T. G., & Taber, C. R. (2011). Inference with “difference in differences” with a small number of policy changes. The Review of Economics and Statistics, 93, 113—125.

Cornwell, C. M., & Mustard, D. B. (2002). Race and the effects of Georgia's HOPE scholarship. In Who Should We Help? The Negative Social Consequences of Merit Scholarships, 57—72.

Cornwell, C. M., Lee, K. H., & Mustard, D. B. (2005). Student responses to merit scholarship retention rules. Journal of Human Resources, 40, 895—917.

_______. (2008). The effects of state-sponsored merit scholarships on course selection and major choice in college. Working paper, University of Georgia, Athens.

DesJardins, S. L., Ahlburg, D. A., & McCall, B. P. (2002). A temporal investigation of factors related to timely degree completion. The Journal of Higher Education, 73, 555—581.

Dong, Y., & Lewbel, A. (2015). Identifying the effect of changing the policy threshold in regression discontinuity models. Review of Economics and Statistics, 97, 1081—1092.

Dynarski, S. (2008). Building the stock of college-educated labor. Journal of Human Resources, 43, 576—610.

Erwin, C. P., & Binder, M. (2018). Does broad-based merit aid improve college completion? Evidence from New Mexico's lottery scholarship. Education Finance and Policy, 1—77.

Fan, J., & Gijbels, I. (1996). Local polynomial regression and its applications. London: Chapman and Hall.

Garibaldi, P., Giavazzi, F., Ichino, A., & Rettore, E. (2012). College cost and time to complete a degree: Evidence from tuition discontinuities. Review of Economics and Statistics, 94, 699—711.

Gelman, A., & Imbens, G. (2018). Why high-order polynomials should not be used in regression discontinuity designs. Journal of Business & Economic Statistics, 1—10.

Gerard, F., M. Rokkanen, & C. Rothe. (2018) Bounds on treatment effects in regression discontinuity designs with a manipulated running variable. February 2018 version. NBER Working Paper 22892.

Hahn, J., Todd, P., & van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. Econometrica, 69, 201—209.

Imbens G. W., & Lemieux, T. (2008). Regression discontinuity designs: a guide to practice. Journal of Econometrics, 142: 615—635.

Jia, N. (2019). Heterogeneous effects of merit scholarships: do program features matter? Applied Economics, 1—17.

Lee, D. S. (2008). Randomized experiments from non-random selection in U.S. House elections. Journal of Econometrics, 142, 675—697.

_______. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. The Review of Economic Studies, 76, 1071—1102.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. Journal of Econometrics, 142, 698—714.

Ness, E. C. (2008). Merit aid and the politics of education. In P. G. Altbach (Ed.), Studies in higher education. New York, NY: Routledge Press.

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.

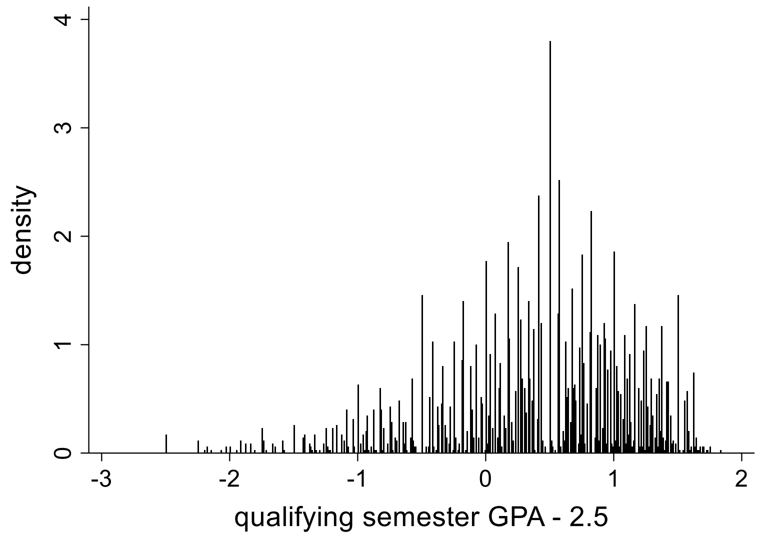
Sjoquist, D. L., & Winters, J. V. (2012). Building the stock of college-educated labor revisited. Journal of Human Resources, 47, 270—285.

_______. (2015a). State merit-based financial aid programs and college attainment. Journal of Regional Science, 55, 364—390.

_______. (2015b). State merit aid programs and college major: A focus on STEM. Journal of Labor Economics, 33, 973—1006.

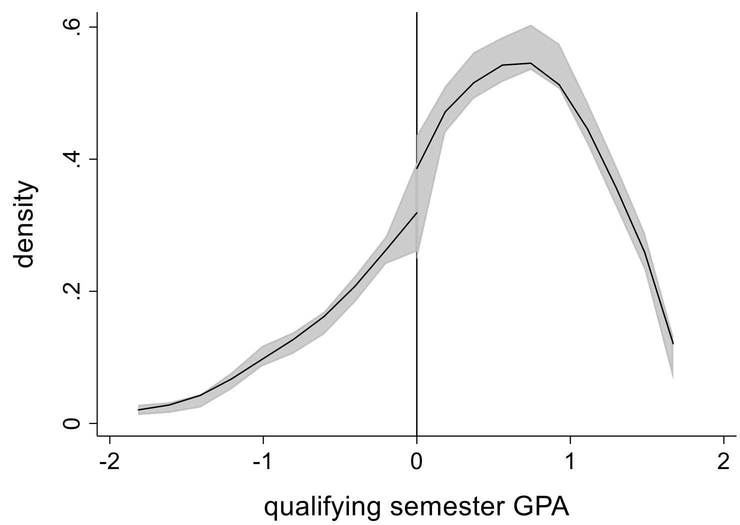
Van der Klaauw, W. (2002). Estimating the effect of financial aid offers on college enrollment: A regression–discontinuity approach. International Economic Review, 43, 1249—1287.

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.



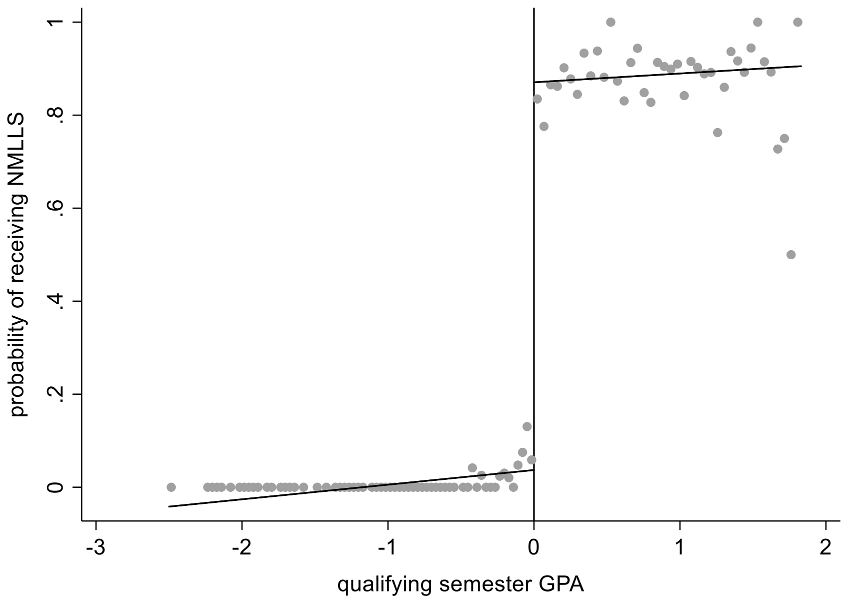
Note: histogram bin widths are .01, as qualifying GPAs are rounded to the nearest hundredth. The running variable is centered at zero.

Figure 1. Histogram of Qualifying Semester GPA



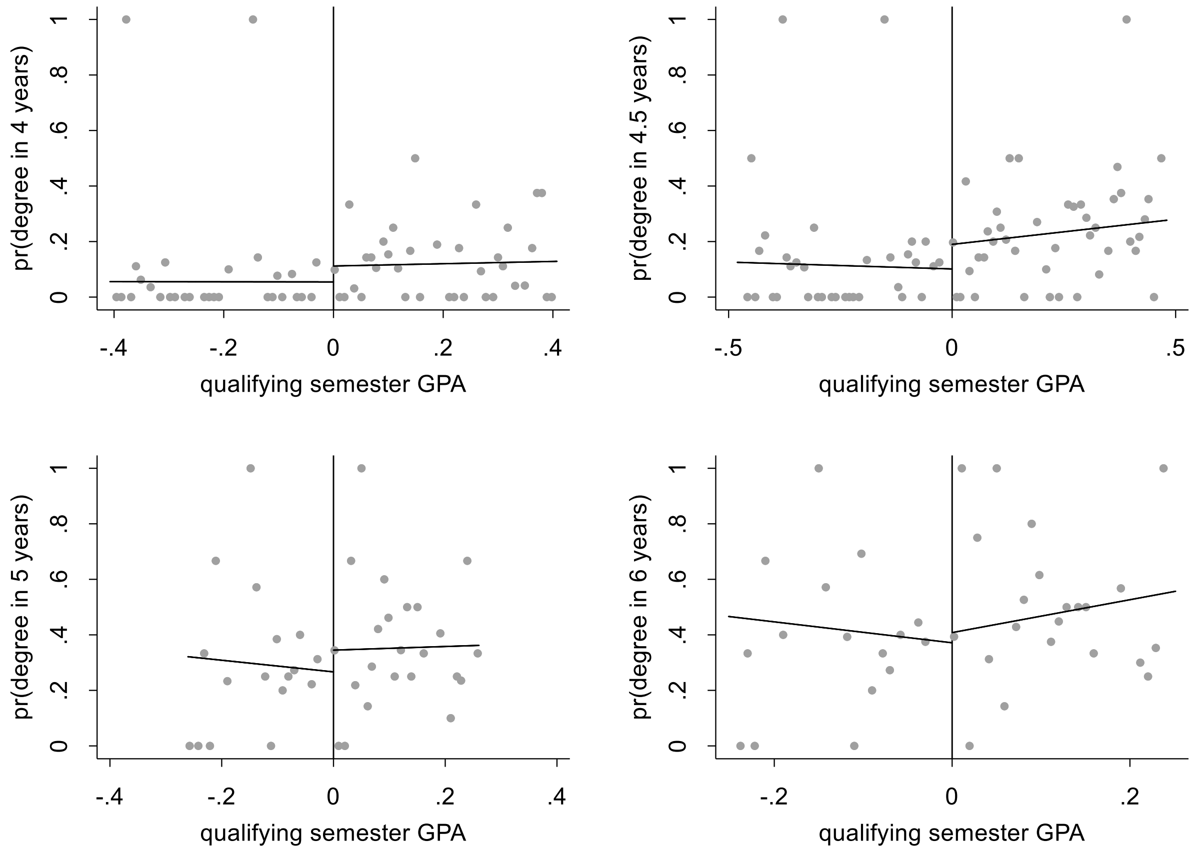
Note: common MSE-optimal data-driven bandwidths are used. A quadratic is fit on each side of the GPA cutoff, which is centered at zero. Ninety-five percent confidence intervals are shown in gray. The triangular kernel is used to construct local polynomial estimators.

Figure 2. Visual Presentation of McCrary's (2008) Test for Running Variable Manipulation



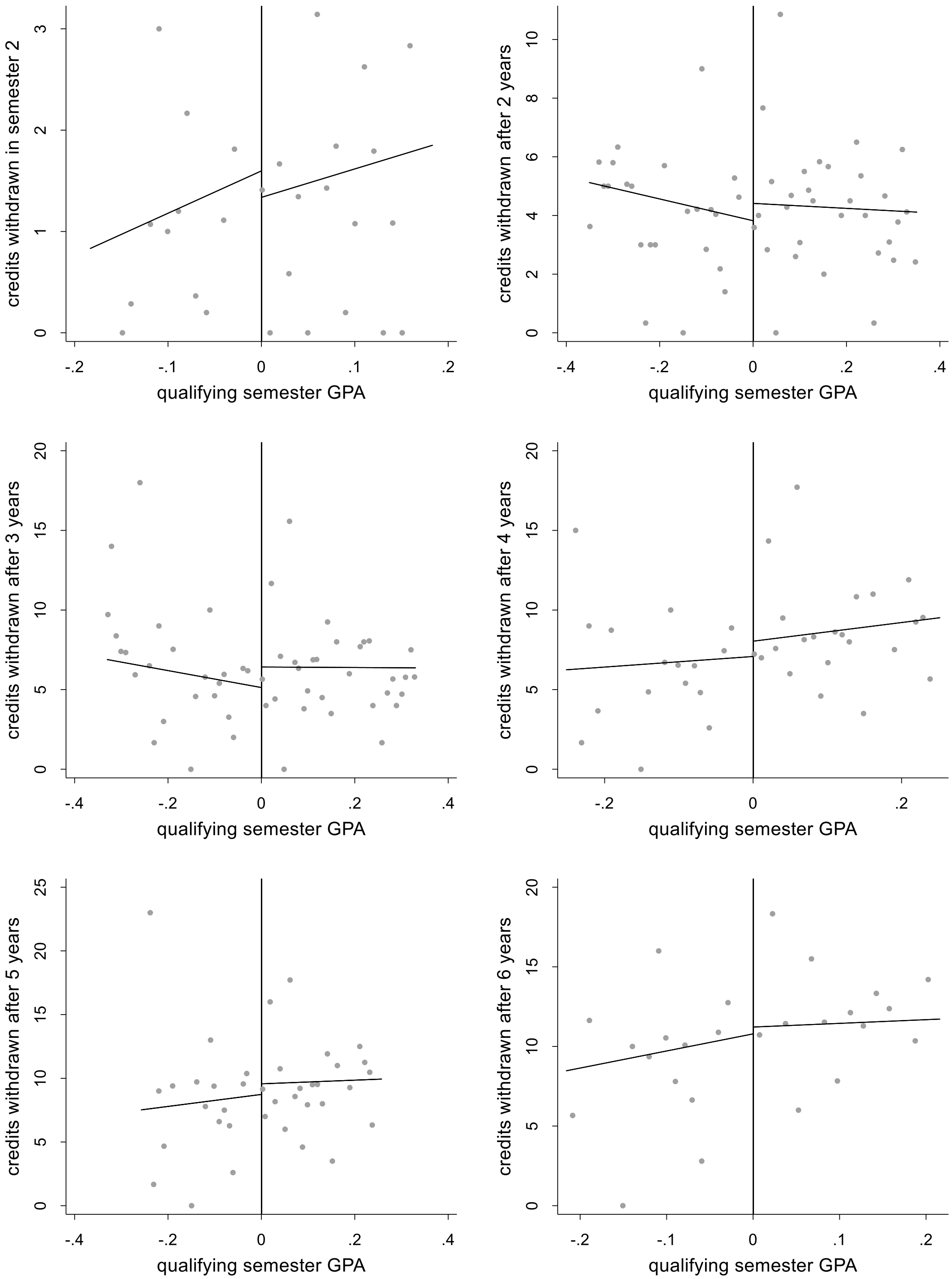
Note: points depict the within-bin sample average of NMLLS receipt probability by qualifying semester GPA. A linear fit has been added below and above the cutoff. Binned means of qualifying semester GPA with evenly spaced bins are chosen optimally to approximate the underlying unknown regression function. The triangular kernel function is used to construct global polynomial estimators.

Figure 3. Jump in Treatment Probability Around the Qualifying Semester GPA Cutoff, 1997 to 1999 cohorts



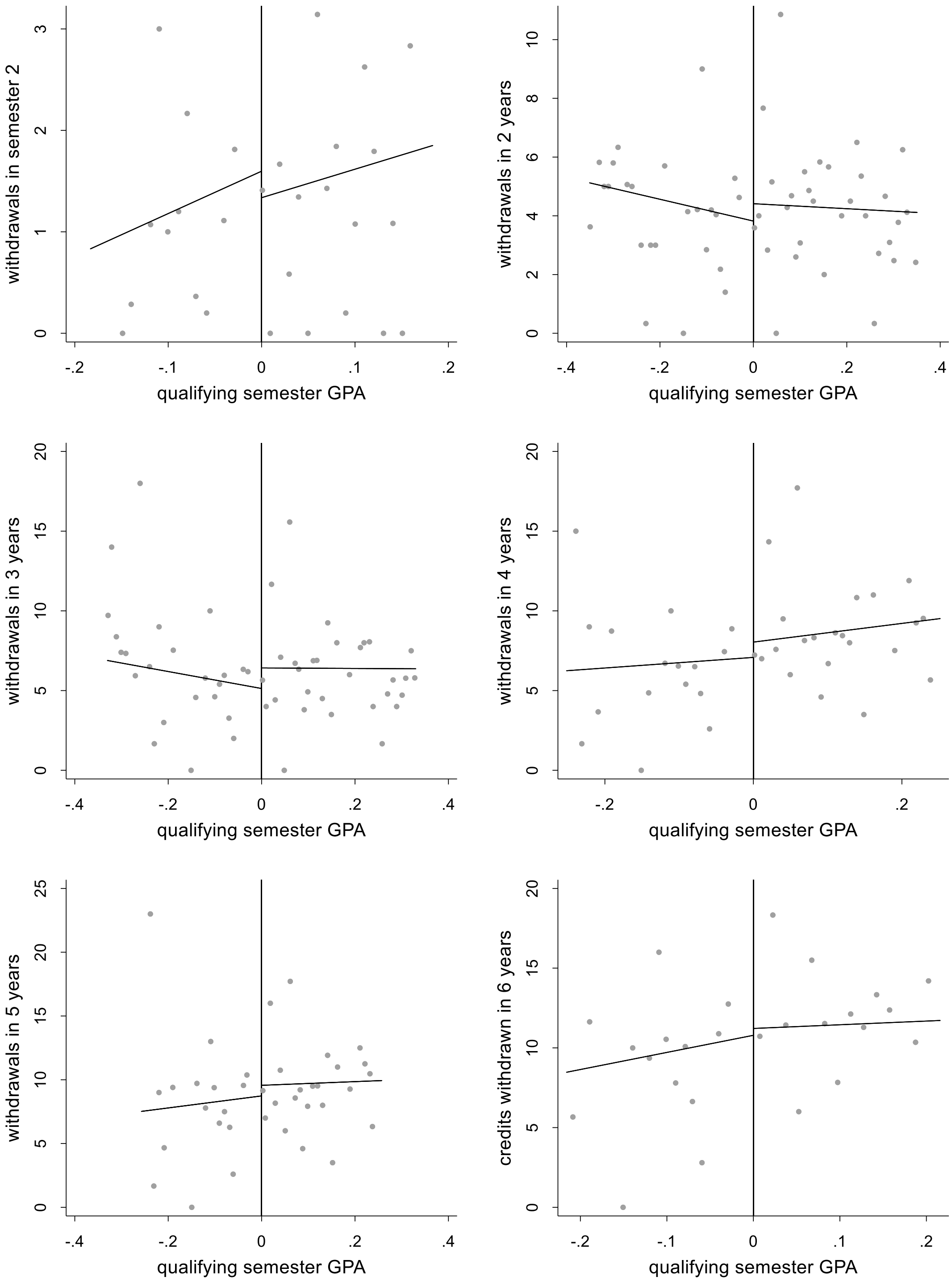
Note: observations nonrandomly heaped at multiples of one-half and one-third have been removed. Binned means of qualifying semester GPA with evenly spaced bins are chosen optimally to approximate the underlying unknown regression function. A linear fit has been added below and above the cutoff centered at zero. The triangular kernel is used to construct the global polynomial estimators.

Figure 4. Visual Presentation of LATE, College Completion, 1997 to 1999



Note: observations nonrandomly heaped at multiples of one-half and one-third have been removed. Binned means of qualifying semester GPA with evenly spaced bins are chosen optimally to approximate the underlying unknown regression function. A linear fit has been added below and above the cutoff centered at zero. The triangular kernel is used to construct the global polynomial estimators.

Figure 5. Visual Presentation of LATE, Credit Accumulation, 1997 to 1999



Note: observations nonrandomly heaped at multiples of one-half and one-third have been removed. Binned means of qualifying semester GPA with evenly spaced bins are chosen optimally to approximate the underlying unknown regression function. A linear fit has been added below and above the cutoff centered at zero. The triangular kernel is used to construct the global polynomial estimators.

Figure 6. Visual Presentation of LATE, Course Withdrawals, 1997 to 1999

Table 1. Descriptive statistics, 1997 to 1999 cohorts

qualifying semester credits:	
registered	14.192 (1.674)
earned	13.771 (1.763)
withdrawn	.570 (1.215)
bachelor's degree within (years):	
4	.164
4.5	.288
5	.460
6	.559
credits earned within (years):	
1	27.049 (4.122)
2	50.069 (13.799)
3	70.403 (24.582)
4	89.592 (36.051)
5	101.312 (42.211)
6	106.507 (44.797)
credits withdrawn within (years):	
1	1.857 (2.622)
2	3.757 (4.253)
3	5.387 (6.626)
4	7.080 (7.033)
5	8.226 (8.154)
6	9.761 (9.534)
female	.553
high school GPA	3.435 (.450)
composite ACT	24.007 (3.240)
required remedial coursework	.066
family income < $40,000	.176
family income < $20,000	.078
white	.576
Hispanic	.319
Asian	.040
American Indian	.034
black	.017
declined to state race-ethnicity	.014
observations	3,495

Source: Freshmen Tracking System, Office of Institutional Analytics, University of New Mexico. Standard deviations are in parentheses.

Table 2. Tests for placebo treatment effects

Outcome		Outcome	
qualifying semester credits registered	-.685*	family income < \$20,000	.016
standard error	.369	standard error	.048
$N_W^-\|N_W^+$	416 | 811	$N_W^-\|N_W^+$	358 | 674
h	.484	h	.406
qualifying semester credits earned	-1.151***	female	-.022
standard error	.438	standard error	.083
$N_W^-\|N_W^+$	414 | 809	$N_W^-\|N_W^+$	416 | 811
h	.451	h	.474
qualifying semester credits withdrawn	.231*	Hispanic	-.078
standard error	.119	standard error	.088
$N_W^-\|N_W^+$	528 | 1181	$N_W^-\|N_W^+$	354 | 660
h	.668	h	.375
high school GPA	-.100	black	.021
standard error	.126	standard error	.054
$N_W^-\|N_W^+$	467 | 944	$N_W^-\|N_W^+$	356 | 669
h	.513	h	.397
composite ACT	-.753	Asian	.010
standard error	.604	standard error	.030
$N_W^-\|N_W^+$	505 | 1096	$N_W^-\|N_W^+$	469 | 951
h	.620	h	.542
required remedial coursework	.057	American Indian	-.023
standard error	.061	standard error	.048
$N_W^-\|N_W^+$	340 | 628	$N_W^-\|N_W^+$	416 | 811
h	.364	h	.496
family income < \$40,000	-.051	declined to state race-ethnicity	.008
standard error	.065	standard error	.021
$N_W^-\|N_W^+$	394 | 763	$N_W^-\|N_W^+$	273 | 504
h	.421	h	.294

Note: estimates are based on fuzzy regression discontinuity models using the 2.5 GPA cutoff during the student's first semester at UNM (i.e., qualifying semester). A triangular kernel is used to construct local polynomial estimators. Common mean squared error-optimal bandwidths, h, determine the neighborhood of the GPA cutoff examined. Local linear regression is used to construct point estimators. Robust standard errors are clustered at the cohort-level. *, **, and *** signify statistical significance at the 10, five, and one percent-levels, respectively.

Table 3. FRD LATE, degree completion, accounting for heaping in qualifying GPA

	(1)	(2)	(3)
	Baseline	Flexible controls	Multiples of 1/3 & 1/4 removed
first stage: NMLLS receipt	.687***	.684***	.682***
standard error	.036	.037	.030
degree within four years	.078	.077	.108**
standard error	.055	.054	.046
$N_W^- \| N_W^+$	358 | 680	355 | 668	272 | 556
h	.419	.381	.413
degree in 4.5 years	.134	.131	.142*
standard error	.089	.088	.075
$N_W^- \| N_W^+$	416 | 811	504 | 1085	330 | 687
h	.496	.592	.482
degree within five years	.102	.103	.082
standard error	.109	.113	.127
$N_W^- \| N_W^+$	469 | 951	504 | 1096	168 | 310
h	.540	.613	.246
degree within six years	.041	.026	.017
standard error	.062	.052	.114
$N_W^- \| N_W^+$	504 | 1092	467 | 948	169 | 313
h	.608	.530	.264
total observations	3,495	3,495	2,653

Note: estimates are based on fuzzy regression discontinuity models using the 2.5 GPA cutoff during the student's first semester (i.e., qualifying semester). A triangular kernel is used to construct local polynomial estimators. Common mean squared error-optimal bandwidths, h, determine the neighborhood of the GPA cutoff examined. Local linear regression is used to construct point estimators. Robust standard errors are clustered at the cohort-level. First-stage results are from models of four-year completion rates. *, **, and *** signify statistical significance at the 10, five, and one percent-levels, respectively.

Table 4. FRD LATE, credits earned, accounting for heaping in qualifying GPA

	(1)	(2)	(3)
	Baseline	Flexible controls	Multiples of 1/3 & 1/4 removed
first stage: NMLLS receipt	.696***	.688***	.664***
standard error	.036	.037	.038
credits earned in 2nd semester	-.165	-.264	-.327
standard error	.265	.288	.448
$N_W^- \| N_W^+$	405 | 799	405 | 799	238 | 482
h	.496	.476	.356
credits earned after two years	2.222	2.033	3.233
standard error	1.862	1.811	3.897
$N_W^- \| N_W^+$	416 | 809	412 | 805	229 | 463
h	.468	.446	.350
credits earned after three years	4.512*	4.882**	5.707
standard error	2.606	2.485	5.929
$N_W^- \| N_W^+$	412 | 805	467 | 945	245 | 487
h	.442	.508	.354
credits earned after four years	8.669*	8.826**	8.907
standard error	4.617	4.381	6.536
$N_W^- \| N_W^+$	416 | 809	467 | 944	269 | 544
h	.460	.510	.387
credits earned after five years	7.645	7.946	7.408
standard error	5.637	5.412	5.482
$N_W^- \| N_W^+$	416 | 811	416 | 811	330 | 685
h	.485	.479	.464
credits earned after six years	9.659*	8.781	9.214
standard error	5.564	5.534	6.213
$N_W^- \| N_W^+$	416 | 811	469 | 951	326 | 681
h	.478	.542	.448
total observations	3,495	3,495	2,653

Note: estimates are based on fuzzy regression discontinuity models using the 2.5 GPA cutoff during the student's first semester (i.e., qualifying semester). A triangular kernel is used to construct local polynomial estimators. Common mean squared error-optimal bandwidths, h, determine the neighborhood of the GPA cutoff examined. Local linear regression is used to construct point estimators. Robust standard errors are clustered at the cohort-level. First-stage results are from models using credits earned during in the second semester. *, **, and *** signify statistical significance at the 10, five, and one percent-levels, respectively.

Table 5. FRD LATE, course withdrawals, accounting for heaping in qualifying GPA

	(1)	(2)	(3)
	Baseline	Flexible controls	Multiples of 1/3 & 1/4 removed
first stage: NMLLS receipt	.682***	.680***	.633***
standard error	.039	.040	.071
credits withdrawn in 2nd semester	.581	.398	-.277
standard error	.550	.421	.877
$N_W^- \mid N_W^+$	315 | 587	315 | 587	129 | 239
h	.348	.336	.186
credits withdrawn after year 2	.790	.360	.861
standard error	1.305	1.147	1.589
$N_W^- \mid N_W^+$	358 | 680	467 | 948	245 | 487
h	.418	.529	.351
credits withdrawn after year 3	2.403	2.101	2.058
standard error	2.072	1.852	2.305
$N_W^- \mid N_W^+$	355 | 668	412 | 805	229 | 463
h	.386	.520	.332
credits withdrawn after year 4	3.413	2.905	1.539
standard error	2.867	2.243	3.867
$N_W^- \mid N_W^+$	273 | 504	315 | 587	168 | 310
h	.299	.331	.252
credits withdrawn after year 5	3.386	3.147	1.517
standard error	2.739	2.350	3.795
$N_W^- \mid N_W^+$	287 | 538	315 | 587	168 | 310
h	.325	.344	.259
credits withdrawn after year 6	3.685	2.504	1.198
standard error	3.784	3.412	5.134
$N_W^- \mid N_W^+$	270 | 483	270 | 483	162 | 286
h	.288	.288	.219
total observations	3,495	3,495	2,653

Note: estimates are based on fuzzy regression discontinuity models using the 2.5 GPA cutoff during the student's first semester (i.e., qualifying semester). A triangular kernel is used to construct local polynomial estimators. Common mean squared error-optimal bandwidths, h, determine the neighborhood of the GPA cutoff examined. Local linear regression is used to construct point estimators. Robust standard errors are clustered at the cohort-level. First-stage results are from models using credits withdrawn during the first year. *, **, and *** signify statistical significance at the 10, five, and one percent-levels, respectively.

Table 6. Bounding exercise, degree completion

	(1)	(2)	(3)
	Main results	Excluding suspected manipulators, $2.5 \leq GPA \leq 2.75$	Excluding suspected manipulators, $2.5 \leq GPA \leq 3$
first stage: NMLLS receipt	.682***	.546***	.560***
standard error	.030	.029	.027
degree within four years	.108**	.172***	.153**
standard error	.046	.062	.064
$N_W^- \| N_W^+$	272 | 556	270 | 470	308 | 471
h	.413	.396	.421
degree in 4.5 years	.142*	.136*	.105
standard error	.075	.073	.077
$N_W^- \| N_W^+$	330 | 687	332 | 619	330 | 515
h	.482	.547	.547
degree within five years	.082	-.054	-.061
standard error	.127	.176	.187
$N_W^- \| N_W^+$	168 | 310	163 | 216	162 | 213
h	.246	.228	.211
degree within six years	.017	-.075	-.054
standard error	.114	.210	.155
$N_W^- \| N_W^+$	169 | 313	159 | 203	168 | 235
h	.264	.208	.246
total observations	2,653	2,578	2,481

Note: both specifications exclude students with heaped qualifying semester GPAs. Suspected manipulators earned the NMLLS with qualifying semester GPAs between 2.5 and 2.75 (column 2) or 2.5 and 3 (column 3) while either 1) registering for the minimum number of credits for scholarship eligibility 2) or dropping at least one course during the qualifying semester. Robust standard errors are clustered at the cohort-level. First-stage results are from models of four-year completion rates. *, **, and *** signify statistical significance at the 10, five, and one percent-levels, respectively.

Table 7. Bounding exercise, credits earned

	(1)	(2)	(3)
	Main results	Excluding suspected manipulators, $2.5 \leq$ GPA ≤ 2.75	Excluding suspected manipulators, $2.5 \leq$ GPA ≤ 3
first stage: NMLLS receipt	.664***	.541***	.478***
standard error	.038	.030	.059
credits earned in 2nd semester	-.327	-.014	.716
standard error	.448	.597	.798
$N_W^- \| N_W^+$	238 | 482	262 | 467	162 | 233
h	.356	.397	.249
credits earned after two years	3.233	10.404	9.339
standard error	3.897	6.355	6.100
$N_W^- \| N_W^+$	229 | 463	184 | 244	192 | 301
h	.350	.290	.306
credits earned after three years	5.707	16.240*	14.053*
standard error	5.929	9.081	8.550
$N_W^- \| N_W^+$	245 | 487	184 | 284	200 | 330
h	.354	.285	.312
credits earned after four years	8.907	18.576**	16.160*
standard error	6.536	9.491	8.430
$N_W^- \| N_W^+$	269 | 544	229 | 388	269 | 420
h	.387	.335	.386
credits earned after five years	7.408	15.664**	13.111**
standard error	5.482	6.985	8.489
$N_W^- \| N_W^+$	330 | 685	330 | 618	254 | 382
h	.464	.534	.367
credits earned after six years	9.214	20.951**	17.924**
standard error	6.213	8.693	7.450
$N_W^- \| N_W^+$	326 | 681	269 | 469	200 | 330
h	.448	.389	.313
total observations	2,653	2,578	2,481

Note: both specifications exclude students with heaped qualifying semester GPAs. Suspected manipulators earned the NMLLS with qualifying semester GPAs between 2.5 and 2.75 (column 2) or 2.5 and 3 (column 3) while either 1) registering for the minimum number of credits for scholarship eligibility 2) or dropping at least one course during the qualifying semester. Robust standard errors are clustered at the cohort-level. First-stage results are from models of second semester credits earned. *, **, and *** signify statistical significance at the 10, five, and one percent-levels, respectively.

Table 8. Bounding exercise, course withdrawals

	(1)	(2)	(3)
	Main results	Excluding suspected manipulators, 2.5 ≤ GPA ≤ 2.75	Excluding suspected manipulators, 2.5 ≤ GPA ≤ 3
first stage: NMLLS receipt	.633***	.463***	.459***
standard error	.071	.070	.072
credits withdrawn in 2nd semester	-.277	-.357	-.438
standard error	.877	1.045	1.086
$N_W^- \| N_W^+$	129 | 239	163 | 216	162 | 213
h	.186	.221	.214
credits withdrawn after year 2	.861	1.028	1.337
standard error	1.589	2.901	2.519
$N_W^- \| N_W^+$	245 | 487	169 | 238	187 | 301
h	.351	.262	.292
credits withdrawn after year 3	2.058	3.643	3.768
standard error	2.305	4.076	3.821
$N_W^- \| N_W^+$	229 | 463	192 | 326	200 | 330
h	.332	.303	.314
credits withdrawn after year 4	1.539	3.515	4.132
standard error	3.867	6.745	5.523
$N_W^- \| N_W^+$	168 | 310	216 | 216	168 | 235
h	.252	.216	.255
credits withdrawn after year 5	1.517	3.448	4.926
standard error	3.795	8.628	6.102
$N_W^- \| N_W^+$	168 | 310	129 | 166	166 | 232
h	.259	.176	.231
credits withdrawn after year 6	1.198	.910	2.585
standard error	5.134	11.906	10.077
$N_W^- \| N_W^+$	162 | 286	121 | 151	129 | 166
h	.219	.137	.188
total observations	2,653	2,578	2,481

Note: both specifications exclude students with heaped qualifying semester GPAs. Suspected manipulators earned the NMLLS with qualifying semester GPAs between 2.5 and 2.75 (column 2) or 2.5 and 3 (column 3) while either 1) registering for the minimum number of credits for scholarship eligibility 2) or dropping at least one course during the qualifying semester. Robust standard errors are clustered at the cohort-level. First-stage results are from models second semester course withdrawal. *, **, and *** signify statistical significance at the 10, five, and one percent-levels, respectively.

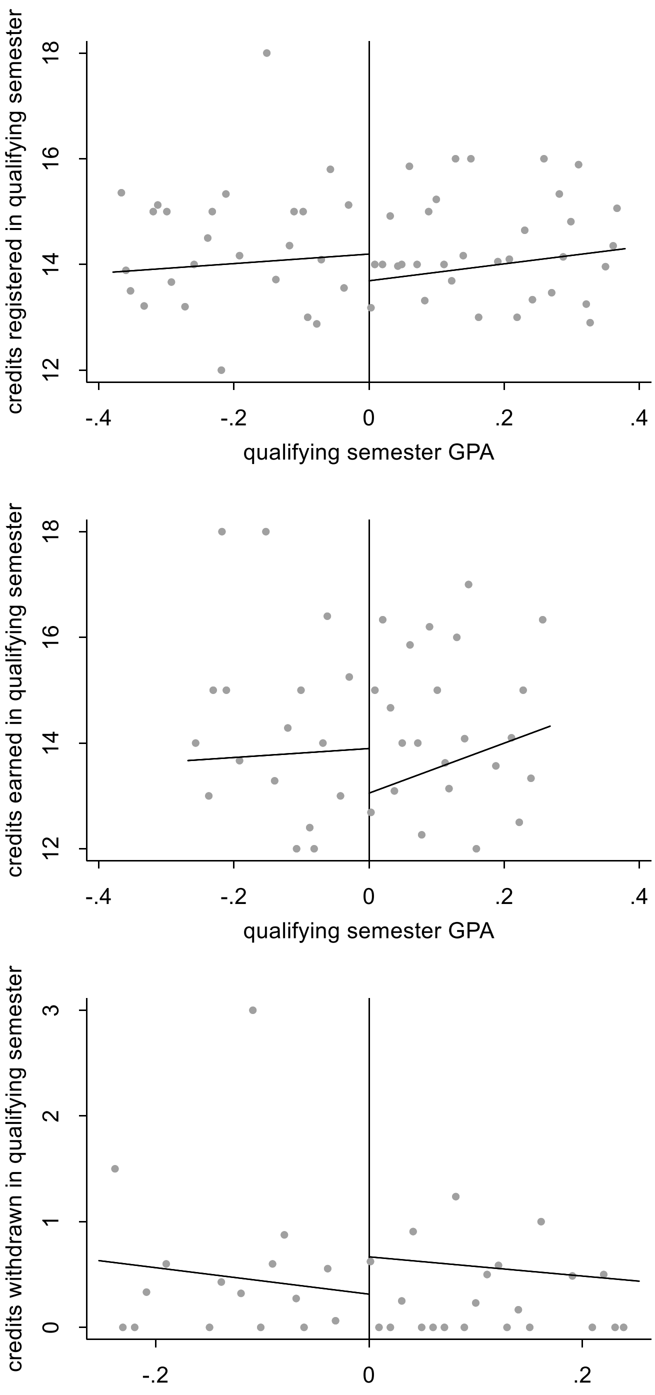
APPENDIX A: TESTS FOR CUTOFF MANIPULATION

Table A1. Testing for manipulation of the qualifying semester GPA cutoff for NMLLS eligibility, 1997 to 1999 cohorts

	Bandwidths		Effective obs.		Conv. test		Robust test	
$h_{-} \neq h_{+}$	left	right	left	right	T	p-value	T	p-value
$T_2(\hat{h}_1)$	.605	.552	504	951	1.124	.261	.110	.912
$T_3(\hat{h}_2)$	1.010	.890	697	1630	.574	.566	.193	.847
$T_4(\hat{h}_3)$	1.156	1.112	736	2043	.527	.599	.081	.935
$h_{-} = h_{+}$								
$T_2(\hat{h}_1)$	.501	.501	467	944	.748	.455	.296	.767
$T_3(\hat{h}_2)$	.890	.890	643	1630	.723	.470	.141	.888
$T_4(\hat{h}_3)$	1.112	1.112	730	2043	.555	.579	.046	.963

Note: results from manipulation tests following McCrary (2008) and Cattaneo, Jansson, and Ma (2016) examining 1997 to 1999 cohorts at UNM. $T_p(h)$ is the manipulation test statistic using the *p*-th order density estimators with bandwidth *h*. $\hat{h}_p$ denotes the MSE-optimal bandwidths for the *p*-th order density estimator. A triangular kernel is used to construct local polynomial estimators. Tests are performed with identical and different data-driven bandwidths. Conventional and robust test statistics examine the null hypothesis of continuity in the qualifying semester GPA around the NMLLS eligibility cutoff.

APPENDIX B: VISUAL PRESENTATION OF PLACEBO TREATMENT EFFECTS



Note: observations nonrandomly heaped at multiples of one-third and one-fourth have been removed. A linear fit has been added below and above the cutoff centered at zero. The triangular kernel is used to construct the global polynomial estimators.

Figure B1. Visual check of threshold manipulation in qualifying semester, 1997-1999

APPENDIX C: FALSIFICATION TESTING USING ALTERNATIVE CUTOFFS

Table C1. Falsification testing using alternative cutoffs, college completion

	(1)	(2)
	c = 2.3	c = 2.7
first stage: NMLLS receipt	-.180***	-.071
standard error	.041	.057
degree within four years	.441	.947
standard error	1.004	.646
$N_W^- \| N_W^+$	434 | 1020	473 | 619
h	.682	.362
degree in 4.5 years	-.038	-2.842
standard error	.781	8.070
$N_W^- \| N_W^+$	336 | 713	344 | 337
h	.497	.212
degree within five years	.131	.568
standard error	.540	.733
$N_W^- \| N_W^+$	258 | 428	697 | 1087
h	.334	.580
degree within six years	-.114	4.601
standard error	.356	8.695
$N_W^- \| N_W^+$	205 | 319	423 | 468
h	.255	.295

Note: estimates are based on fuzzy regression discontinuity models using the 2.5 GPA cutoff during the student's qualifying semester. A triangular kernel is used to construct local polynomial estimators. Common mean squared error-optimal bandwidths, *h*, determine the neighborhood of the GPA cutoff examined. Local linear regression is used to construct point estimators. N_W^- are effective observations below the cutoff and N_W^+ are those above. Robust standard errors are clustered at the cohort-level. *, **, and *** signify statistical significance at the 10, five, and one percent-levels, respectively.

Table C2. Falsification testing using alternative cutoffs, credit accumulation

	(1)	(2)
	c = 2.3	c = 2.7
first stage: NMLLS receipt	-.141***	-.132**
standard error	.049	.057
credits earned in 2nd semester	1.719	.734
standard error	6.630	4.855
$N_W^-\|N_W^+$	205 | 316	546 | 797
h	.268	.445
credits earned after two years	5.070	12.072
standard error	27.165	17.333
$N_W^-\|N_W^+$	205 | 319	597 | 828
h	.257	.455
credits earned after three years	10.549	167.92
standard error	33.458	177.26
$N_W^-\|N_W^+$	207 | 371	437 | 601
h	.280	.310
credits earned after four years	-5.555	252.44
standard error	49.673	268.90
$N_W^-\|N_W^+$	205 | 319	437 | 601
h	.256	.314
credits earned after five years	-13.715	143.89
standard error	76.283	92.325
$N_W^-\|N_W^+$	197 | 286	522 | 653
h	.239	.377
credits earned after six years	-10.585	194.84
standard error	46.154	140.06
$N_W^-\|N_W^+$	207 | 326	472 | 608
h	.268	.349

Note: estimates are based on fuzzy regression discontinuity models using the 2.5 GPA cutoff during the student's qualifying semester. A triangular kernel is used to construct local polynomial estimators. Common mean squared error-optimal bandwidths, h, determine the neighborhood of the GPA cutoff examined. Local linear regression is used to construct point estimators. N_W^- are effective observations below the cutoff and N_W^+ are those above. Robust standard errors are clustered at the cohort-level. *, **, and *** signify statistical significance at the 10, five, and one percent-levels, respectively.

Table C3. Falsification testing using alternative cutoffs, credit withdrawal

	(1)	(2)
	c = 2.3	c = 2.7
first stage: NMLLS receipt	-.187***	-.079
standard error	.040	.058
credits withdrawn in 2nd semester	.972	1.883
standard error	4.052	2.350
$N_W^- \| N_W^+$	258 | 428	473 | 619
h	.335	.360
credits withdrawn after year 2	-1.609	8.958
standard error	10.467	11.134
$N_W^- \| N_W^+$	258 | 428	472 | 608
h	.333	.341
credits withdrawn after year 3	-1.170	4.137
standard error	15.259	5.875
$N_W^- \| N_W^+$	258 | 428	630 | 954
h	.339	.528
credits withdrawn after year 4	1.861	3.384
standard error	17.979	9.965
$N_W^- \| N_W^+$	266 | 515	749 | 1223
h	.376	.632
credits withdrawn after year 5	2.998	8.994
standard error	18.983	10.165
$N_W^- \| N_W^+$	258 | 426	555 | 753
h	.322	.412
credits withdrawn after year 6	4.601	6.891
standard error	21.903	10.494
$N_W^- \| N_W^+$	258 | 426	597 | 828
h	.325	.457

Note: estimates are based on fuzzy regression discontinuity models using the 2.5 GPA cutoff during the student's qualifying semester. A triangular kernel is used to construct local polynomial estimators. Common mean squared error-optimal bandwidths, h, determine the neighborhood of the GPA cutoff examined. Local linear regression is used to construct point estimators. N_W^- are effective observations below the cutoff and N_W^+ are those above. Robust standard errors are clustered at the cohort-level. *, **, and *** signify statistical significance at the 10, five, and one percent-levels, respectively.

APPENDIX D: PREFERRED SPECIFICATION USING ALTERNATIVE BANDWIDTHS

Table D1. Sensitivity of results to various bandwidths, college completion

	(1)	(2)	(3)	(4)
	One common MSE-optimal	Two separate MSE-optimal	One common CER-optimal	Two separate CER-optimal
first stage: NMLLS eligibility	.682***	.679***	.678***	.674***
standard error	.030	.038	.032	.042
degree within four years	.108**	.104**	.111**	.100**
standard error	.046	.049	.044	.047
$N_W^- \| N_W^+$	272 | 556	270 | 463	268 | 536	254 | 401
$h^- \| h^+$	.413 | .413	.494 | .573	.377 | .377	.366 | .306
degree in 4.5 years	.142*	.145*	.132*	.142*
standard error	.075	.078	.076	.078
$N_W^- \| N_W^+$	330 | 687	330 | 639	326 | 681	330 | 544
$h^- \| h^+$	.482 | .482	.536 | .421	.440 | .440	.490 | .385
degree within five years	.082	.087	.079	.100
standard error	.127	.092	.131	.104
$N_W^- \| N_W^+$	168 | 310	162 | 1074	163 | 290	159 | 977
$h^- \| h^+$	.246 | .246	.215 | .734	.225 | .225	.196 | .671
degree within six years	.017	.068	.021	.076
standard error	.114	.131	.121	.150
$N_W^- \| N_W^+$	169 | 313	159 | 685	168 | 310	129 | 556
$h^- \| h^+$	.264 | .264	.207 | .452	.242 | .242	.189 | .413

Note: all models exclude multiples of one-third and one-fourth. Estimates are based on fuzzy regression discontinuity models using the 2.5 GPA cutoff during the student's first semester (i.e., qualifying semester). A triangular kernel is used to construct local polynomial estimators. Common mean squared error-optimal bandwidths, *h*, determine the neighborhood of the GPA cutoff examined. Local linear regression is used to construct point estimators. Robust standard errors are clustered at the cohort-level. First-stage results are from models using credits withdrawn during the first year. Models employ local linear regression and use mean squared error (MSE)-optimal and coverage error ration (CER)-optimal bandwidths. *, **, and *** signify statistical significance at the 10, five, and one percent-levels, respectively.

Table D2. Sensitivity of results to various bandwidths, credit accumulation

	(1)	(2)	(3)	(4)
	One common MSE-optimal	Two separate MSE-optimal	One common CER-optimal	Two separate CER-optimal
first stage: NMLLS eligibility	.664***	.657***	.661***	.654***
standard error	.038	.040	.043	.041
credits after second semester	-.327	.088	-.086	.203
standard error	.448	.494	.471	.580
$N_W^-\|N_W^+$	238 | 482	195 | 678	195 | 409	181 | 632
h^- | h^+	.356 | .356	.321 | .464	.326 | .326	.293 | .425
credits after two years	3.233	5.028	4.148	5.768
standard error	3.897	4.269	4.002	4.493
$N_W^-\|N_W^+$	229 | 463	192 | 842	200 | 410	184 | 739
h^- | h^+	.350 | .350	.301 | .625	.320 | .320	.275 | .571
credits after three years	5.707	8.509	6.498	9.915
standard error	5.929	6.664	6.158	7.027
$N_W^-\|N_W^+$	245 | 487	192 | 826	201 | 414	184 | 693
h^- | h^+	.354 | .354	.303 | .587	.324 | .324	.277 | .537
credits after four years	8.907	10.423	8.941	11.882*
standard error	6.536	7.188	6.817	7.093
$N_W^-\|N_W^+$	269 | 544	245 | 694	245 | 487	201 | 687
h^- | h^+	.387 | .387	.356 | .544	.354 | .354	.326 | .497
credits after five years	7.408	6.745	6.876	6.729
standard error	5.482	5.268	5.875	5.240
$N_W^-\|N_W^+$	330 | 685	367 | 826	308 | 639	332 | 693
h^- | h^+	.464 | .464	.598 | .581	.424 | .424	.547 | .531
credits after six years	9.214	9.151	8.924	8.682
standard error	6.213	6.340	6.544	5.701
$N_W^-\|N_W^+$	326 | 681	269 | 544	272 | 550	229 | 487
h^- | h^+	.448 | .448	.380 | .385	.410 | .410	.348 | .352

Note: all models exclude multiples of one-third and one-fourth. Estimates are based on fuzzy regression discontinuity models using the 2.5 GPA cutoff during the student's first semester (i.e., qualifying semester). A triangular kernel is used to construct local polynomial estimators. Common mean squared error-optimal bandwidths, *h*, determine the neighborhood of the GPA cutoff examined. Local linear regression is used to construct point estimators. Robust standard errors are clustered at the cohort-level. First-stage results are from models using credits withdrawn during the first year. Models employ local linear regression and use mean squared error (MSE)-optimal and coverage error ration (CER)-optimal bandwidths. *, **, and *** signify statistical significance at the 10, five, and one percent-levels, respectively.

Table D3. Sensitivity of results to various bandwidths, credit withdrawal

	(1)	(2)	(3)	(4)
	One common MSE-optimal	Two separate MSE-optimal	One common CER-optimal	Two separate CER-optimal
first stage: NMLLS eligibility	.633***	.637***	.633***	.639***
standard error	.071	.062	.070	.057
credits withdrawn after second semester	-.277	-.222	-.271	-.191
standard error	.877	.549	.880	.556
$N_W^- \| N_W^+$	129 | 239	129 | 545	129 | 239	128 | 487
$h^- \| h^+$	.186 | .186	.155 | .392	.170 | .170	.142 | .359
credits withdrawn after two years	.861	.812	.670	.768
standard error	1.589	2.290	1.758	2.416
$N_W^- \| N_W^+$	245 | 487	168 | 694	201 | 414	166 | 687
$h^- \| h^+$	.351 | .351	.260 | .550	.321 | .321	.238 | .503
credits withdrawn after three years	2.058	2.102	1.717	2.009
standard error	2.305	3.142	2.714	3.416
$N_W^- \| N_W^+$	229 | 463	184 | 739	192 | 401	168 | 691
$h^- \| h^+$	.332 | .332	.277 | .575	.304 | .304	.253 | .526
credits withdrawn after four years	1.539	.750	1.255	.382
standard error	3.867	5.847	4.207	6.146
$N_W^- \| N_W^+$	168 | 310	159 | 694	166 | 307	129 | 687
$h^- \| h^+$	.252 | .252	.193 | .556	.230 | .230	.176 | .509
credits withdrawn after five years	1.517	.185	1.280	-.133
standard error	3.795	5.968	4.096	6.053
$N_W^- \| N_W^+$	168 | 310	129 | 687	166 | 307	129 | 687
$h^- \| h^+$	.259 | .259	.164 | .518	.237 | .237	.150 | .474
credits withdrawn after six years	1.198	-.920	.503	-1.796
standard error	5.134	7.270	5.703	7.202
$N_W^- \| N_W^+$	162 | 286	121 | 687	159 | 276	121 | 681
$h^- \| h^+$	.219 | .219	.137 | .485	.201 | .201	.125 | .443

Note: all models exclude multiples of one-third and one-fourth. Estimates are based on fuzzy regression discontinuity models using the 2.5 GPA cutoff during the student's first semester (i.e., qualifying semester). A triangular kernel is used to construct local polynomial estimators. Common mean squared error-optimal bandwidths, h, determine the neighborhood of the GPA cutoff examined. Local linear regression is used to construct point estimators. Robust standard errors are clustered at the cohort-level. First-stage results are from models using credits withdrawn during the first year. Models employ local linear regression and use mean squared error (MSE)-optimal and coverage error ration (CER)-optimal bandwidths. *, **, and *** signify statistical significance at the 10, five, and one percent-levels, respectively.

APPENDIX E: TESTS FOR EXTERNAL VALIDITY

Table E1. Tests of external validity using treatment effect derivatives and complier probability derivatives

	(1)	(2)	(3)
	LATE	TED	CPD
degree within four years	.080* (.045)	.075 (.249)	.003 (.171)
degree within 4.5 years	.129** (.058)	.365 (.263)	.057 (.141)
degree within five years	.078 (.071)	.242 (.357)	.033 (.154)
degree within six years	.028 (.076)	.357 (.347)	.055 (.143)
credits earned in second semester	-.341 (.472)	1.777 (3.416)	-.078 (.253)
credits earned after two years	1.514 (2.151)	6.697 (13.509)	-.054 (.204)
credits earned after three years	3.046 (3.499)	11.011 (19.242)	.005 (.170)
credits earned after four years	6.198 (5.474)	26.119 (32.793)	-.035 (.191)
credits earned after five years	4.675 (6.381)	22.172 (35.738)	-.002 (.174)
credits earned after six years	6.266 (6.971)	14.386 (38.942)	-.001 (.174)
total observations			3,495

Note: models exclude students with qualifying semester GPA heaped at multiples of one-third and one-fourth. Estimates are based on fuzzy regression discontinuity models using the 2.5 GPA cutoff during the student's first semester (i.e., qualifying semester) and are not adjusted for covariates. TED is the treatment effect derivative and CPD is the complier probability derivative. A triangular kernel is used to construct local polynomial estimators. Common mean squared error-optimal bandwidths, *h*, determine the neighborhood of the GPA cutoff examined. Local linear regression is used to construct point estimators. The user-written Stata command *ted*, by Giovanni Cerulli, produces the results above. Robust standard errors are reported in parentheses. *, **, and *** signify statistical significance at the 10, five, and one percent-levels, respectively.

Table E1. Tests of external validity using treatment effect derivatives and complier probability derivatives (continued)

	(1)	(2)	(3)
	LATE	TED	CPD
credits withdrawn during second semester	.480 (.386)	2.187 (2.806)	-.094 (.249)
credits withdrawn during first two years	.802 (.687)	.357 (4.086)	-.032 (.189)
credits withdrawn during first three years	1.728* (.912)	1.783 (4.588)	.033 (.155)
credits withdrawn during first four years	3.317*** (1.144)	4.756 (7.034)	-.046 (.198)
credits withdrawn during first five years	3.121** (1.443)	4.623 (9.469)	-.064 (.215)
credits withdrawn during first six years	3.596** (1.683)	7.770 (11.490)	-.074 (.226)
total observations			3,495

Note: models exclude students with qualifying semester GPA heaped at multiples of one-third and one-fourth. Estimates are based on fuzzy regression discontinuity models using the 2.5 GPA cutoff during the student's first semester (i.e., qualifying semester) and are not adjusted for covariates. TED is the treatment effect derivative and CPD is the complier probability derivative. A triangular kernel is used to construct local polynomial estimators. Common mean squared error-optimal bandwidths, *h*, determine the neighborhood of the GPA cutoff examined. Local linear regression is used to construct point estimators. The user-written Stata command *ted*, by Giovanni Cerulli, produces the results above. Robust standard errors are reported in parentheses. *, **, and *** signify statistical significance at the 10, five, and one percent-levels, respectively.

APPENDIX F: DESCRIPTIVE STATISTICS BY SUBSAMPLE

Table F1. Descriptive statistics, 1997 to 1999 cohorts, by subsample

	(1)	(2)	(3)
	Full sample	(1) Less heaped	(2) Less suspected manipulators, 2.5 ≤ GPA ≤ 2.75
qualifying semester credits:			
registered	14.192 (1.674)	14.267 (1.642)	14.325 (1.632)
earned	13.771 (1.763)	13.944 (1.735)	13.998 (1.729)
withdrawn	.570 (1.215)	.516 (1.165)	.504 (1.156)
bachelor's degree within (years):			
4	.164	.173	.176
4.5	.288	.299	.301
5	.460	.469	.471
6	.559	.571	.574
credits earned within (years):			
1	27.049 (4.122)	27.391 (3.970)	27.478 (3.956)
2	50.069 (13.799)	50.704 (13.577)	50.889 (13.569)
3	70.403 (24.582)	71.334 (24.287)	71.668 (24.261)
4	89.592 (36.051)	90.890 (35.588)	91.364 (31.512)
5	101.312 (42.211)	102.601 (41.652)	103.090 (41.544)
6	106.507 (44.797)	107.849 (44.266)	108.337 (44.165)
credits withdrawn within (years):			
1	1.857 (2.622)	1.745 (2.508)	1.721 (2.493)
2	3.757 (4.253)	3.597 (4.165)	3.570 (4.157)
3	5.387 (6.626)	5.194 (5.544)	5.173 (5.555)
4	7.080 (7.033)	6.885 (6.947)	6.868 (6.975)
5	8.226 (8.154)	8.051 (8.120)	8.045 (8.158)
6	9.761 (9.534)	9.525 (9.500)	9.508 (9.544)
female	.553	.562	.561
high school GPA	3.435 (.450)	3.448 (.448)	3.455 (.447)
composite ACT	24.007 (3.240)	24.120 (3.275)	24.164 (3.282)
required remedial coursework	.066	.062	.061
family income < $40,000	.176	.171	.170
family income < $20,000	.078	.077	.077
white	.576	.578	.581
Hispanic	.319	.311	.309
Asian	.040	.044	.043
American Indian	.034	.037	.037
black	.017	.017	.018
declined to state race-ethnicity	.014	.013	.012
total observations	3,495	2,653	2,578

Source: Freshmen Tracking System, Office of Institutional Analytics, University of New Mexico. Standard deviations are in parentheses.