The Impact of Merit Aid on College Choice and Degree Attainment:

Reexamining Florida's Bright Futures Program

Forthcoming, *Educational Evaluation and Policy Analysis*

Oded Gurantz, Assistant Professor
Truman School of Public Affairs, University of Missouri
216A Middlebush Hall, Columbia, MO 65201
gurantzo@missouri.edu
Research Areas: Higher Education and Economics

Taylor K. Odle, University of Pennsylvania
Higher Education Division, University of Pennsylvania
208 South 37th Street, Stiteler Hall, Suite 200, Philadelphia, PA 19104
todle@upenn.edu
Research Areas: Student Financial Aid, College Access and Success, Program Evaluation

Abstract: We replicate and extend prior work on Florida's Bright Futures merit aid scholarship to consider its effect on college enrollment and degree completion. We estimate causal impacts using a regression discontinuity design to exploit SAT thresholds that strongly determine eligibility. We find no positive impacts on attendance or attainment, and instrumental variable results generally reject estimates as small as 1-2 percentage points. Across subgroups, we find that eligibility slightly reduces six-year associate degree attainment for lower-SES students, and may induce small enrollment shifts among Hispanic and White students toward four-year colleges. Our findings of these minimal-at-best impacts contrast those of prior works, attributable in part to methodological improvements and more robust data, and further underscore the importance of study replication. (*JEL*: H75, I21, I22, I23, I28)

Keywords: Bright Futures; financial aid; merit aid; regression discontinuity; replication

This research was also supported in part through funding from the Institute of Education Sciences, U.S. Department of Education, under Grant R305B200035 to the University of Pennsylvania. Any views or opinions expressed are those of the authors alone and do not represent the views or opinions of the Institute or the U.S. Department of Education.

Introduction

Financial aid has a causal impact on a variety of postsecondary outcomes, including college enrollment, degree completion, and labor-market earnings (Bettinger, Gurantz, Kawano, Sacerdote, & Stevens, 2019; Denning, Marx, & Turner, 2019; Nguyen, Kramer, & Evans, 2019). The magnitude and direction of these effects largely vary by program design and delivery, with merit-based grant programs typically producing weaker effects on enrollment and completion than need-based programs (Domina, 2014; Dynarski & Scott-Clayton, 2013; Herbaut & Geven, 2020; Page & Scott-Clayton, 2016). Within this body of evidence, a large portion has focused on the causal impacts of state merit aid. This interest is due in part to programs’ large public expenditures but also because of many states’ large-scale shifts from need to merit programs in the 1990s and early 2000s (Dynarski, 2004). As of 2018, 23 states operated at least one merit aid program and spent nearly $2.4 billion on merit awards (National Association of State Student Grant and Aid Programs, 2018). Among those 2018 expenditures, nearly $383 million came from Florida’s Bright Futures (BF) program (Florida Department of Education, 2018).

Though the preponderance of empirical evidence promotes need-based programs as producing stronger positive impacts than their merit-based counterparts (see Herbaut and Geven (2020) and Nguyen et al. (2019) for reviews), debate still exists among state policymakers regarding the costs and benefits of each form of aid, as many ask: For whom does aid “work,” and on what margins? While aid programs generally produce positive effects, some have shown zero or even negative impacts, and repeated evaluations of even the same programs have produced conflicting results (e.g., Bettinger et al. (2019) and Kane (2003); Dynarski (2008) and Sjoquist and Winters (2012)). In this light, study replication and extension help detect heterogeneous effects across program types and student groups while also adding new outcome measures and strengthening or correcting prior inferences (Granger & Maynard, 2015; Hedges, 2018). In the policy arena, replications and extensions are particularly important to building a robust body of evidence for policymakers to consider the design and implementation of aid programs, and to weigh investments of scarce state resources (Wong, Anglin, & Steiner, 2021).

This study reexamines the effects of Florida's Bright Futures merit aid program, previously studied by Zhang, Hu, Sun, and Pu (2016). We similarly estimate impacts using a regression discontinuity (RD) design that leverages eligibility criteria requiring students to, in part, score above specific SAT benchmarks. We expand Zhang et al. (2016) and related prior works by: observing three times as many high school cohorts, including students in private high schools; using new, longer-term outcomes on degree attainment up to six years after high school graduation; and relying on national-level postsecondary enrollment data, rather than state-level data. Most importantly, we make methodological improvements to alleviate the threats to internal validity present in these prior studies, discussed in detail below.

In contrast to prior work, we generally find no impacts of Bright Futures eligibility on postsecondary enrollment or completion outcomes. Using students' SAT scores as a proxy for scholarship eligibility, we show the offer of an award does not generally result in statistically significant changes in (1) whether a student attends college overall; (2) the sector, level, or geographic location of postsecondary enrollment; or (3) associate or bachelor's degree attainment. One consistent finding, however, is that students near the program's lower SAT threshold, which would provide eligibility for a partial rather than full tuition scholarship, are roughly 1 to 2 percentage points less likely to earn an associate degree from an in-state institution, with this effect concentrated among students living in zip codes with lower socioeconomic status. Although a negative outcome on its face, this may be accompanied by a small positive impact on bachelor's degree completion, though this is effect is statistically insignificant and slightly smaller in magnitude. These results provide further evidence that, although some merit aid programs have been shown to increase degree attainment, others may have negative, null, or smaller effects, particularly when compared to their need-based counterparts (Cohodes & Goodman, 2014; Fitzpatrick & Jones, 2016; Herbaut & Geven, 2020; Nguyen et al., 2019; Sjoquist & Winters, 2012). When considering possible effects across racial groups, we do find suggestive evidence that BF may have induced small enrollment shifts for Hispanic and White students away from two-year colleges and toward in-state, four-year and private institutions, respectively. When judged against enrollment and completion goals, however, these minimal-at-best impacts suggest the costs of BF may not outweigh its short-term benefits, though impacts on long-run outcomes (e.g., lower loan burdens or improved job-market outcomes) are possible.

Our study's findings have important implications for research and policy. In addition to our empirical contribution to the ongoing need versus merit aid debate, our replication and extension serves to correct prior inferences on Bright Futures and shows that some prevailing findings on aid programs may warrant revisiting in the presence of microdata, newer methods, larger sample sizes, and longer observation periods. Indeed, reviews and replications of prior studies into merit aid suggest prior findings may have been overstated and that these programs have minimal if any impacts on enrollment and completion (Fitzpatrick & Jones, 2016; Sjoquist & Winters, 2012). Further, one of our study's methodological improvements serves as a strong example of the importance of proper implementation of regression discontinuity designs that rely upon test scores given endogenous retaking behaviors at ex ante known eligibility thresholds. Our findings also stand to inform public policy in Florida and states with similar merit programs. Florida has consistently decreased the availability of BF over the last decade alone by raising eligibility thresholds each year from 2011-12 through 2013-14, in 2016-17, and again in 2019. These actions cut the proportion of all high school students in the state who would be eligible for a BF award nearly in half: from 38% in 2010-11 to 20% in 2016-17 (Florida Department of Education, 2019). Stronger evidence on the effectiveness of BF provides policymakers with better information on whether their program "works," and if such changes are supported. In all, our paper underscores the need for researchers and policymakers alike to take care when considering existing evidence on program efficacy, suggesting that they should draw implications from the entirety of available evidence rather than individual studies. Knowledge of this nature is particularly important as policy leaders continue to make tradeoffs between need- and merit-based programs, and our findings support a growing body of evidence suggesting that merit programs may do little to affect access and success outcomes given their allocation of funds to students who are already likely to enroll and succeed. This counters the traditional logic of financial aid, which is to reduce constraints or provide incentives for students who would otherwise have not enrolled or succeeded (Dynarski, 2004; Long & Riley, 2007).

The remainder of the paper is organized as follows. First, we provide an overview of Florida's Bright Futures program and explicitly outline the study's research questions. Next, we review relevant literature on merit aid programs to situate our study and discuss prior works on BF itself. We then describe our data and empirical strategy, noting how both make considerable improvements upon existing evidence. We then present our primary results, robustness checks, and findings from a heterogeneity analysis across students' socioeconomic status and ethnicity. Here, we also dedicate considerable effort in working to replicate prior studies as close as possible given our more robust data, and to carefully discuss why our inferences substantively vary from existing works. Finally, we conclude with a discussion of our findings, their contributions to extant literature, and their implications for policy.

Background

Florida's Bright Futures scholarship is one of the nation's largest merit programs, expending over $423 million in 2010-11 (within our study window) to more than 179,000 students (Florida Department of Education, 2011). The scholarship is available to recent high school graduates who enroll in a degree program at an in-state public or private institution for at least six non-remedial credit hours. From 2004 through 2011, students with an SAT score of 970 or greater were eligible for the Florida Medallion Scholars (FMS) award, covering 75% of tuition and fees; this threshold increased to 980 in 2012. From 2004 through 2012, students scoring 1270 or higher were eligible for the more generous Florida Academic Scholars (FAS) award, covering 100% of tuition and fees plus a $600 annual stipend for additional educational expenses.[1] BF award levels are not dependent upon the receipt of other aid (i.e., not a "last-dollar" scholarship), so awards can be combined with federal Pell grants and other aid. In practice, the average annual award for FMS and FAS students in 2010-11 was $2,124 and $3,250, respectively (Florida Department of Education, 2011). Thus, crossing the 970/980 SAT threshold, which shifts students from no award to FMS (i.e., from $0 to coverage for 75% of tuition and fees), appears to be a more substantial change in aid than crossing the 1270 threshold that shifts students from FMS to FAS (i.e., from 75% to 100%). For initial eligibility, students are also required to meet a weighted GPA threshold (3.00 for FMS and 3.50 for FAS) and complete community service hours (75 for FAS students in 2011, then 100 for FAS and 75 for FMS in 2012), which we are unable to observe. BF funding

[1] These SAT thresholds correspond to ACT scores of 28 for FAS and 20 for FMS (or 21 in 2012). Zhang et al. (2016) find that the SAT was more popular than the ACT in the mid-2000s, though both exams were popular.

lasts through 100% of a program of study (120 credit hours or 5 years), and awards at private institutions are based upon credit hour costs at a comparable public college.

In this study, we leverage the Bright Futures SAT eligibility criteria described above to estimate causal impacts of the scholarship program on students' college enrollment and degree completion outcomes. Specifically, we answer the following research questions:

1. Does eligibility for a Bright Futures award on the SAT margin affect college enrollment overall or by level (two or four year), sector (public or private), or location (in or out of state)?
2. Does eligibility for a Bright Futures award on the SAT margin affect degree attainment overall or by type (associate or bachelor's) or timeframe (four or six years)?

We further consider these questions across student subgroups, including ethnicity and a measure of socioeconomic status.

Literature Review

Financial aid is generally meant to influence students' college cost-benefit decisions. By reducing affordability constraints, aid may lead students considering postsecondary education to ultimately enroll and may similarly alter students' colleges choice by affecting where they enroll. Indeed, grant aid reduces the costs associated with an investment in human capital (i.e., postsecondary education), allowing students to expect larger net benefits (Becker, 1975). A robust body of evidence has documented the generally positive effect of aid on college-going (Dynarski & Scott-Clayton, 2013; Page & Scott-Clayton, 2016). While we know financial aid impacts access and attendance in these ways, it also affects outcomes like persistence, completion, and labor-market earnings (Bettinger et al., 2019; Denning et al., 2019; Nguyen et al., 2019). Evidence on these effects of grant aid focuses on aid from many sources, including states, offering various need- and merit-based awards (Bettinger et al., 2019; Dynarski, 2008; Sjoquist & Winters, 2012).

Recent meta-analytic evidence suggests need-based aid may produce stronger effects than merit aid, particularly for student persistence and completion. Across 43 studies, Nguyen et al. (2019) found that grant aid improved persistence and attainment by 1.5 to 2 percentage points overall but found effects for merit-only awards were consistently lower than need-only awards for year-to-year persistence and on-time completion (0-2.1 percentage points for merit awards compared to a stable 2.5 points for need-based awards). In a systematic review of quasi-experimental studies on financial aid programs, Herbaut and Geven (2020) similarly observed lower average effect sizes of merit aid programs compared to need-based counterparts for enrollment and completion outcomes. Despite the generally positive, albeit small effects in the merit aid literature, there is also rigorous evidence to suggest merit aid could have negative impacts on students' outcomes by way of enrollment diversion toward lower-quality or under-resourced institutions (e.g., Cohodes and Goodman (2014)).

State merit aid programs have many goals and rely upon multiple mechanisms to influence students' academic outcomes. Bettinger et al. (2019) describe these goals to include signaling academic performance standards, which may increase academic effort and high school completion; reducing liquidity constraints to propel postsecondary enrollment and attainment rates while simultaneously reducing hazards to completion (e.g., the need to work); and decreasing "brain drain" by increasing the likelihood that academically-talented students remain in-state for college and subsequent workforce participation. A large body of evidence has documented the generally successful record of merit aid in accomplishing these aims (e.g., Castleman and Long (2016); Cornwell, Mustard, and Sridhar (2006); Pallais (2009); Scott-Clayton (2011)). The notable exception is the post-baccalaureate or long-term retention of awardees, where effects remain null or very small for large state programs (Bettinger et al., 2019; Fitzpatrick & Jones, 2016; Sjoquist & Winters, 2014). Our study contributes to this body of evidence by replicating and extending prior works on one of the nation's largest merit programs and by examining its impacts across many of these primary aims: enrollment, out-migration for college, and attainment—all of which we further consider across students' ethnicity and socioeconomic status.

Summarizing studies that estimated impacts of state merit programs on college enrollment, Deming and Dynarski (2010) found that, on average, reducing college costs by $1,000 increases the likelihood of enrollment by 4 percentage points. These studies include estimates of many large state programs, including the BF-peer HOPE scholarships in Georgia and Tennessee (Dynarski & Scott-Clayton, 2013; Singell & Stone, 2002). Across several other programs in the southern United States, Dynarski (2003) estimated a 3.6 percentage point increase in the probability of attending college given an offer of aid. The Herbaut and Geven (2020) systematic review also considered the effects of merit programs on overall college enrollment of low-income student populations and noted six study effects ranging from -2.6 to +4.2 percentage points. On the college choice margin, Cohodes and Goodman (2014) found large effects of the merit-based Adams Scholarship in Massachusetts, suggesting it induced diversion of 6.9% of students at the eligibility threshold away from award-ineligible institutions toward eligible ones. Yet, even for the universal (not need- or merit-based) Knox Achieves program, Carruthers and Fox (2016) found diversion effects of 1 to 5 percentage points away from ineligible four-year institutions, and these effects were even larger for students with the strongest academic credentials. On the attainment front, the Nguyen et al. (2019) meta-analysis found merit aid programs increased the likelihood of on-time graduation by approximately 2.1 percentage points. Dynarski (2008) similarly found that merit programs in

Georgia and Arkansas increased attainment by 3 to 4 percentage points. While some studies show that merit aid has no impacts (e.g., Fitzpatrick & Jones, 2016; Sjoquist & Winters, 2012), the preponderance of existing work suggests aid does affect college enrollment and completion—typically providing impacts of 2-4 percentage points across studies on any given margin.

Bright Futures

Given the size of Florida's Bright Futures program, many prior student- and institution-level studies have explored the scholarship's effects on high school course-taking behaviors, college enrollment, and out-of-state migration for college. At the high school level, Harkreader, Hughes, Tozzi, and Vanlandingham (2008) found, based upon a multivariate regression, that BF may have increased students' enrollment in college preparatory courses. At the extensive enrollment margin, Harkreader et al. (2008) also used a logistic regression to suggest students eligible for a BF award were roughly 9 percentage points more likely to attend college and noted descriptive decreases in the state's rates of out-migration. Zhang, Hu, and Sensenig (2013) similarly estimated impacts on postsecondary enrollment by sector and location using institution-level data from the Integrated Postsecondary Education Data System in a difference-in-differences framework. The authors found BF to be associated with increased first-time, full-time enrollment counts of 22-24% by campus and reduced out-of-state migration of 8.1 percentage points in aggregate. For each of these studies, effects varied by students' socioeconomic status and/or ethnicity, and other prior work has suggested the program's merit-based design theoretically favors higher-income households who, on average, have higher SAT scores (increasing the likelihood they earn a full FAS award) and spend less on the state lottery that partially funds the program. Stranahan and Borg (2004) estimated a net benefit of $2,200 for these higher SES households compared to a net loss of $700 for lower SES households (given lower SAT scores, which led to a reduced likelihood of scholarship receipt overall or to receipt of the partial FMS award).

The most relevant prior work to our study is that of Zhang et al. (2016), whose study of the Bright Futures program produced some of the largest effects on college choice found in the literature. The authors similarly used a regression discontinuity design to estimate impacts of Bright Futures on students' attendance at any public college in Florida, though their estimation strategy and data limitations might bias results, as we discuss below. First, they examine the sector of college attendance for students in the 2004 through 2006 high school cohorts. In their preferred specification, Zhang et al. (2016) reported that eligibility for the higher FAS award on the SAT margin increased the likelihood of enrollment at a Florida public college by 6 percentage points. On the college choice margin, the authors concluded FMS- and FAS-eligible students were a sizeable 11.5 and 6.9 percentage points more likely to enroll at a public four-year institution (compared to a public two-year, though FAS and FMS awards can be used at both), respectively. Indeed, the authors note that such an "increase represents an approximately 50% increase in the probability of attending four-year vs. two-year colleges" (p. 132).[2] There are, however, a number of potential limitations with this prior quasi-experimental evaluation of BF. Zhang et al. (2016) relied on administrative data from the Florida Education Data Warehouse, preventing them from observing private high school graduates, out-of-state college enrollments, and in-state enrollments at private colleges. As subsidies might induce students to stay in-state or attend more expensive, private colleges, this limitation may bias inferences. Additionally, the primary running variable used in their RD analysis was each students' highest SAT score, rather than initial SAT. Given

[2] Given our observation of SAT scores only, we present Zhang et al.'s (2016) SAT results here and make any comparisons to that portion of their analysis only. The authors find similar impacts on the ACT margin, though there were some differences across sample selections and model specifications. In all, their results consistently point to significant and substantial impacts of Bright Futures on in-state and four-year college enrollment.

that half of SAT-takers retake the exam, students' manipulation of the assignment mechanism by test retaking given ex ante known scholarship thresholds threatens the internal validity of the RD design (Goodman, Gurantz, & Smith, 2020).

We replicate and extend these prior works on Bright Futures to consider its effect on college enrollment and degree completion given conflicting evidence in extant literature on the direction and magnitude of effects of state merit programs broadly. While we strengthen this body of evidence overall, we also contribute to emerging findings suggesting the effects of need-based programs may dwarf those of their merit-based counterparts. Our study also seeks to inform policy in an environment that continues to debate the costs and benefits of state investment in need versus merit awards. In all, we help answer whether BF impacts students' college-going and completion outcomes, and, if so, on what margins and for which students. To accomplish these aims, we make several improvements upon these prior works. Our data encompass additional and more recent cohorts of high school students, which we additionally observe for an extended period to estimate impacts on longer-run outcomes. We are also able to observe students in public and private high schools and the full picture of postsecondary enrollment and attainment outcomes: enrollment by level (two- or four-year), sector (public or private), and location (in-state or out-of-state), as well as attainment of any degree by level (associate or bachelor's) and location (in-state or out-of-state) up to six years after high school graduation. This more robust set of data allows us to observe students and outcomes that have been blind to prior works. Further, our empirical strategy also alleviates prior threats to internal validity by using students' first SAT score as the primary assignment mechanism, which allows us to yield valid causal inferences on the effect of BF.

Data and Methods

Our study leverages student-unit data from the College Board and covers the universe of SAT test takers in Florida's public and private high schools. College Board population data allow us to observe students' entire SAT test history and a variety of other factors, including gender, ethnicity, SAT writing performance (a third component of the SAT which is not used as award eligibility criteria), and PSAT test taking. We match these records to National Student Clearinghouse data on postsecondary outcomes, which also allow us to observe in-state and out-of-state enrollment at public and private postsecondary institutions nationwide, as well as any degree outcome by level and year (Dynarski, Hemelt, & Hyman, 2015). We focus our sample on the 2004 through 2012 high school cohorts so that we can observe degree completion outcomes up to six years after high school graduation.

Table 1 presents selected descriptive statistics for students in our analytic sample whose first SAT score falls within 60 points of an eligibility threshold. This subset captures more than 220,000 students near the FMS SAT threshold (970, or 980 in 2012) and over 71,000 students near the 1270 FAS SAT threshold. Roughly half of each sample are female, and the demographic composition reflects that of Florida's population; the majority are White, with Hispanics representing the second largest ethnic group. Additionally, as expected, students near the higher-scoring FAS threshold have higher 10th- and 11th-grade PSAT outcomes and higher SAT writing scores than the (relatively) lower-scoring FMS students. Across our outcomes of interest, students near the lower FMS threshold are equally split across attendance at in-state, two- and four-year public institutions where Bright Futures awards can be used, whereas FAS students enroll in public, in-state four-year and out-of-state (ineligible) institutions at substantially higher rates. Finally, while both groups appear more likely to earn a bachelor's degree from an in-state institution, nearly 60% of the FMS subsample earned any degree (associate or bachelor's) within six years of high school graduation compared to over 72% of the FAS subsample.

Regression Discontinuity Design

We estimate the causal impact of Florida's Bright Futures program on this sample of students using a regression discontinuity design that exploits variation in eligibility for an award around the FMS (970/980 SAT) and FAS (1270 SAT) test score thresholds. This design allows us to estimate the impact of BF eligibility on the SAT margin by comparing outcomes for BF-eligible students just above a respective threshold to similar yet BF-ineligible peers just below the threshold. As we cannot observe scholarship receipt, we begin by estimating the intent-to-treat (ITT) parameter, interpreted as the causal effect of the scholarship offer for students local to these cutoffs, based on their initial SAT exam. Formally, we first estimate a reduced-form specification given by equation (1):

$$Y_{it} = \beta_0 + \beta_1 Eligible_{it} + \beta_2 SAT_{it} + \beta_3 Eligible_{it} \times SAT_{it} + \theta_t + \mathbf{X}_i' \phi + \varepsilon_{it} \quad (1)$$

where Y_{it} is an outcome for student i in cohort t, $Eligible_{it}$ equals 1 if a student's first SAT score was above the relevant SAT cutoff, and SAT_{it} is a student's first SAT score, centered at the respective cutoff. θ_t are high school cohort dummies to restrict comparisons to students in the same graduation year, and $\mathbf{X}_i$ is a vector of baseline observable characteristics included only for robustness (gender, ethnicity, PSAT participation and scores, and SAT writing score). β_1 is the parameter of interest, the ITT effect of eligibility for a BF award for SAT test takers. We estimate equation (1) separately at each eligibility threshold and report results using a 60-point SAT bandwidth to mirror Zhang et al. (2016), linear specifications, and heteroscedastic-robust standard errors, though our results are robust to a series of alternate bandwidths and functional forms described later.

A regression discontinuity design yields valid causal inferences under several conditions, which are satisfied in our sample. First, though we include background characteristics in our estimation for robustness, our sample achieves covariate balance across these factors at the eligibility thresholds as shown in Appendix Table 1. The only observed difference reaching traditional significance is between PSAT scores for students at the FAS threshold when samples are conditioned on PSAT-taking, though these differences amount to less than 1 test point. Second, when using initial SAT as the primary running variable, we observe a smooth distribution of students' SAT scores in Figure 1. This distribution presents with no evidence of bunching that might indicate students are able to sort themselves in relation to the identified thresholds, even though the SAT eligibility criterion is ex ante known to students. We do, however, observe that students are aware of these thresholds, as those scoring below the lower FMS and higher FAS thresholds are roughly 6 and 9 percentage points more likely to retake the exam than students just above, respectively (Appendix Figures 1 and 2), with no evidence that retake rates change discontinuously at any other point in the SAT distribution. These increased retake rates among students just below the eligibility thresholds lead to a distribution of maximum SAT scores shown in Figure 2, which resembles the primary distribution of scores in Zhang et al.'s (2016) study. This presents a clear violation of a fundamental RD assumption and supports our use of students' initial SAT score as the primary assignment mechanism. Finally, we are unaware of other interventions occurring concurrently at these thresholds which would confound estimates and have not found information or concerns in prior literature on Bright Futures to suggest otherwise.

Instrumental Variable (Fuzzy RD) Approach

As is customary with regression discontinuity designs implemented in the presence of multiple assignment mechanisms or with unknown fidelity of implementation, we also implement a “fuzzy” RD using a two-stage least-squares or instrumental variable (IV) approach that leverages our running variable from the sharp design above (Imbens & Lemieux, 2008; Lee & Lemieux, 2010). Our ITT estimates from equation (1) assume that students who surpassed a respective SAT threshold were ultimately offered either the FMS or FAS award. That is, given our inability to observe ACT scores and other qualifiers, we assume that being eligible for the scholarship on the SAT margin is synonymous with receiving the scholarship. This is a reasonable assumption given evidence from Zhang et al. (2016), who were able to observe scholarship receipt, along with the GPA qualifier. Comparing students’ presence in Bright Futures award files against a classifier based solely on ACT/SAT scores resulted in minimal misclassification, motivating the authors’ use of scores as their primary running variable without the need to explicitly take GPA into account. For their overall analysis, Zhang et al. (2016) found that test scores accurately classified over 99% of students in each of their three cohorts and over 97% when they focused only on students with SAT scores.

Given additional eligibility factors, our inability to observe actual aid receipt, and endogenous SAT test retaking behaviors noted above, we use students’ first SAT score as an instrument for FMS or FAS award eligibility (i.e., ever receiving an SAT score above the exogenous eligibility threshold). That is, we first model the likelihood a student is ever eligible for a BF award on the SAT margin as a function of their first SAT score and then use that predicted likelihood of an FMS or FAS offer to estimate impacts on our outcomes of interest. This approach is given by the series in equation (2):

$$EverBF_{it} = \alpha_0 + \alpha_1 Eligible_{it} + \alpha_2 SAT_{it} + \alpha_3 Eligible_{it} \times SAT_{it} + \mathbf{X}_i'\Omega + \varepsilon_{it} \quad (2)$$

$$Y_{it} = \pi_0 + \pi_1 \widehat{EverB}F_{it} + \pi_2 SAT_{it} + \pi_3 Eligible_{it} \times SAT_{it} + \mathbf{X}_i'\Gamma + \varepsilon_{it} \ .$$

Here, the first stage predicts the likelihood student i in cohort t was ever eligible for a BF scholarship ($EverBF_{it}$) at the SAT margin as a function of their first SAT score (SAT_{it}), and α_1 represents the difference in the probability of being eligible for BF between students who are just above and just below the SAT threshold, given their first score. The second stage predicts the same Y_{it} outcomes but instruments for BF eligibility on the SAT margin. Here, π_1 is the Local Average Treatment Effect (LATE) of BF eligibility for students near the SAT thresholds.

This fuzzy RD approach allows us to approximate the change in probability of receiving an FMS or FAS offer at the SAT threshold, which we find varies from roughly 28 to 37 percentage points (see Results below), though this approach requires a few additional assumptions. One is that unobservable GPA and community service requirements are not strong predictors of grant receipt, which was shown to be true by Zhang et al. (2016). The more challenging assumption is that students who continue to score below the SAT threshold do not receive the grant, but this is complicated by the fact that we cannot observe ACT scores, which can independently lead to aid eligibility for these students. We again rely on Zhang et al. (2016), who show that about half of students who took the SAT did not take the ACT. In practice then, the first-stage we identify produces a lower bound estimate of the LATE, and under extreme assumptions could be significantly larger. Even in a "worst case" scenario, for example, that shrinks our first-stage by half (or doubles our IV estimates), our point estimates are generally small and still reject larger estimates found in prior work on Bright Futures.[3]

[3] A number of pieces of evidence suggest that these extreme scenarios are unlikely to pass. For example, on page 127, Zhang et al. (2016) note that roughly equal numbers of students took only the SAT or both the SAT and ACT, with significantly fewer students only taking the ACT. Although this refers to the whole population and not just to students near the cutoff, these results provide a useful benchmark or a "worst case" scenario. The largest concern is if only students who score below the SAT are the ones who also take the ACT, but Zhang et al. (2016) also show this is not the case in a number of ways (e.g., the cumulative distributions of SAT and ACT scores in Table 1 for students who took both exams largely overlap; the regression estimates in Tables 4 and 5 rely on students who took both exams and are large in number).

Instrumental variable methods produce estimates applicable to a “complier” subset of the population. Compliers are students who, in essence, comply with their original treatment assignment, meaning those who are only eligible for the BF scholarship by virtue of initially scoring above the SAT threshold. In our setting, compliers below the threshold includes students who either did not re-take the exam or whose subsequent SAT scores remained below the threshold. While students below the threshold do re-take the exam—plausibly to gain eligibility—roughly 45% of students just below the FMS threshold and 35% below the FAS threshold in our sample never receive a sufficiently high SAT score, and so are disqualified from receiving a BF award (Appendix Figures 3 and 4). As noted above, we assume scoring above these SAT thresholds is equivalent to being eligible for a BF award. Thus, our estimates can be interpreted under this assumption as the outcome difference between students who were made eligible for a BF award as a function of their first SAT score and those who remain ineligible on the SAT margin. This complier population is likely to include the types of students that states target in their financial aid and outreach efforts. Students who scored below the threshold and never met the benchmark attended college 88% of the time, compared to a 96% college-going rate among those who initially scored below the benchmark but who later exceeded the SAT threshold. Even more stark are the differences in where they attend, as 51% of these students who remain below the benchmark attend an in-state, two-year college compared to only 29% of those with lower initial scores who eventually score above the SAT threshold. We do not find large ethnic differences in the complier population; 60% of students who later met the SAT threshold were White or Asian, compared to 55% of students who did not.

Like RD designs, instrumental variable methods produce valid causal inferences under certain assumptions, including relevance, exclusion, exchangeability, and monotonicity (Angrist, Imbens, & Rubin, 1996). Our instrument is relevant to treatment assignment given that students must pass an SAT threshold on any attempt to eventually qualify for a Bright Futures award. This is true whether students also surpass the weighted GPA and community service thresholds or not as all criteria must be met. Our instrument also plausibly meets the exclusion restriction given that surpassing any given SAT threshold should not affect students' college enrollment or degree attainment outcomes except through the receipt of aid or other resources associated with such an achievement (i.e., as modeled here through the receipt of a BF award). In this context, the guarantee of aid to cover postsecondary expenses in whole or in part via an FMS or FAS award (through qualification via the instrument) likely sends a positive signal to students of their academic abilities and promise, providing motivation to enroll and complete. One possible violation of the exclusion restriction in our case could be effects of BF on students' general motivations or preparations for college regardless of aid eligibility or receipt, though studies generally find that the financial payment is typically the motivating factor (e.g., Gurantz, Hurwitz, and Smith (2017)), and the presence of aid programs that produce null results also point against this being a common occurrence. Next, our instrument meets the exchangeability (or independence) assumption given that the possibility of achieving a SAT score above a respective threshold is available to all students. For robustness, we also condition our estimates on baseline characteristics to control differences in students' observable demographic and academic outcomes and to test this ignorability, and our inferences remain unchanged to this inclusion. Finally, we assume monotonicity or "no defiers" in that there should be no students who (1) are not offered or decline the award if they surpass the SAT threshold or (2) would somehow be offered or accept an award if they fell below. We note again in our interpretations below that IV impacts of the local average treatment effect are relevant just to such compliers and not the full population.

Results

Across our reduced-form and instrumental variable estimates, we generally find that eligibility for a Bright Futures award does not result in statistically significant changes in (1) whether a student attends college overall; (2) the sector, level, or geographic location of postsecondary enrollment; or (3) associate or bachelor's degree attainment. Table 2 presents these results for our primary college enrollment outcomes. The largest positive reduced-form point estimates are a 0.3 percentage point increase in four-year public enrollment at the lower FMS threshold and a 0.6 percentage point increase in four-year private enrollment at the higher FAS threshold, with both estimates precisely estimated yet statistically insignificant. Regression discontinuity plots for in-state enrollment by sector and level, as well as any out-of-state enrollment, are included in Figure 3 for the FMS threshold and Figure 4 for FAS. As expected, all figures depicting effects at the enrollment margins show no discontinuities between eligible and just-ineligible peers. Instrumental variable estimates are presented in the same place (Table 2). These estimates are still small, yet even with tens or hundreds of thousands of observations, a modest 28-37 percentage-point first-stage precludes us from rejecting some small but potentially meaningful impacts. We can, however, confidently reject the large effects observed in prior studies. The largest impacts—statistically insignificant but consistent with reduced-form estimates—show that higher-scoring FAS students may have shifted in-state enrollment out of four-year, public colleges into private colleges by 2 percentage points.

Table 3 presents reduced-form and instrumental variable results for Bright Futures impacts on degree attainment, with regression discontinuity plots for six-year degree completion outcomes by level and location presented in Figures 5 (FMS) and 6 (FAS). The largest positive reduced-form impact on degree completion is a statistically insignificant 0.5 percentage point increase in earning an in-state bachelor's degree after six years for lower-scoring FMS students. Yet the only statistically significant impact observed suggests that BF eligibility actually reduced in-state associate degree attainment by 0.8 percentage points for these same students, suggesting a potential tradeoff with declines in associate degree completion only somewhat matched by increases in bachelor's degrees. The related IV estimates for this result show a negative impact for lower-scoring FMS students on in-state associate attainment of roughly 2.2 percentage points, with a statistically insignificant but positive impact on in-state bachelor's degree completion of 1.5 percentage points. Altogether, these results reject prior findings on improved college-going, four-year enrollment, and reduced out-migration for college. Although these are a select sample of relatively high-scoring SAT students, the baseline six-year completion rates are 55% and 72% for FMS and FAS students, respectively, suggesting significant room for improvement—which does not appear to be achieved through BF.

Robustness

Our regression discontinuity results are robust to specifications that vary the bandwidth and functional form, as well as to the inclusion of covariates in the estimation. Appendix Tables 2-4 present these robustness tests for enrollment outcomes, including in-state enrollment at two- and four-year institutions, and any out-of-state enrollments, respectively. Appendix Tables 5-6 present the same robustness tests for in-state associate and bachelor's degree attainment outcomes. For each outcome, we vary the bandwidth from 30 to 100 points, model linear and quadratic forms, and alter the inclusion of covariates. These robustness checks support the primary analysis' generally null findings and provide consistent evidence of Bright Futures' small but negative effect on students' in-state associate degree attainment, where all reduced form estimates are significant and range from negative 0.6 to negative 3.2 percentage points (Appendix Table 5).

Heterogeneity by SES and Ethnicity

As noted, prior studies on Bright Futures have argued that the benefits of the program may favor higher-income households (Harkreader et al., 2008; Stranahan & Borg, 2004). We test for these heterogeneous effects by linking students' home addresses to a zip code level socioeconomic status (SES) indicator yielded by multiple factors in the American Community Survey 2012 5-year estimates. Guided by Jang and Reardon (2019) and Cowen et al. (2012), we developed a standardized SES scale that includes median income, housing costs, educational attainment, unemployment rate, and poverty, which achieved an internal reliability coefficient of 0.84. We disaggregate our sample into three SES terciles and re-estimated our instrumental variable models within these separate groups. We present IV results because reduced-form comparisons across terciles are complicated by different first-stage estimates in ever surpassing the eligibility threshold; essentially, students in high-income areas are more likely to retake the exam, improve their performance, and earn an SAT score above the threshold, which results in a smaller first-stage impact on aid eligibility.

Instrumental variable results by SES tercile of Bright Futures' effects on select attendance and completion outcomes are presented in Table 4. We find no statistically significant differences in enrollment outcomes at either threshold overall or by sector, level, or location. Similarly, there are generally no impacts of BF eligibility on six-year degree attainment with one exception. We observe that the negative impacts on in-state associate degree completion are driven by relatively lower-scoring (FMS) students living in the lowest or middle SES zip codes. For students in middle SES areas, we estimate BF reduced in-state associate degree completion by 3.4 percentage points, but in-state bachelor's attainment appears to substitute for this reduction with a marginally significant 3 percentage point increase. For students in the lowest SES areas, however, we observe a 4.6 percentage point reduction in in-state associate degree completion with no plausible substitutions. This suggests our main findings of reduced in-state associate degree attainment are primarily driven by completion outcomes of low-SES students. These findings counter the assertion by Harkreader et al. (2008) that BF increased in-state enrollment among lower-income students but do support Stranahan and Borg (2004) in part by suggesting BF may have differential impacts by socioeconomic status.

We also test for heterogeneous effects across ethnic groups given that a robust body of evidence documents such differential impacts of state merit programs, including prior studies on Bright Futures (Chen & DesJardins, 2010; Dynarski, 2000; Harkreader et al., 2008; Pallais, 2009; Zhang et al., 2013). Instrumental variable results by ethnicity are presented in Table 5. Overall, we generally find no differential effects across attendance or completion outcomes. We do observe a potential enrollment shift among Hispanic students at the FMS threshold of approximately 4.5 percentage points away from in-state two-year institutions toward public, in-state four-year institutions. Similarly, we estimate a 3.8 percentage point increase in enrollment at private, in-state four-year institutions for higher-scoring White students at the FAS threshold, which appears to be driven by a diversion from in-state publics rather than a reduction in out migration. These generally null impacts are encouraging in that BF may not differentially privilege students from one ethnic group over another but are also consistent with our primary analyses in that the program appears to have minimal-at-best effects on students' college-going and completion outcomes, even after disaggregating by socioeconomic status and ethnicity.

Replication of Prior Work

Why do our results differ from the prior RD analysis in Zhang et al. (2016)? The key difference between our studies is our use of the initial SAT score as the primary running variable versus their use of the maximum SAT score. Using the maximum SAT, after students have had opportunities to retake the exam and sort relative to the eligibility threshold, is problematic in practice as it leads to comparisons of students with different unobservable levels of achievement or motivation, as evidenced by the bunching in Figure 2. To more closely compare these differences across papers, we first implement a series of biased regressions on students' enrollment outcomes (as analyzed by Zhang et al. (2016)) using the endogenous maximum SAT score as the assignment mechanism for our newer and larger analytic sample. Unlike Zhang et al. (2016), we can also estimate impacts on in-state private and out-of-state enrollments using their same bandwidth and linear specification with no controls, which is shown in Table 6. Across almost all enrollment outcomes—overall and by sector, level, and location—we detect significant effects with this biased running variable. These effects mirror Zhang et al. (2016) by suggesting positive impacts on four-year public college enrollment, though our biased results are still substantially smaller than their estimates. When analyzing students with only SAT scores, Zhang et al. (2016) reported increases in overall public college enrollment of 6.3 percentage points at the FAS threshold and impacts on four-year (versus two-year) enrollment of 9.8 and 6.6 percentage points at the FMS and FAS thresholds. We detect impacts using maximum SAT on overall college going, but to the effect of 0.9 percentage points at the 1270 threshold, and corresponding impacts on in-state four-year college enrollment of roughly 1.2 and 4.2 percentage points at the same FMS and FAS thresholds. Thus, a main difference in findings between the papers seems driven by the use of initial versus maximum SAT as the assignment variable, though there are still meaningful differences in the size of these effects.

A key improvement of our study is also the incorporation of additional and newer cohorts, and our ability to detect students from in-state public and private high schools who attend in-state public or private institutions or any out-of-state college. This greater coverage could in part further explain our studies' differences. As an additional comparison, we attempt to remove these differences by replicating the Zhang et al. (2016) sample as close as possible by (1) restricting our sample to the 2004-2006 cohorts using only students in public high schools and (2) assuming we cannot observe in-state private or any out-of-state postsecondary enrollments. Doing so allows us to also achieve roughly equivalent baseline college-going rates between our samples, suggesting that any remaining sample differences between the two studies is not likely to explain differences between their findings.[4] Table 7 shows the results of this analysis using both the initial and maximum SAT scores. Even these biased impacts show no effect on overall college-going at either threshold, and only a 3.1 point increase in public four-year enrollment at the FAS threshold. These again are much smaller than Zhang et al.'s (2016) estimates of over 6 percentage points. When using the initial SAT score as the (unbiased) running variable, no significant impacts are observed. Thus, even upon closer replication, neither the more appropriate running variable, the recent time period, nor reliance on Florida's administrative data appear to fully explain these differences as (1) we still do not find that Bright Futures impacts overall in-state enrollment, (2) using only public high school students produces identical estimates to our own above, and (3) results are similarly null when using only the 2004 through 2006 cohorts.

Perhaps the closest comparison to our study, however, comes from the Zhang et al. (2016) study itself. In a later section on test retaking, the authors show significant bunching based on maximum SAT, and thus estimated IV impacts based on initial SAT scores, quadratic slopes, and controls. In their paper, these new, "unbiased" regressions produce point estimates and standard

[4] Zhang et al. (2016) reported baseline in-state, public college-going rates of approximately 78% at the lower FMS and 75-80% at the higher FAS threshold (Figure 1, Panel A, p. 128). In our restricted sample where we mimic their approach, we observe these rates to be roughly 76% and 71%, respectively, though the true enrollment rate is above 90% once out-of-state and in-state private colleges are taken into account.

errors that are mostly unchanged from, or remain identical to, their earlier tables based on the biased regressions. Thus, they surprisingly find similar effects when using both initial and maximum SAT scores under separate regression discontinuity and instrumental variable designs. Their paper does not discuss how they construct their IV estimates, and they do not show the reduced form impacts or first-stage estimates of this specification which would allow us to compare results. Our most similar model is likely our robustness test in Appendix Table 3 for the in-state, four-year enrollment outcome, where we incorporate quadratic slopes and controls. With the same specification, we estimate marginally significant effects at the FMS threshold only of roughly 1.3 percentage points compared to Zhang et al.'s (2016) highly significant estimate of nearly 12 points. In all, it is clear that methodological improvements and more robust data help to account for many but not all differences between our studies.

Beyond our study's improvements, as discussed above, we are confident any sample differences are not able to explain these impact differences. Furthermore, we are also confident that any differences in outcome measurements are not driving these differences. Zhang et al. (2016) do not explicitly identify their enrollment horizon of interest (e.g., immediate upon high school graduation or delayed), yet given Bright Futures' availability for a first award to begin up to three years after high school graduation, we estimate impacts for any eventual enrollment and assume Zhang et al. (2016) do the same. However, our results also remain unchanged when we condition outcomes on first year (immediate) enrollment, as most higher-scoring SAT students choose to enroll in college upon high school graduation. Thus, even when mirroring their sample and specification as closely as possible and considering other possible differences between our studies, larger-than-can-be-explained differences remain. Under the causal replication framework developed by Wong et al. (2021), after establishing equivalence of populations, treatment, outcomes, and identifying assumptions in such close replications, any remaining differences between studies could be explained in part by what they refer to as "inaccuracies in reporting", which contribute to the reproducibility and replicability crises in the social sciences (Bollen, Cacioppo, Kaplan, Kronsnick, & Olds, 2015).

Discussion

In this study, we replicated and extended prior works on Florida's Bright Futures merit aid program to consider its effect on college enrollment and degree completion. We tested for heterogeneous impacts across students' ethnicity and socioeconomic status while additionally making important improvements upon prior studies through the use of more robust data and better methodological decisions to guard against threats to internal validity. While we correct prior inferences on BF, we also serve as a strong example of the importance of proper implementation of regression discontinuity designs in scenarios with ex ante known eligibility thresholds. Our study makes important contributions not only as the first to estimate impacts of BF on students' ultimate degree attainment, but also as an addition to the growing body of evidence considering the efficacy of state merit aid programs and the relative size of their impacts given need-based alternatives (Herbaut & Geven, 2020; Nguyen et al., 2019). Our findings also stand to inform public policy in Florida and states with similar merit programs by providing stronger evidence on the effectiveness of BF.

We find that eligibility for a Bright Futures award on the SAT margin does not result in statistically significant or positive changes in (1) whether a student attends college overall; (2) the sector, level, or geographic location of postsecondary enrollment; or (3) associate or bachelor's degree attainment. The one consistent finding is that students near the lower SAT threshold for FMS are roughly 1 to 2 percentage points less likely to earn an associate degree from an in-state institution, with this effect concentrated among students in low-SES zip codes. Combined with a positive but slightly smaller and statistically insignificant increase in the probability of earning an in-state bachelor’s degree (0.5 to 1.5 percentage points), these changes in degree attainment suggest BF may result in a simple even-trade in the best-case scenario (i.e., by inducing minimal shifts of students from in-state associate degrees to in-state bachelor’s degrees). Further, we also observe suggestive BF-induced enrollment shifts for Hispanic and White students away from two-year colleges and toward in-state, four-year and private institutions, respectively. Yet, judged against the enrollment and completion goals of BF and other large-scale merit programs, these minimal-at-best impacts indicate the costs of BF may not outweigh such small and short-run benefits. In all, our results confidently reject prior studies suggesting that BF has large impacts on enrollment and out-migration for college. One possible explanation of this may be that the eligibility criteria were already too high. Overall college attendance rates at the FMS and FAS thresholds in our sample, once we include out of state and private college enrollment, were already 92% and 96%, respectively, suggesting little room for improvement. (Florida has raised the SAT eligibility thresholds multiple times since 2011-12, suggesting awards now go to students with even higher levels of baseline enrollment.) Degree completion rates fall well short of these benchmarks, suggesting aid could potentially benefit students if they translate into higher graduation rates, but we find no positive impacts on attainment in our study. Though we find no meaningful impacts on students’ enrollment or completion outcomes, BF could influence other outcomes, including by limiting the need to work while enrolled, reducing borrowing, or increased long-run labor-market earnings (Chapman, 2016; Evans & Nguyen, 2019). Future research on these outcomes could contribute to this debate and help researchers and policymakers alike better understand the full extent of the impacts of this merit aid program—and the horizon in which these impacts could be expected.

While our findings differ from prior works on Bright Futures, they align with more emerging notions of state merit aid and study replication. Our results provide further evidence that merit aid programs may have negative, null, or small effects on students' enrollment and attainment outcomes, particularly when considered against need-based aid (Cohodes & Goodman, 2014; Fitzpatrick & Jones, 2016; Herbert & Geven, 2020; Nguyen et al., 2019; Sjoquist & Winters, 2012). These findings have important implications given the stated purposes of state merit aid in increasing access, enrollment, and attainment while also reducing out-migration for college. Our findings are also stark given shifting state support for merit versus need-based aid and the "best and brightest" rhetoric that surrounds these programs, despite recurring evidence that merit programs may do little for students who were not already likely to enroll and succeed. This further underscores the need for researchers and policymakers alike to take care when considering what "works" and shows that some prevailing findings on aid programs may warrant revisiting in the presence of better microdata, newer methods or improved implementation strategies, larger sample sizes, and longer observation periods. This is at the heart of replication and extension (Granger & Maynard, 2015; Hedges, 2018; Wong et al., 2021). Indeed, previous reviews and replications of merit aid studies suggested prior findings may have been overstated and that these programs indeed had minimal—if any at all—impacts on enrollment and completion (Fitzpatrick & Jones, 2016; Sjoquist & Winters, 2012).

Limitations

Our study is not without notable limitations. First, as is the case for any regression discontinuity analysis, our results cannot account for general effects on outcomes like student motivation that might arise from the implementation of Bright Futures. Indeed, the presence of merit aid programs has been shown to increase student achievement in high schools, which could have improved students’ likelihood of being eligible for a BF award (Pallais, 2009; Scott-Clayton, 2011). Second, our estimation of a local average treatment effect applies only to students with a first SAT score near a respective FMS or FAS threshold, who are already relatively high achieving, and in particular should be thought of as an effect on compliers who would not receive aid simply by virtue of scoring below the eligibility threshold on their initial exam. It is possible BF had effects across the ability spectrum where our inferences cannot be extrapolated. Finally, we are methodologically limited by a modest 27-38 percentage point first stage, which is likely to be slightly lower in practice given unobservable ACT eligibility criterion. Thus, our findings again likely represent lower-bound estimates of the effect of BF. Our instrumental variable impact estimates are, however, rejecting findings as small as 1-2 percentage points, and descriptive results in Zhang et al. (2016) on the prevalence and distribution of scores for students taking both exams suggest we can confidently rejecting the large estimates observed in prior studies. Even in the face of these limitations, our paper provides the strongest causal estimates of Bright Futures’ impacts on college enrollment, out-migration, and attainment, and consistently suggests the program’s monetary payoff has little impact for students at these margins.

Conclusion

Research on Bright Futures and other financial aid programs continues to inform public policy, particularly concerning ways to increase college access and completion (Dynarski & Scott-Clayton, 2013; Page & Scott-Clayton, 2016). While we find no effects of Florida’s BF program on college enrollment, out-migration, and degree attainment, need-based financial aid remains a viable mechanism for states to increase students’ college access, persistence, degree-attainment, and labor-market outcomes (Bettinger et al., 2019; Deming & Dynarski, 2010; Nguyen et al., 2013). Our work suggests Florida and other states strongly consider investments in need-based aid given these consistently proven benefits and programs' abilities to reduce constraints or provide incentives for students who would have otherwise not enrolled or succeeded; this is especially true in light of prior research that found the need-based Florida Student Access Grant did indeed improve student outcomes (Castleman & Long, 2016).

Ultimately, the diversity of existing aid programs, factors endogenous to aid program adoption and receipt, and data and measurement constraints make rigorous evaluations of these policies challenging. Our paper provides an expanded view of students' outcomes as they pertain to the Bright Futures scholarship and supports a growing body of work suggested prior findings on merit aid may have been overstated upon replication with little or no real impacts on enrollment and completion (Fitzpatrick & Jones, 2016; Sjoquist & Winters, 2012). This underscores the need for researchers and policymakers to take care when considering existing evidence on programs' efficacy, drawing implications from the entirety of available evidence rather than individual studies. Our study calls for further replications and extensions, particularly given the advent of better data and research tools.

References

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association, 91*(434), 444-455.

Becker, G. (1975). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education, 2nd edition*: National Bureau of Economic Research.

Bettinger, E. P., Gurantz, O., Kawano, L., Sacerdote, B. I., & Stevens, M. (2019). The Long Run Impacts of Financial Aid: Evidence from California's Cal Grant. *American Economic Journal: Economic Policy, 11*(1), 64-94.

Bollen, K., Cacioppo, J. T., Kaplan, R., Kronsnick, J., & Olds, J. L. (2015). *Social, behavioral, and economic sciences perspectives on robust and reliable science*. National Science Foundation. Washington DC. Retrieved from https://www.nsf.gov/sbe/SBE_Spring_2015_AC_Meeting_Presentations/Bollen_Report_on_Replicability_SubcommitteeMay_2015.pdf

Carruthers, C. K., & Fox, W. F. (2016). Aid for all: College coaching, financial aid, and post-secondary persistence in Tennessee. *Economics of Education Review, 51*, 97-112.

Castleman, B. L., & Long, B. T. (2016). Looking beyond Enrollment: The Causal Effect of Need-Based Grants on College Access, Persistence, and Graduation. *Journal of Labor Economics, 34*(4), 1023-1073.

Chapman, S. (2016). *Student loans and the labor market: Evidence from merit aid programs*. Washington Center for Equitable Growth. Washington DC.

Chen, R., & DesJardins, S. L. (2010). Investigating the impact of financial aid on student dropout risks: Racial and ethnic differences. The Journal of Higher Education, 81(2), 179-208. *Journal of Higher Education, 81*(2), 179-208.

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*(4), 251-285.

Cornwell, C. M., Mustard, D. B., & Sridhar, D. J. (2006). The Enrollment Effects of Merit-Based Financial Aid: Evidence from Georgia's HOPE Program. *Journal of Labor Economics, 24*(4), 761-786.

Cowen, C. D., Hauser, R. M., Kominski, R. A., Levin, H. M., Lucas, S. R., Morgan, S. L., . . . Chapman, C. (2012). *Improving the measurement of socioeconomic status for the National Assessment of Educational Progress: A theoretical foundation*. National Center for Education Statistics. Washington DC.

Deming, D. J., & Dynarski, S. (2010). College aid. In P. B. Levine & D. J. Zimmerman (Eds.), *Targeting investments in children: Fighting poverty when resources are limited*. Chicago, IL: University of Chicago Press.

Denning, J. T., Marx, B. M., & Turner, L. J. (2019). ProPelled: The Effects of Grants on Graduation, Earnings, and Welfare. *American Economic Journal: Applied Economics, 11*(3), 193-224.

Domina, T. (2014). Does Merit Aid Program Design Matter? A Cross-Cohort Analysis. *Research in Higher Education, 55*(1), 1-26.

Dynarski, S. M. (2000). Hope for Whom? Financial Aid for the Middle Class and Its Impact on College Attendance. *National Tax Journal, 53*(3), 629-662.

Dynarski, S. M. (2003). Does Aid Matter? Measuring the Effect of Student Aid on College Attendance and Completion. *American Economic Review, 93*(1), 279-288.

Dynarski, S. M. (2004). The New Merit Aid. In C. Hoxby (Ed.), *College Choice: The Economics of Where to Go, When to Go, and How to Pay for It* (pp. 63-97). Chicago, IL: University of Chicago Press.

Dynarski, S. M. (2008). Building the stock of college-educated labor. *Journal of Human Resources, 43*(3), 576–610.

Dynarski, S. M., Hemelt, S. W., & Hyman, J. M. (2015). The missing manual: Using National Student Clearinghouse data to track postsecondary outcomes. *Educational Evaluation and Policy Analysis, 37*, 53S-79S.

Dynarski, S. M., & Scott-Clayton, J. (2013). Financial aid policy: Lessons from research. *The Future of Children, 23*(1), 67-91.

Evans, B. J., & Nguyen, T. D. (2019). Monetary Substitution of Loans, Earnings, and Need-based Aid in Postsecondary Education: The Impact of Pell Grant Eligibility. *Economics of Education Review, 70*, 1-19.

Fitzpatrick, M. D., & Jones, D. (2016). Post-baccalaureate migration and merit-based scholarships. *Economics of Education Review, 54*, 155-172.

Florida Department of Education. (2011). *Florida Bright Futures Scholarship Program (BF): End-of-Year Report (2010-11)*. Office of Student Financial Assistance, Florida Department of Education. Tallahassee, FL.

Florida Department of Education. (2018). *Florida Bright Futures Scholarship Program (BF): End-of-Year Report (2017-18)*. Office of Student Financial Assistance, Florida Department of Education. Tallahassee, FL.

Florida Department of Education. (2019). *Florida Bright Futures scholarship statistical reports: B-Florida high school graduates eligible for and receiving Bright Futures*. Office of Student Financial Assistance, Florida Department of Education. Tallahassee, FL.

Goodman, J. S., Gurantz, O., & Smith, J. (2020). Take Two! SAT Retaking and College Enrollment Gaps. *American Economic Journal: Economic Policy, 12*(2), 115-158.

Granger, R. C., & Maynard, R. (2015). Unlocking the Potential of the "What Works" Approach to Policymaking and Practice:Improving Impact Evaluations. *American Journal of Evaluation, 36*(4), 558-569.

Gurantz, O., Hurwitz, M., & Smith, J. (2017). College Enrollment and Completion Among Nationally Recognized High-Achieving Hispanic Students. *Journal of Policy Analysis and Management, 36*(1), 126-153.

Harkreader, S., Hughes, J., Tozzi, M. H., & Vanlandingham, G. (2008). The impact of Florida's Bright Futures scholarship program on high school performance and college enrollment. *Journal of Student Financial Aid, 38*(1), 5-16.

Hedges, L. V. (2018). Challenges in Building Usable Knowledge in Education. *Journal of Research on Educational Effectiveness, 11*(1), 1-21.

Herbaut, E., & Geven, K. (2020). What works to reduce inequalities in higher education? A systematic review of the (quasi-)experimental literature on outreach and financial aid. *Research in Social Stratification and Mobility, 65*, 100442.

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics, 142*(2), 615-635.

Jang, H., & Reardon, S. F. (2019). States as Sites of Educational (In)Equality: State Contexts and the Socioeconomic Achievement Gradient. *AERA Open, 5*(3), 2332858419872459.

Kane, T. J. (2003). *A Quasi-Experimental Estimate of the Impact of Financial Aid on College-Going*. NBER Working Paper No. 9703.

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature, 48*(2), 281-355.

Long, B. T., & Riley, E. (2007). Financial aid: A broken bridge to college access? *Harvard Educational Review, 77*(1), 39-63.

National Association of State Student Grant and Aid Programs. (2018). *49th annual survey report on state-sponsored student financial aid: 2017-18 academic year*. Author. Washington, DC.

Nguyen, T. D., Kramer, J. W., & Evans, B. J. (2019). The Effects of Grant Aid on Student Persistence and Degree Attainment: A Systematic Review and Meta-Analysis of the Causal Evidence. *Review of Educational Research, 89*(6), 831-874.

Page, L. C., & Scott-Clayton, J. (2016). Improving college access in the United States: Barriers and policy responses. *Economics of Education Review, 51*, 4-22.

Pallais, A. (2009). Taking a Chance on College: Is the Tennessee Education Lottery Scholarship Program a Winner? *Journal of Human Resources, 44*(1), 199-222.

Scott-Clayton, J. (2011). On Money and Motivation: A Quasi-Experimental Analysis of Financial Incentives for College Achievement. *Journal of Human Resources, 46*(3), 614-646.

Singell, L. D., & Stone, J. A. (2002). The Good, the Poor, and the Wealthy: Who Responds Most to College Financial Aid? *Bulletin of Economic Research, 54*(4), 393-407.

Sjoquist, D. L., & Winters, J. V. (2012). Building the Stock of College-Educated Labor Revisited. *Journal of Human Resources, 47*(1), 270-285.

Sjoquist, D. L., & Winters, J. V. (2014). Merit aid and post-college retention in the state. *Journal of Urban Economics, 80*, 39-50.

Stranahan, H. A., & Borg, M. O. (2004). Some Futures are Brighter than Others: the Net Benefits Received by Florida Bright Futures Scholarship Recipients. *Public Finance Review, 32*(1), 105-126.

Wong, V. C., Anglin, K., & Steiner, P. M. (2021). *Design-based approaches to causal replication studies* EdWorking Paper: 20-311. Annenberg Institute, Brown University. Providence, RI. Retrieved from https://doi.org/10.26300/xsqw-c323

Zhang, L., Hu, S., & Sensenig, V. (2013). The Effect of Florida's Bright Futures Program on College Enrollment and Degree Production: An Aggregated-Level Analysis. *Research in Higher Education, 54*(7), 746-764.

Zhang, L., Hu, S., Sun, L., & Pu, S. (2016). The Effect of Florida's Bright Futures Program on College Choice: A Regression Discontinuity Approach. *The Journal of Higher Education, 87*(1), 115-146.

Table 1. Descriptive statistics.

	Florida Medallion Scholar (FMS)	Florida Academic Scholar (FAS)
SAT Threshold	970/980	1270
Analytic Sample Size	220245	71201
Background Characteristics		
Female	56.1%	45.3%
Asian	4.2%	6.4%
African-American	12.0%	3.6%
Hispanic	23.4%	16.1%
White	54.8%	69.5%
Other Race/Ethnicity	5.6%	4.5%
Took 11th Grade PSAT	44.7%	74.2%
11th Grade PSAT score (60-240)	142	182
Took 10th Grade PSAT	72.0%	81.6%
10th Grade PSAT score (60-240)	131	169
SAT Writing Score (200-800)	465	594
National Student Clearinghouse Data		
Attendance		
In-State, Two-Year	38.1%	9.9%
In-State, Four-Year Public	38.2%	56.7%
In-State, Four-Year Private	5.6%	7.3%
Out-of-State	10.4%	21.7%
Degree Attainment		
Any Degree in 6 Years	55.9%	72.1%
In-State Associate Degree in 4 Years	14.0%	5.6%
In-State Associate Degree in 6 Years	19.5%	7.7%
In-State Bachelor's Degree in 4 Years	18.6%	32.7%
In-State Bachelor's Degree in 6 Years	34.4%	48.8%

Note: Includes all of Florida's SAT takers who graduated high school from 2004 through 2012 with initial SAT within 60 points of an eligibility threshold.

Table 2. Impact of Bright Futures on college attendance.

			In State				Out of State			
		Attend College	Any	Two-Year	Four-Year, Public	Four-Year, Private	Any	Two-Year	Four-Year, Public	Four-Year, Private
Eligibility for Florida Medallion Scholars (FMS) 970/980 SAT Threshold										
Reduced Form		-0.0017	0.0008	-0.0043	0.0033	0.0018	-0.0025	-0.0013	-0.0015	0.0003
		(0.0023)	(0.0033)	(0.0042)	(0.0042)	(0.0020)	(0.0026)	(0.0011)	(0.0016)	(0.0019)
	First Stage									
Instrumental Variable	0.3704**	-0.0046	0.0022	-0.0116	0.0089	0.0048	-0.0067	-0.0036	-0.0040	0.0008
	(0.0033)	(0.0063)	(0.0090)	(0.0113)	(0.0113)	(0.0053)	(0.0071)	(0.0031)	(0.0043)	(0.0051)
Baseline Means		92.1%	81.7%	39.5%	36.9%	5.3%	10.4%	1.8%	3.6%	5.0%
Eligibility for Florida Academic Scholars (FAS) 1270 SAT Threshold										
Reduced Form		0.0022	0.0011	0.0003	-0.0049	0.0058	0.0011	0.0010	0.0016	-0.0015
		(0.0031)	(0.0067)	(0.0045)	(0.0075)	(0.0040)	(0.0063)	(0.0011)	(0.0035)	(0.0055)
	First Stage									
Instrumental Variable	0.2837**	0.0079	0.0038	0.0009	-0.0174	0.0204	0.0040	0.0037	0.0058	-0.0054
	(0.0055)	(0.0109)	(0.0236)	(0.0158)	(0.0266)	(0.0139)	(0.0222)	(0.0039)	(0.0125)	(0.0195)
Baseline Means		95.7%	74.4%	10.1%	57.2%	7.0%	21.3%	0.6%	5.5%	15.2%

Notes: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. Reduced form estimates are regression discontinuity treatment effects using initial SAT score as the running variable, linear slopes, and a bandwidth of 60 SAT points. Instrumental variable estimates are treatment effects using initial SAT score to predict eligibility based on an individual ever scoring above the relevant SAT threshold. Sample includes all of Florida's SAT takers who graduated high school from 2004 through 2012. Sample sizes at the 970/980 and 1270 threshold are 220,245 and 71,201 observations, respectively. Baseline means include all observations whose initial SAT score was 10 or 20 points below the eligibility threshold.

Table 3. Impact of Bright Futures on four- and six-year degree attainment.

			4-Year				6-Year			
		Any Degree in Six Years	Associate, In-State	Associate, Out-State	Bachelor, In-State	Bachelor, Out-State	Associate, In-State	Associate, Out-State	Bachelor, In-State	Bachelor, Out-State
Eligibility for Florida Medallion Scholars (FMS)										
970/980 SAT Threshold										
Reduced Form		-0.0006	-0.0056+	0.0005	0.0010	0.0001	-0.0080*	-0.0003	0.0054	0.0005
		(0.0042)	(0.0030)	(0.0007)	(0.0033)	(0.0014)	(0.0034)	(0.0009)	(0.0040)	(0.0018)
	First Stage									
Instrumental Variable	0.3704**	-0.0017	-0.0150+	0.0013	0.0028	0.0003	-0.0216*	-0.0009	0.0145	0.0014
	(0.0033)	(0.0112)	(0.0080)	(0.0019)	(0.0089)	(0.0039)	(0.0092)	(0.0024)	(0.0109)	(0.0050)
Baseline Means		55.3%	14.5%	0.6%	17.8%	2.8%	20.2%	1.1%	33.4%	4.8%
Eligibility for Florida Academic Scholars (FAS)										
1270 SAT Threshold										
Reduced Form		-0.0075	0.0014	0.0008	0.0027	0.0004	0.0031	0.0006	-0.0066	-0.0034
		(0.0061)	(0.0034)	(0.0011)	(0.0070)	(0.0051)	(0.0040)	(0.0013)	(0.0073)	(0.0056)
	First Stage									
Instrumental Variable	0.2837**	-0.0264	0.0049	0.0027	0.0096	0.0013	0.0110	0.0021	-0.0233	-0.0119
	(0.0055)	(0.0216)	(0.0120)	(0.0037)	(0.0246)	(0.0179)	(0.0141)	(0.0045)	(0.0258)	(0.0198)
Baseline Means		72.4%	5.4%	0.4%	32.7%	12.7%	7.5%	0.7%	49.4%	16.3%

Notes: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. Reduced form estimates are regression discontinuity treatment effects using initial SAT score as the running variable, linear slopes, and a bandwidth of 60 SAT points. Instrumental variable estimates are treatment effects using initial SAT score to predict eligibility based on an individual ever scoring above the relevant SAT threshold. Sample includes all of Florida's SAT takers who graduated high school from 2004 through 2012. Sample sizes at the 970/980 and 1270 threshold are 220,245 and 71,201 observations, respectively. Baseline means include all observations whose initial SAT score was 10 or 20 points below the eligibility threshold.

Table 4. Instrumental variable impacts of Bright Futures on selected college attendance and degree attainment outcomes, by SES terciles.

				Attendance				6-Year Degree Attainment			
		N	Attend College	In-State, Two-Year	In-State, Four-Year, Public	In-State, Four-Year, Private	Out of State	Associate, In-State	Associate, Out-State	Bachelor, In-State	Bachelor, Out-State
FMS: 970/980 Threshold	First-Stage										
SES Tercile 1 (Low)	0.4646**	36468	-0.0025	-0.0030	0.0114	-0.0010	-0.0100	-0.0459*	-0.0042	-0.0009	-0.0008
	(0.0080)		(0.0132)	(0.0224)	(0.0216)	(0.0108)	(0.0130)	(0.0182)	(0.0046)	(0.0207)	(0.0080)
Baseline Means			90.8%	41.9%	33.8%	5.9%	9.2%	21.4%	1.1%	29.3%	3.1%
SES Tercile 2 (Middle)	0.4062**	69571	-0.0044	-0.0166	0.0219	0.0040	-0.0137	-0.0336*	0.0017	0.0304+	-0.0077
	(0.0059)		(0.0104)	(0.0186)	(0.0179)	(0.0085)	(0.0110)	(0.0153)	(0.0037)	(0.0174)	(0.0073)
			91.8%	43.7%	33.6%	5.1%	9.4%	22.0%	0.9%	30.7%	4.1%
SES Tercile 3 (High)	0.3138**	110877	-0.0059	-0.0079	-0.0071	0.0071	0.0019	-0.0015	0.0009	0.0051	0.0114
	(0.0047)		(0.0099)	(0.0185)	(0.0191)	(0.0088)	(0.0120)	(0.0150)	(0.0039)	(0.0185)	(0.0089)
			92.9%	36.4%	40.5%	5.3%	10.7%	18.9%	1.0%	37.0%	5.5%
FAS: 1270 Threshold											
SES Tercile 1 (Low)	0.4098**	7005	0.0073	-0.0076	0.0066	-0.0391	0.0474	0.0038	-0.0051	-0.0535	-0.0175
	(0.0180)		(0.0276)	(0.0423)	(0.0592)	(0.0317)	(0.0442)	(0.0359)	(0.0118)	(0.0577)	(0.0375)
Baseline Means			94.0%	15.3%	54.8%	8.6%	15.4%	9.8%	1.1%	46.7%	11.0%
SES Tercile 2 (Middle)	0.3339**	18741	0.0025	-0.0003	-0.0697	0.0254	0.0470	-0.0178	0.0085	-0.0537	0.0305
	(0.0110)		(0.0183)	(0.0299)	(0.0443)	(0.0240)	(0.0337)	(0.0259)	(0.0078)	(0.0429)	(0.0293)
			95.5%	13.8%	57.4%	7.3%	16.9%	9.8%	0.7%	48.7%	12.2%
SES Tercile 3 (High)	0.2393**	44408	0.0078	-0.0029	0.0004	0.0338	-0.0235	0.0261	0.0004	-0.0012	-0.0259
	(0.0068)		(0.0154)	(0.0210)	(0.0396)	(0.0206)	(0.0343)	(0.0194)	(0.0063)	(0.0385)	(0.0311)
			96.2%	7.8%	58.3%	6.7%	23.4%	6.2%	0.6%	50.8%	18.4%

Notes: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. Estimates are treatment effects using initial SAT score as an instrumental variable to predict eligibility based on any SAT score exceeding a threshold. Sample includes all of Florida's SAT takers who graduated high school from 2004 through 2012. Baseline means include all observations whose initial SAT score was 10 or 20 points below the eligibility threshold.

Table 5. Instrumental variable impacts of Bright Futures on selected college attendance and degree attainment outcomes, by ethnicity.

				Attendance				6-Year Degree Attainment			
		N	Attend College	In-State, Two-Year	In-State, Four-Year, Public	In-State, Four-Year, Private	Out of State	Associate, In-State	Associate, Out-State	Bachelor, In-State	Bachelor, Out-State
FMS: 970/980 Threshold	First-Stage										
White	0.3672**	120645	0.0000	-0.0091	0.0035	0.0102	-0.0046	-0.0203	0.0030	0.0210	-0.0025
	(0.0046)		(0.0086)	(0.0156)	(0.0152)	(0.0073)	(0.0099)	(0.0126)	(0.0033)	(0.0149)	(0.0072)
Baseline Means			92.1%	41.6%	34.4%	5.3%	10.8%	20.6%	1.1%	33.4%	5.3%
Hispanic	0.3688**	51626	-0.0009	-0.0390+	0.0450+	-0.0008	-0.0060	-0.0230	-0.0027	0.0264	0.0016
	(0.0068)		(0.0131)	(0.0233)	(0.0232)	(0.0105)	(0.0134)	(0.0196)	(0.0046)	(0.0226)	(0.0093)
			92.0%	41.1%	37.7%	4.9%	8.4%	22.2%	1.0%	33.6%	3.8%
Black	0.4234**	26438	-0.0008	-0.0044	0.0300	-0.0045	-0.0219	-0.0079	-0.0069	0.0119	0.0079
	(0.0091)		(0.0141)	(0.0262)	(0.0289)	(0.0143)	(0.0192)	(0.0203)	(0.0059)	(0.0271)	(0.0127)
			93.7%	29.0%	45.6%	6.4%	12.8%	14.3%	1.1%	32.9%	4.9%
Asian	0.2490**	9228	-0.1065*	0.0712	-0.0885	-0.0691+	-0.0201	-0.0306	-0.0045	-0.0503	0.0220
	(0.0159)		(0.0485)	(0.0813)	(0.0858)	(0.0375)	(0.0479)	(0.0682)	(0.0139)	(0.0831)	(0.0330)
			91.9%	32.8%	44.5%	5.5%	9.1%	21.4%	0.7%	39.7%	4.0%
FAS: 1270 Threshold											
White	0.2925**	49471	0.0056	-0.0005	-0.0284	0.0381*	-0.0036	0.0007	-0.0005	0.0017	-0.0284
	(0.0067)		(0.0122)	(0.0192)	(0.0308)	(0.0156)	(0.0250)	(0.0167)	(0.0051)	(0.0298)	(0.0222)
Baseline Means			96.0%	11.2%	59.0%	6.1%	19.7%	7.8%	0.7%	50.2%	15.1%
Hispanic	0.2687**	11439	0.0347	0.0505	-0.0444	-0.0462	0.0748	0.0462	0.0263*	-0.1324+	0.0686
	(0.0137)		(0.0292)	(0.0376)	(0.0709)	(0.0423)	(0.0626)	(0.0360)	(0.0119)	(0.0697)	(0.0574)
			95.1%	7.5%	51.8%	10.8%	25.1%	7.0%	0.4%	48.5%	19.8%
Black	0.3489**	2568	0.0128	0.0447	-0.0418	0.0397	-0.0297	0.0632	-0.0178	-0.1503	0.0465
	(0.0284)		(0.0455)	(0.0568)	(0.1144)	(0.0607)	(0.1051)	(0.0531)	(0.0273)	(0.1120)	(0.0944)
			95.8%	6.7%	53.8%	7.1%	28.2%	5.9%	1.7%	43.9%	21.3%
Asian	0.1651**	4544	-0.0687	-0.0595	0.0513	-0.0180	-0.0425	0.0400	-0.0162	0.2812	-0.0599
	(0.0207)		(0.0826)	(0.0775)	(0.1785)	(0.0955)	(0.1525)	(0.0841)	(0.0262)	(0.1750)	(0.1381)
			94.9%	5.5%	60.0%	7.3%	22.1%	5.1%	0.5%	51.5%	17.3%

Notes: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. Estimates are treatment effects using initial SAT score as an instrumental variable to predict eligibility based on any SAT score exceeding a threshold. Sample includes all of Florida's SAT takers who graduated high school from 2004 through 2012. Baseline means include all observations whose initial SAT score was 10 or 20 points below the eligibility threshold.

Table 6. Biased regressions of impact of Bright Future on college attendance using maximum SAT score.

		In State				Out of State			
	Attend College	Any	Two-Year	Four-Year, Public	Four-Year, Private	Any	Two-Year	Four-Year, Public	Four-Year, Private
FMS: 970/980 Threshold									
Reduced Form	0.0069**	0.0157**	0.0002	0.0119**	0.0036+	-0.0088**	-0.0024+	-0.0037*	-0.0027
	(0.0025)	(0.0035)	(0.0044)	(0.0041)	(0.0020)	(0.0027)	(0.0013)	(0.0016)	(0.0019)
Baseline Means	91.0%	80.4%	46.3%	28.8%	5.3%	10.5%	2.1%	3.5%	4.9%
FAS: 1270 Threshold									
Reduced Form	0.0086**	0.0288**	-0.0132**	0.0421**	-0.0001	-0.0202**	-0.0001	-0.0033	-0.0168**
	(0.0026)	(0.0053)	(0.0040)	(0.0063)	(0.0034)	(0.0049)	(0.0010)	(0.0030)	(0.0041)
Baseline Means	95.6%	78.0%	11.3%	59.3%	7.4%	17.6%	0.6%	5.7%	11.3%

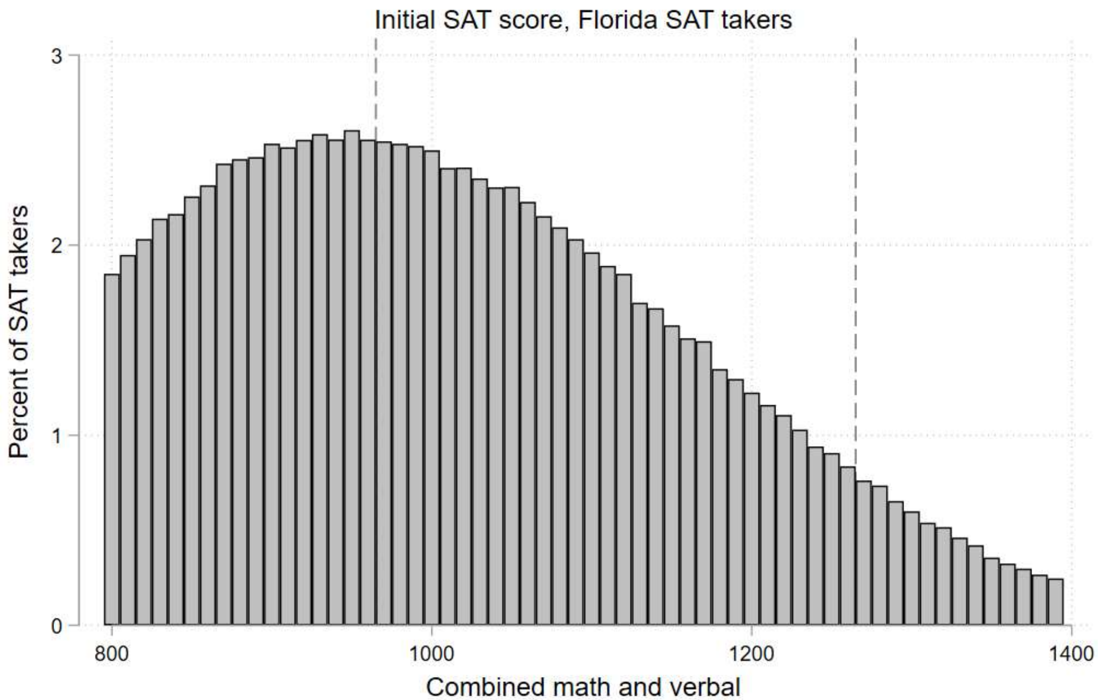
Notes: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. Estimates are regression discontinuity treatment effects using maximum SAT score as the running variable, linear slopes, and a bandwidth of 60 SAT points. The use of maximum SAT score leads to biased estimates due to endogenous retaking that leads to significant bunching (Figure 2), but is presented as a comparison to the method using in Zhang, et. al. (2016). Sample includes all of Florida's SAT takers who graduated high school from 2004 through 2012, and includes 210,661 observations at the FMS threshold and 97,339 observations at the FAS threshold. Baseline means include all observations whose initial SAT score was 10 or 20 points below the eligibility threshold.

Table 7. Replication of results based on Zhang subsample.

	Biased				Unbiased			
		In State				In State		
	N	Any	Two-Year	Four-Year, Public	N	Any	Two-Year	Four-Year, Public
FMS: 970/980 Threshold								
Reduced Form	51781	-0.0025	-0.0086	0.0061	53542	-0.0020	-0.0073	0.0053
		(0.0074)	(0.0090)	(0.0086)		(0.0071)	(0.0085)	(0.0087)
FAS: 1270 Threshold								
Reduced Form	23366	0.0096	- 0.0215**	0.0311*	17006	-0.0115	-0.0043	-0.0073
		(0.0112)	(0.0083)	(0.0126)		(0.0139)	(0.0094)	(0.0151)

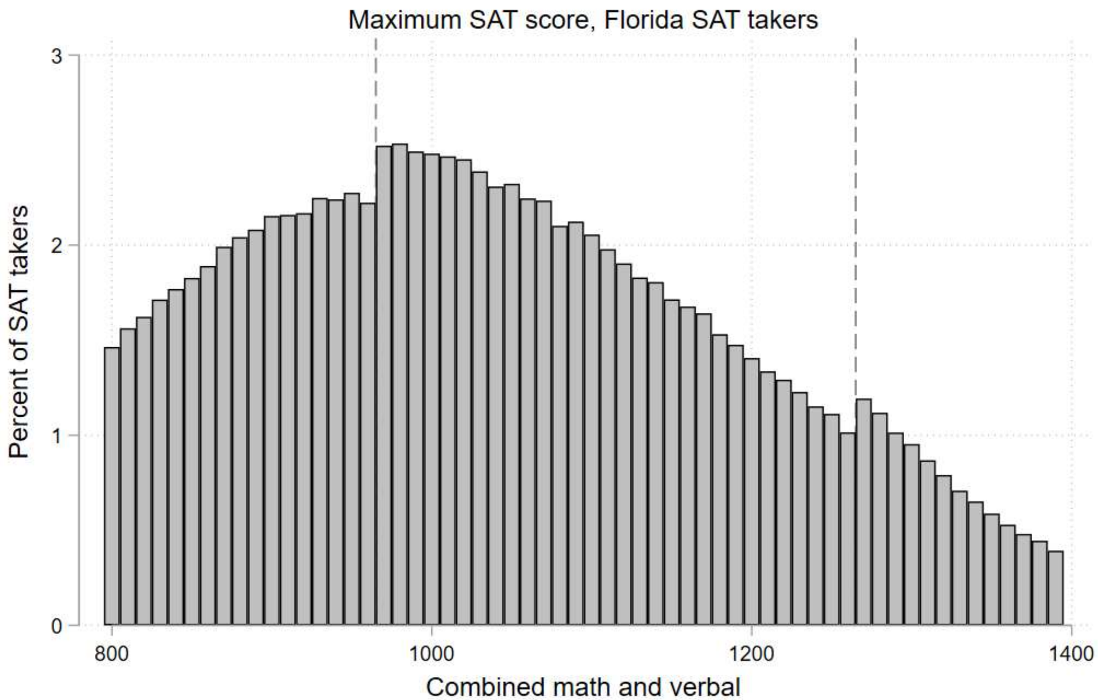
Notes: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. We mirror Zhang et al. (2016) by (1) restricting the sample to all Florida SAT takers who graduated high school from 2004 through 2006 and attended a public high school and (2) assume we cannot observe in-state, private college or any out of state college enrollment. Biased estimates are regression discontinuity treatment effects using the endogenous (incorrect) maximum SAT score as the running variable, linear slopes, and a bandwidth of 60 SAT points. Unbiased estimates are derived in the same manner except using initial SAT score as the running variable.

Figure 1. Density of observations, initial SAT.



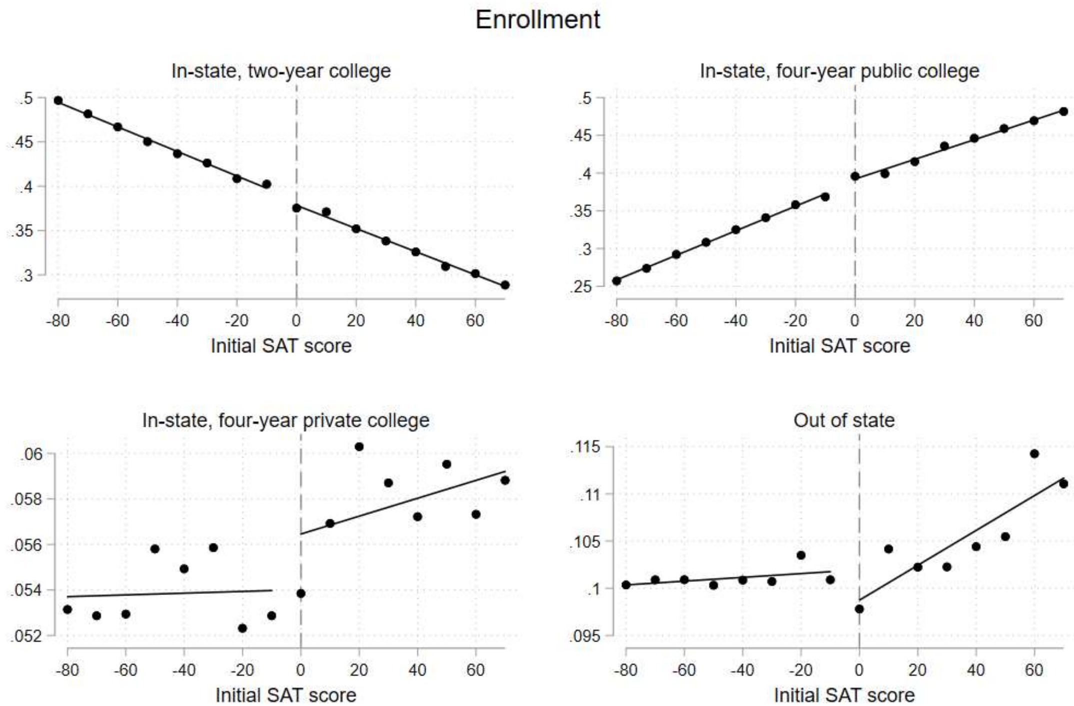
Notes: Sample includes all of Florida's SAT takers who graduated high school from 2004 through 2011. Dashed lines identify the 970 and 1270 eligibility thresholds. 2012 data was removed from figure due to changing SAT eligibility threshold.

Figure 2. Density of observations, maximum SAT.



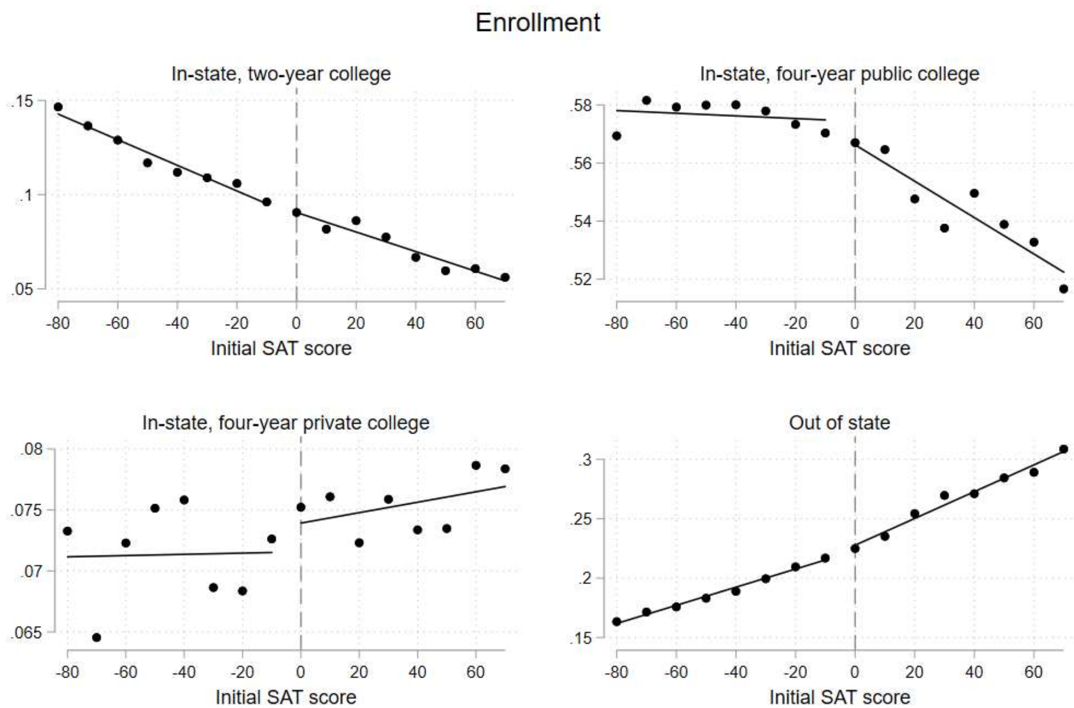
<u>Notes</u>: Sample includes all of Florida’s SAT takers who graduated high school from 2004 through 2012. Dashed lines identify the 970 and 1270 eligibility thresholds. 2012 data was removed from figure due to changing SAT eligibility threshold.

Figure 3. Enrollment outcomes for Florida Medallion Scholars, by initial SAT.



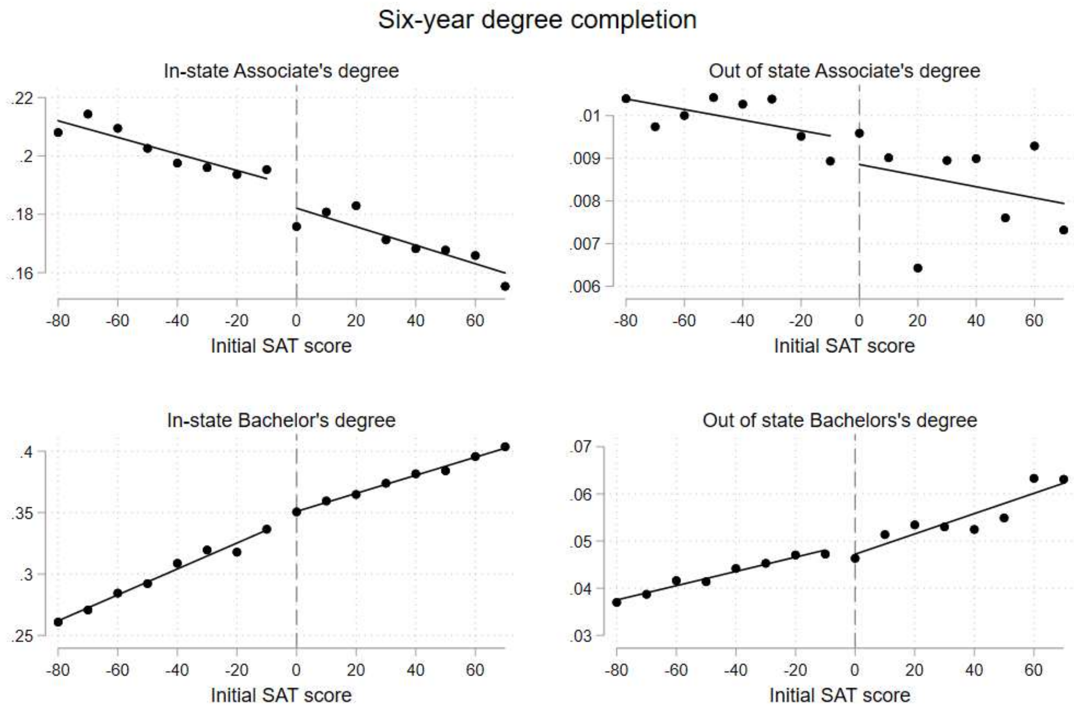
Notes: Plot of regression discontinuity effect at 970 threshold using initial SAT as the running variable, linear slopes, and a bandwidth of 80 SAT points. Sample includes all of Florida's SAT takers who graduated high school from 2004 through 2011. 2012 data was removed from figure due to changing SAT eligibility threshold. Corresponding regression results in Table 2.

Figure 4. Enrollment outcomes for Florida Academic Scholars, by initial SAT.



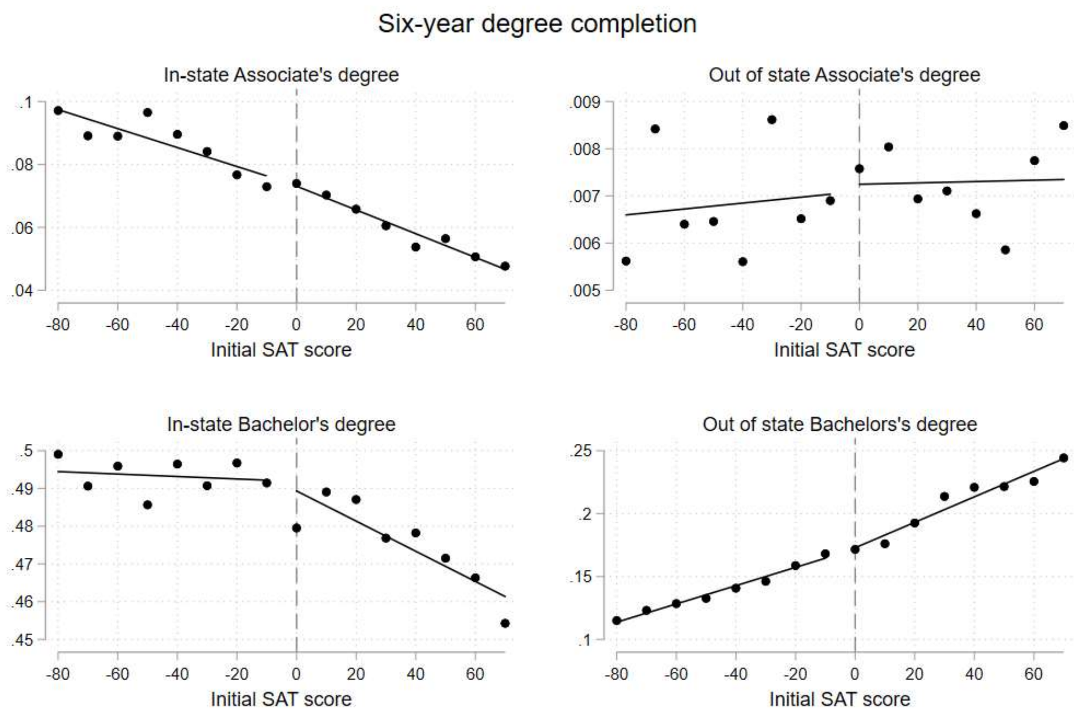
Notes: Plot of regression discontinuity effect at 1270 threshold using initial SAT as the running variable, linear slopes, and a bandwidth of 80 SAT points. Sample includes all of Florida's SAT takers who graduated high school from 2004 through 2012. Corresponding regression results in Table 2.

Figure 5. Degree completion outcomes for Florida Medallion Scholars, by initial SAT.



Notes: Plot of regression discontinuity effect at 970 threshold using initial SAT as the running variable, linear slopes, and a bandwidth of 80 SAT points. Sample includes all of Florida’s SAT takers who graduated high school from 2004 through 2011. 2012 data was removed from figure due to changing SAT eligibility threshold. Corresponding regression results in Table 3.

Figure 6. Degree completion outcomes for Florida Academic Scholars, by initial SAT.



Notes: Plot of regression discontinuity effect at 1270 threshold using initial SAT as the running variable, linear slopes, and a bandwidth of 80 SAT points. Sample includes all of Florida's SAT takers who graduated high school from 2004 through 2012. Corresponding regression results in Table 3.

ONLINE ONLY APPENDICES

Appendix Table 1. Balance tests, by specification.

	Female	Asian	African-American	Hispanic	White	Other Race	Took 11th Grade PSAT	11th Grade PSAT score	Took 10th Grade PSAT	10th Grade PSAT score	SAT Writing Score
FMS: 970/980 Threshold											
Linear	-0.0054	0.0014	0.0020	0.0009	-0.0035	-0.0007	-0.0048	-0.2893+	0.0027	-0.1324	-0.0493
	(0.0043)	(0.0017)	(0.0028)	(0.0037)	(0.0043)	(0.0020)	(0.0043)	(0.1503)	(0.0039)	(0.1209)	(0.5425)
Quadratic	-0.0078	0.0043	-0.0063	0.0003	0.0071	-0.0055	-0.0054	-0.0758	0.0095	-0.3173	-0.4736
	(0.0073)	(0.0029)	(0.0049)	(0.0063)	(0.0073)	(0.0034)	(0.0073)	(0.2578)	(0.0066)	(0.2058)	(0.9196)
N	220245	220245	220245	220245	220245	220245	220245	98353	220245	158606	170420
FAS: 1270 Threshold											
Linear	0.0061	0.0062+	-0.0000	-0.0020	-0.0008	-0.0034	0.0039	0.3925+	0.0023	0.5195*	-0.4007
	(0.0076)	(0.0037)	(0.0028)	(0.0056)	(0.0070)	(0.0032)	(0.0066)	(0.2128)	(0.0059)	(0.2093)	(1.0677)
Quadratic	0.0164	0.0011	-0.0062	0.0059	0.0052	-0.0059	0.0032	0.8353*	0.0061	0.8164*	-0.9298
	(0.0126)	(0.0062)	(0.0048)	(0.0092)	(0.0117)	(0.0053)	(0.0111)	(0.3584)	(0.0099)	(0.3500)	(1.7761)
N	71201	71201	71201	71201	71201	71201	71201	52810	71201	58112	53647

Notes: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. Regressions test for differences in observable characteristics between just-eligible and just-ineligible students. Includes all of Florida's SAT takers who graduated high school from 2004 through 2012 with initial SAT within 60 points of an eligibility threshold.

Appendix Table 2. Robustness tests (bandwidth and specification): Reduced-form impacts on in-state, two-year college attendance.

Bandwidth	30	40	50	60	70	80	90	100
FMS: 970/980 Threshold								
N	111457	148300	184443	220245	255654	290406	325153	358847
Linear	-0.0075	-0.0065	-0.0058	-0.0043	-0.0042	-0.0039	-0.0049	-0.0058+
	(0.0063)	(0.0053)	(0.0046)	(0.0042)	(0.0039)	(0.0036)	(0.0034)	(0.0032)
Linear with Covariates	-0.0230	-0.0123	-0.0100	-0.0095	-0.0081	-0.0072	-0.0051	-0.0036
	(0.0159)	(0.0105)	(0.0084)	(0.0072)	(0.0064)	(0.0058)	(0.0054)	(0.0050)
Quadratic	-0.0097	-0.0079	-0.0067	-0.0050	-0.0052	-0.0049	-0.0056+	-0.0062*
	(0.0062)	(0.0051)	(0.0045)	(0.0041)	(0.0038)	(0.0035)	(0.0033)	(0.0031)
Quadratic with Covariates	-0.0273+	-0.0163	-0.0132	-0.0119+	-0.0093	-0.0087	-0.0071	-0.0057
	(0.0155)	(0.0103)	(0.0082)	(0.0070)	(0.0062)	(0.0057)	(0.0053)	(0.0049)
FAS: 1270 Threshold								
N	35179	47032	59008	71201	83466	95958	108358	121599
Linear	-0.0024	-0.0042	-0.0033	0.0003	0.0013	0.0027	0.0031	0.0040
	(0.0068)	(0.0056)	(0.0049)	(0.0045)	(0.0041)	(0.0039)	(0.0036)	(0.0035)
Linear with Covariates	0.0112	0.0050	-0.0006	-0.0065	-0.0058	-0.0056	-0.0035	-0.0028
	(0.0169)	(0.0111)	(0.0088)	(0.0076)	(0.0067)	(0.0062)	(0.0057)	(0.0054)
Quadratic	-0.0019	-0.0032	-0.0023	0.0011	0.0026	0.0039	0.0049	0.0057+
	(0.0065)	(0.0054)	(0.0048)	(0.0043)	(0.0040)	(0.0038)	(0.0035)	(0.0034)
Quadratic with Covariates	0.0051	0.0029	-0.0008	-0.0060	-0.0054	-0.0044	-0.0032	-0.0023
	(0.0164)	(0.0108)	(0.0085)	(0.0073)	(0.0065)	(0.0060)	(0.0056)	(0.0052)

Notes: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. Estimates are regression discontinuity treatment effects using initial SAT by bandwidth and specification. Sample includes all of Florida's SAT takers who graduated high school from 2004 through 2012.

Appendix Table 3. Robustness tests (bandwidth and specification): Reduced-form impacts on in-state, four-year college attendance.

Bandwidth	30	40	50	60	70	80	90	100
FMS: 970/980 Threshold								
N	111457	148300	184443	220245	255654	290406	325153	358847
Linear	0.0092	0.0070	0.0065	0.0051	0.0048	0.0043	0.0054	0.0062+
	(0.0064)	(0.0054)	(0.0047)	(0.0043)	(0.0039)	(0.0037)	(0.0034)	(0.0033)
Linear with Covariates	0.0121	0.0125	0.0110	0.0105	0.0096	0.0088	0.0065	0.0046
	(0.0161)	(0.0106)	(0.0085)	(0.0072)	(0.0064)	(0.0059)	(0.0054)	(0.0051)
Quadratic	0.0114+	0.0082	0.0074	0.0057	0.0058	0.0055	0.0063+	0.0065*
	(0.0062)	(0.0052)	(0.0046)	(0.0041)	(0.0038)	(0.0035)	(0.0033)	(0.0032)
Quadratic with Covariates	0.0159	0.0168	0.0138+	0.0128+	0.0106+	0.0101+	0.0082	0.0068
	(0.0155)	(0.0103)	(0.0082)	(0.0070)	(0.0062)	(0.0057)	(0.0053)	(0.0049)
FAS: 1270 Threshold								
N	35179	47032	59008	71201	83466	95958	108358	121599
Linear	0.0052	0.0096	0.0049	0.0008	-0.0042	-0.0066	-0.0088	-0.0105+
	(0.0110)	(0.0091)	(0.0080)	(0.0073)	(0.0067)	(0.0063)	(0.0059)	(0.0056)
Linear with Covariates	-0.0080	-0.0056	0.0068	0.0115	0.0138	0.0114	0.0089	0.0061
	(0.0270)	(0.0179)	(0.0142)	(0.0121)	(0.0108)	(0.0099)	(0.0092)	(0.0086)
Quadratic	0.0033	0.0076	0.0038	-0.0004	-0.0046	-0.0066	-0.0088	-0.0103+
	(0.0109)	(0.0091)	(0.0080)	(0.0072)	(0.0067)	(0.0062)	(0.0059)	(0.0056)
Quadratic with Covariates	-0.0057	-0.0061	0.0043	0.0099	0.0114	0.0093	0.0076	0.0049
	(0.0268)	(0.0177)	(0.0141)	(0.0120)	(0.0107)	(0.0098)	(0.0091)	(0.0085)

Notes: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. Estimates are regression discontinuity treatment effects using initial SAT by bandwidth and specification. Sample includes all of Florida's SAT takers who graduated high school from 2004 through 2012.

Appendix Table 4. Robustness tests (bandwidth and specification): Reduced-form impacts on out-of-state college attendance.

Bandwidth	30	40	50	60	70	80	90	100
FMS: 970/980 Threshold								
N	111457	148300	184443	220245	255654	290406	325153	358847
Linear	-0.0021	-0.0033	-0.0032	-0.0025	-0.0035	-0.0030	-0.0023	-0.0023
	(0.0039)	(0.0033)	(0.0029)	(0.0026)	(0.0024)	(0.0023)	(0.0021)	(0.0020)
Linear with Covariates	0.0085	0.0022	-0.0012	-0.0030	-0.0022	-0.0035	-0.0042	-0.0035
	(0.0100)	(0.0065)	(0.0052)	(0.0044)	(0.0040)	(0.0036)	(0.0034)	(0.0032)
Quadratic	-0.0017	-0.0029	-0.0030	-0.0023	-0.0033	-0.0028	-0.0022	-0.0022
	(0.0039)	(0.0033)	(0.0029)	(0.0026)	(0.0024)	(0.0022)	(0.0021)	(0.0020)
Quadratic with Covariates	0.0094	0.0028	-0.0004	-0.0027	-0.0018	-0.0032	-0.0036	-0.0031
	(0.0099)	(0.0065)	(0.0052)	(0.0044)	(0.0039)	(0.0036)	(0.0033)	(0.0031)
FAS: 1270 Threshold								
N	35179	47032	59008	71201	83466	95958	108358	121599
Linear	-0.0025	-0.0041	-0.0007	0.0011	0.0043	0.0047	0.0055	0.0068
	(0.0094)	(0.0079)	(0.0069)	(0.0063)	(0.0058)	(0.0054)	(0.0051)	(0.0049)
Linear with Covariates	0.0031	0.0013	-0.0046	-0.0049	-0.0062	-0.0033	-0.0024	-0.0017
	(0.0231)	(0.0153)	(0.0121)	(0.0104)	(0.0093)	(0.0085)	(0.0079)	(0.0074)
Quadratic	-0.0018	-0.0037	-0.0008	0.0013	0.0032	0.0036	0.0037	0.0047
	(0.0092)	(0.0077)	(0.0068)	(0.0061)	(0.0057)	(0.0053)	(0.0050)	(0.0048)
Quadratic with Covariates	0.0082	0.0036	-0.0027	-0.0043	-0.0046	-0.0030	-0.0017	-0.0012
	(0.0225)	(0.0149)	(0.0118)	(0.0101)	(0.0090)	(0.0082)	(0.0077)	(0.0072)

Notes: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. Estimates are regression discontinuity treatment effects using initial SAT by bandwidth and specification. Sample includes all of Florida's SAT takers who graduated high school from 2004 through 2012.

Appendix Table 5. Robustness tests (bandwidth and specification): Reduced-form impacts on six-year, in-state, associate degree attainment.

Bandwidth	30	40	50	60	70	80	90	100
FMS: 970/980 Threshold								
N	111457	148300	184443	220245	255654	290406	325153	358847
Linear	-0.0172**	-0.0127**	-0.0098**	-0.0080*	-0.0067*	-0.0069*	-0.0070*	-0.0062*
	(0.0051)	(0.0043)	(0.0038)	(0.0034)	(0.0031)	(0.0029)	(0.0027)	(0.0026)
Linear with Covariates	-0.0302*	-0.0236**	-0.0206**	-0.0183**	-0.0156**	-0.0112*	-0.0088*	-0.0095*
	(0.0129)	(0.0085)	(0.0068)	(0.0058)	(0.0052)	(0.0047)	(0.0044)	(0.0041)
Quadratic	-0.0179**	-0.0133**	-0.0101**	-0.0081*	-0.0069*	-0.0071*	-0.0071**	-0.0062*
	(0.0051)	(0.0042)	(0.0037)	(0.0034)	(0.0031)	(0.0029)	(0.0027)	(0.0026)
Quadratic with Covariates	-0.0320*	-0.0248**	-0.0218**	-0.0193**	-0.0161**	-0.0117*	-0.0096*	-0.0102*
	(0.0128)	(0.0085)	(0.0068)	(0.0058)	(0.0051)	(0.0047)	(0.0043)	(0.0041)
FAS: 1270 Threshold								
N	35179	47032	59008	71201	83466	95958	108358	121599
Linear	0.0082	0.0084+	0.0092*	0.0031	-0.0004	-0.0006	-0.0011	0.0004
	(0.0060)	(0.0050)	(0.0044)	(0.0040)	(0.0037)	(0.0034)	(0.0033)	(0.0031)
Linear with Covariates	0.0030	0.0070	0.0070	0.0159*	0.0154*	0.0096+	0.0062	0.0019
	(0.0148)	(0.0099)	(0.0079)	(0.0067)	(0.0060)	(0.0055)	(0.0051)	(0.0047)
Quadratic	0.0083	0.0086+	0.0094*	0.0034	0.0001	-0.0001	-0.0002	0.0012
	(0.0059)	(0.0050)	(0.0044)	(0.0039)	(0.0036)	(0.0034)	(0.0032)	(0.0031)
Quadratic with Covariates	-0.0008	0.0058	0.0066	0.0158*	0.0152*	0.0097+	0.0060	0.0020
	(0.0146)	(0.0098)	(0.0078)	(0.0066)	(0.0059)	(0.0054)	(0.0050)	(0.0047)

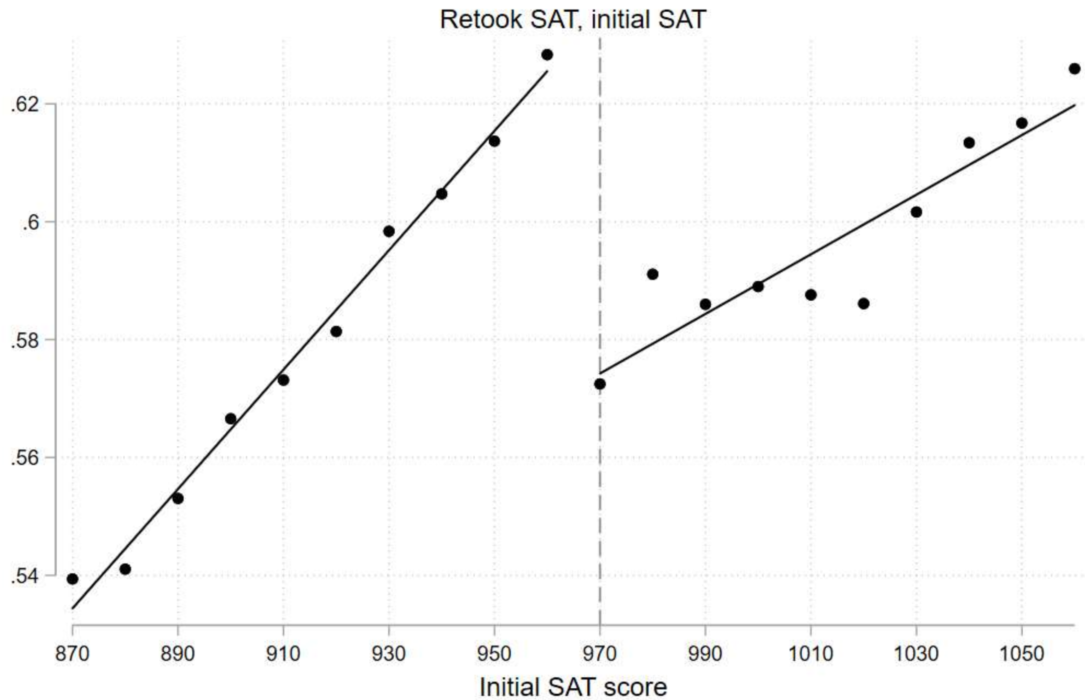
Notes: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. Estimates are regression discontinuity treatment effects using initial SAT by bandwidth and specification. Sample includes all of Florida's SAT takers who graduated high school from 2004 through 2012.

Appendix Table 6. Robustness tests (bandwidth and specification): Reduced-form impacts on six-year, in-state, bachelor's degree attainment.

Bandwidth	30	40	50	60	70	80	90	100
FMS: 970/980 Threshold								
N	111457	148300	184443	220245	255654	290406	325153	358847
Linear	0.0031	0.0055	0.0032	0.0054	0.0031	0.0022	0.0033	0.0031
	(0.0061)	(0.0051)	(0.0045)	(0.0040)	(0.0037)	(0.0035)	(0.0033)	(0.0031)
Linear with Covariates	-0.0294+	-0.0090	0.0032	0.0008	0.0057	0.0060	0.0033	0.0027
	(0.0152)	(0.0101)	(0.0080)	(0.0069)	(0.0061)	(0.0056)	(0.0052)	(0.0048)
Quadratic	0.0051	0.0063	0.0044	0.0064	0.0044	0.0035	0.0045	0.0039
	(0.0060)	(0.0050)	(0.0044)	(0.0039)	(0.0036)	(0.0034)	(0.0032)	(0.0030)
Quadratic with Covariates	-0.0266+	-0.0050	0.0048	0.0024	0.0067	0.0074	0.0048	0.0047
	(0.0148)	(0.0098)	(0.0078)	(0.0067)	(0.0059)	(0.0054)	(0.0050)	(0.0047)
FAS: 1270 Threshold								
N	35179	47032	59008	71201	83466	95958	108358	121599
Linear	-0.0083	-0.0052	-0.0092	-0.0066	-0.0069	-0.0036	-0.0057	-0.0065
	(0.0110)	(0.0092)	(0.0081)	(0.0073)	(0.0067)	(0.0063)	(0.0059)	(0.0056)
Linear with Covariates	0.0151	-0.0046	0.0005	-0.0081	-0.0070	-0.0110	-0.0058	-0.0043
	(0.0271)	(0.0179)	(0.0142)	(0.0122)	(0.0109)	(0.0099)	(0.0092)	(0.0086)
Quadratic	-0.0100	-0.0077	-0.0106	-0.0078	-0.0080	-0.0043	-0.0067	-0.0071
	(0.0109)	(0.0091)	(0.0080)	(0.0072)	(0.0067)	(0.0062)	(0.0059)	(0.0056)
Quadratic with Covariates	0.0215	-0.0032	-0.0015	-0.0103	-0.0091	-0.0132	-0.0070	-0.0058
	(0.0268)	(0.0178)	(0.0141)	(0.0121)	(0.0108)	(0.0098)	(0.0091)	(0.0086)

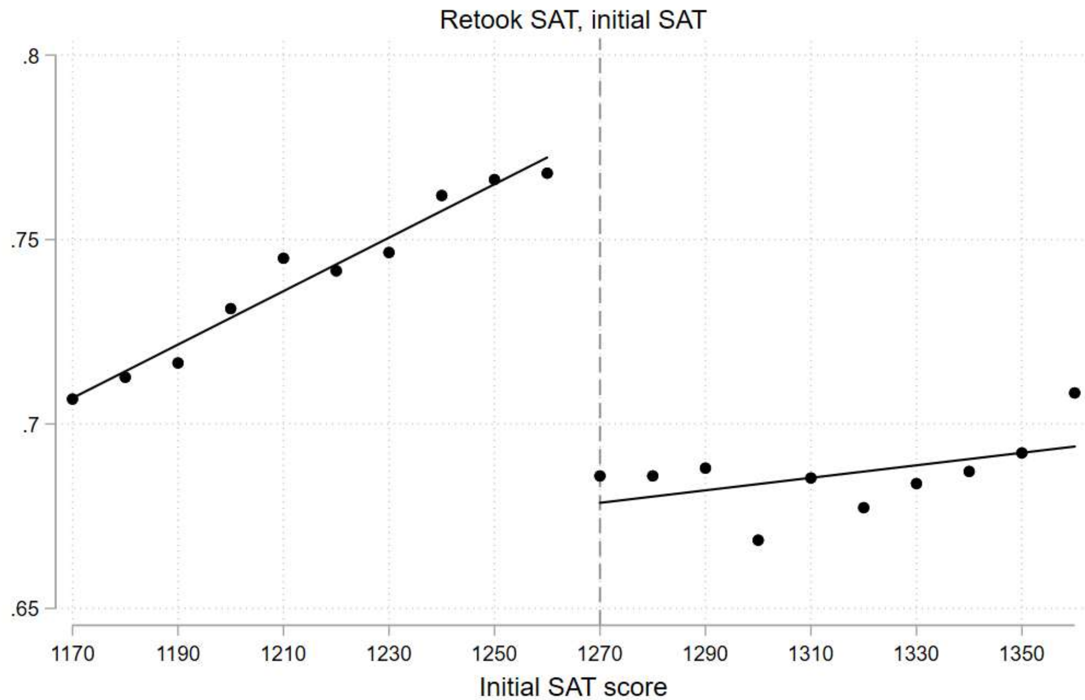
Notes: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. Estimates are regression discontinuity treatment effects using initial SAT by bandwidth and specification. Sample includes all of Florida's SAT takers who graduated high school from 2004 through 2012.

Appendix Figure 1. Retaking at 970 (FMS) threshold.



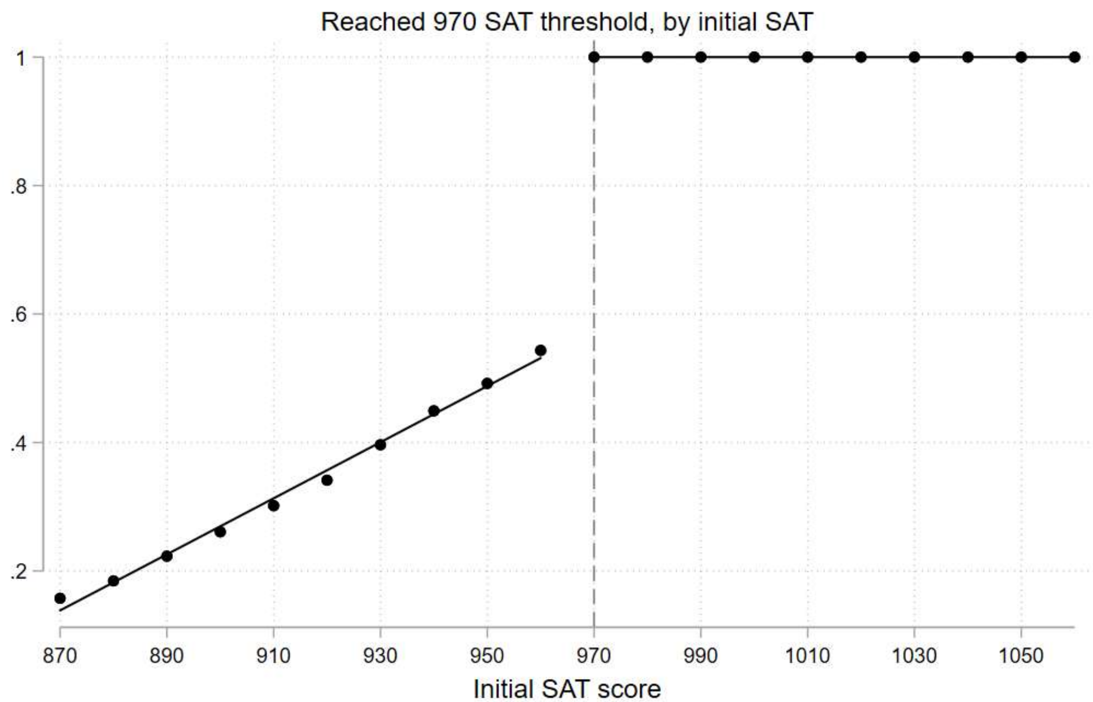
Notes: Plot is probability of SAT test retaking behavior at initial SAT test score using linear slopes and 60 SAT point bandwidth. Sample includes all of Florida's SAT takers who graduated high school from 2004 through 2011. 2012 data was removed from figure due to changing SAT eligibility threshold.

Appendix Figure 2. Retaking at 1270 (FAS) threshold.



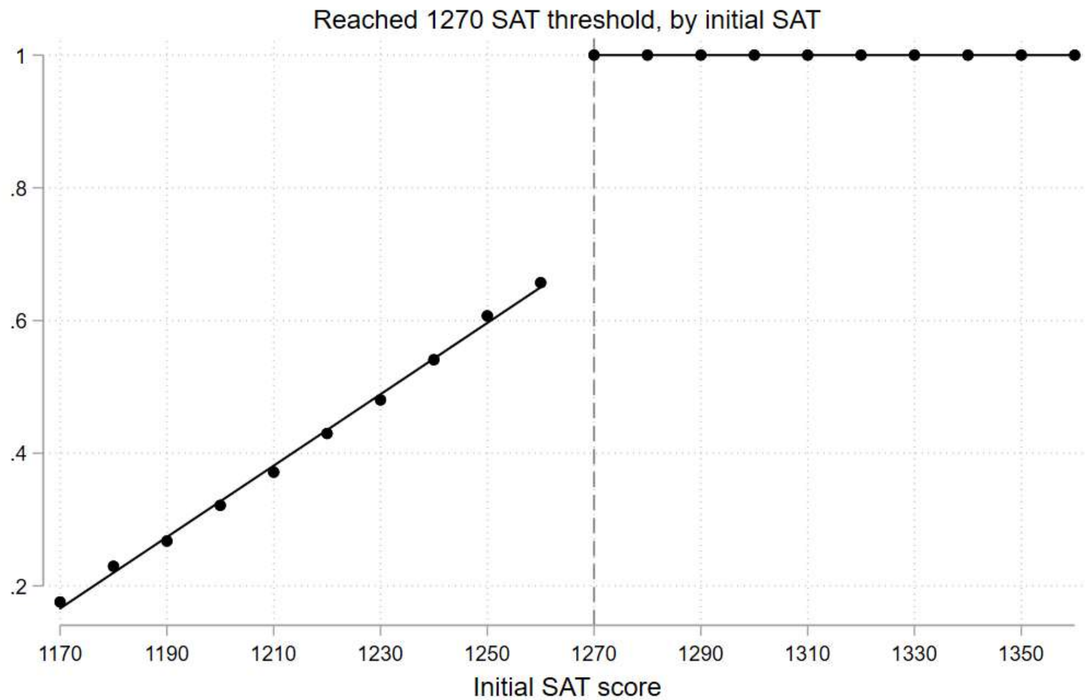
Notes: Plot is probability of SAT test retaking behavior at initial SAT test score using linear slopes and 60 SAT point bandwidth. Sample includes all of Florida's SAT takers who graduated high school from 2004 through 2012.

Appendix Figure 3. Ever surpassed 970 (FMS) threshold, by initial SAT.



Notes: Plot is probability of ever exceeding eligibility threshold at initial SAT test score using linear slopes and 60 SAT point bandwidth. Sample includes all of Florida's SAT takers who graduated high school from 2004 through 2011. 2012 data was removed from figure due to changing SAT eligibility threshold.

Appendix Figure 4. Ever surpassed 1270 (FAS) threshold, by initial SAT.



<u>Notes</u>: Plot is probability of ever exceeding eligibility threshold at initial SAT test score Using linear slopes and 60 SAT point bandwidth. Sample includes all of Florida's SAT takers who graduated high school from 2004 through 2012.