Optional ERIC Coversheet — Only for Use with U.S. Department of Education Grantee Submissions

This coversheet should be completed by grantees and added to the PDF of your submission if the information required in this form **is not included on the PDF to be submitted**.

INSTRUCTIONS

- Before beginning submission process, download this PDF coversheet if you will need to provide information not on the PDF.
- Fill in all fields—information in this form **must match** the information on the submitted PDF and add missing information.
- Attach completed coversheet to the PDF you will upload to ERIC [use Adobe Acrobat or other program to combine PDF files]—do not upload the coversheet as a separate document.
- Begin completing submission form at https://eric.ed.gov/submit/ and upload the full-text PDF with attached coversheet when indicated. Your full-text PDF will display in ERIC after the 12-month embargo period.

GRANTEE SUBMISSION REQUIRED FIELDS

Title of article, paper, or other content

Changing College Choices with Personalized Admissions Information at Scale: Evidence on Naviance

All author name(s) and affiliations on PDF. If more than 6 names, ERIC will complete the list from the submitted PDF.

Last Name, First Name	Academic/Organizational Affiliation	ORCID ID
Mulhern, Christine	Harvard University	0000-0002-7199-1900

Publication/Completion Date—(if *In Press,* enter year accepted or completed) 2021

Check type of content being submitted and complete one of the following in the box below:

- ☑ If article: Name of journal, volume, and issue number if available
- ☐ If paper: Name of conference, date of conference, and place of conference
- ☐ If book chapter: Title of book, page range, publisher name and location
- ☐ If book: Publisher name and location
- ☐ If dissertation: Name of institution, type of degree, and department granting degree

Journal of Labor Economics, 39, 1

DOI or URL to published work (if available) https://doi.org/10.1086/708839

Acknowledgement of Funding— Grantees should check with their grant officer for the preferred wording to acknowledge funding. If the grant officer does not have a preference, grantees can use this suggested wording (adjust wording if multiple grants are to be acknowledged). Fill in Department of Education funding office, grant number, and name of grant recipient institution or organization.

"This work was supported by U.S. Department of Education **[Office name]** Institute of Education Sciences through **[Grant number]** R305B150010 to **Institution]** Harvard University .The opinions expressed are those of the authors and do not represent views of the **[Office name]** Institute or the U.S. Department of Education.

Changing College Choices with Personalized Admissions Information at Scale: Evidence on Naviance

Christine Mulhern *
Harvard University
Mulhern@g.harvard.edu

March 2, 2020

Abstract

Choosing where to apply to college is a complex problem with long-term consequences, but many students lack the guidance necessary to make optimal choices. I show that a technology which provides low-cost personalized college admissions information to over forty percent of high schoolers significantly alters college choices. Students shift applications and attendance to colleges for which they can observe information on schoolmates' admissions experiences. Responses are largest when such information suggests a high admissions probability. Disadvantaged students respond the most, and information on in-state colleges increases their four-year college attendance. Data features and framing, however, deter students from selective colleges.

*I thank Christopher Avery, Joshua Goodman, Thomas Kane, Amanda Pallais, Eric Taylor, Carly Robinson, Rebecca Sachs, Johanna Brownell, and seminar participants at Harvard, UChicago Urban Labs, APPAM and AEFP for valuable feedback. I am grateful to the partner school district for providing data and guidance. The research reported here was supported, in part, by the Institute of Education Sciences, U.S. Department of Education, through grant R305B150010 for the Partnering in Education Research Fellowship in collaboration with the Center for Education Policy Research at Harvard University. The opinions expressed are those of the author and do not represent the views of the Institute or the U.S. Department of Education. All errors are my own.

1 Introduction

Choosing where to apply to college is a complex problem which many students struggle to navigate. In the U.S., students can choose from among more than 4,000 colleges, and traditionally disadvantaged students often lack information about the application process, admissions criteria, and benefits and costs of colleges (Avery & Kane, 2004; Hoxby & Avery, 2013; Hastings, Neilson & Zimmerman, 2015; Radford, 2013). Improving application choices is important because these choices have large impacts on college enrollment, degree attainment, and labor market outcomes (Hoxby & Avery, 2013; Chetty et al., 2017; Cohodes & Goodman, 2014; Smith, 2018; Zimmerman, 2014). This paper provides the first evidence on how a popular low-cost technology can change where students apply to and attend college by providing personalized admissions information.

Traditionally, students have gathered information about their college options and admissions probabilities from their social networks, school counselors, or general resources (Hoxby & Avery, 2013; Radford, 2013; Roderick et al., 2008). Many students lack social networks which can provide this type of information and thus have turned to these other resources or made uninformed choices (Hoxby & Avery, 2013). School counselors are well positioned to provide high touch personalized guidance, but there is considerable variation in their effectiveness, and their large caseloads may make it difficult to scale the high touch nature of their support (Hurwitz & Howell, 2014; Mulhern, 2019). General or online college resources, such as the College Scorecard, are more scalable solutions, but are not personalized.

The technology Naviance bridges these gaps by providing low-cost personalized college admissions information to over forty percent of U.S. high schoolers (Shellenbarger, 2017).[1] Naviance is an online platform that districts can purchase to help with college choices and counseling. It contains college and career search tools, basic college information, and a portal to contact counselors and request application materials. Schools are encouraged to introduce it to students in 9th or 10th grade so they can explore careers and the scores needed for college admission. Students access it more during 11th grade, when taking college entrance exams, and usage peaks during

[1] It is also used by students in over 100 countries. Naviance reports that more than 40% of high schoolers use the platform. The fraction who have access to it, through their school, may be higher. https://www.naviance.com/resources/entry/press-ann-arbor-public-schools-selects-naviance-to-increase-college-and-car

12th grade, when students choose where to apply, submit applications and enroll in college.

A primary feature of Naviance shows students how, for individual colleges, their academic profiles compare to previously admitted schoolmates. This information is conveyed in scattergrams, which are scatterplots showing the GPA and SAT (or ACT) scores of prior applicants from a student's school to a specific college, as well as the admissions decision each of these applicants received. Figure 1 shows an example. Scattergrams are visible for colleges which received at least five applications from a high school. A dashboard also summarizes how a student's scores compare to the average admitted student (from her high school) at each college the student has saved, and the student's scores are color-coded based on whether they are above or below the average admit. I examine how access to this admissions information, and the signals it sends about a student's probability of admission, impact where students apply to and attend college.

I study the college choices of students in a Mid-Atlantic school district, with 10-15 high schools and approximately 4,000 graduates per year, in the first three years students could access Naviance. The district purchased Naviance just before the 2013-2014 school year and first made scattergrams available at the end of the school year, when they had collected admissions data. These scattergrams were based on the class of 2014, and they were updated in June 2015 to also include data on the class of 2015. Thus, as 12th graders, the class of 2016 had access to a different set of scattergrams than the class of 2015. On average, students could access 47 scattergrams.

I examine how access to scattergrams, and the average acceptance criteria they convey, influence where students apply to and attend college. I identify the causal effects of access to admission information and perceived admissions probabilities using regression discontinuity designs and fixed effects approaches. These approaches exploit idiosyncratic variation across high schools and years in what students see. The paper contains four main findings.

First, access to a college's admissions information increases applications and attendance at that college, especially for students with a high admissions probability. I use a regression discontinuity design to show that students are 20% more likely to apply to a college if its information is visible than if it had too few datapoints to pass the visibility threshold. Gaining access to a college's admissions information has the largest impact on students most similar to previous admits, as well as Black, Hispanic, and low-income students. I also find larger effects for in-state public colleges, possibly because these are the most commonly viewed scattergrams, or because students are most interested in nearby and inexpensive colleges (Radford, 2013).

Second, students prefer to apply to colleges where they are most similar to previous admits. I use variation in the average admit's scores, across high schools and years, to measure how application choices vary based on the signals a student receives about her probability of admission. Application rates are decreasing in a student's distance from the average admit's scores. Thus, students prefer to apply where they have a reasonable chance of admission, but not where the signals indicate they can be accepted at a much more selective college.

Third, students use the average admissions lines, and the color-coding of their scores, as heuristics to simplify their application choices. Students just below the average admit's GPA are 8% less likely to apply to a college than students just above it. I find no discontinuity at the average SAT, possibly because there are many information sources for SAT admissions criteria. Students seem to interpret being below the mean GPA as a negative signal, which leads them to reduce the selectivity of the colleges they apply to and attend. Reactions are largest for students who can see the most scattergrams, possibly because they need the most help simplifying their choices.

Finally, the information in Naviance leads application portfolios and attendance choices to reflect the set of colleges with visible and relevant information.[2] The set of colleges to which students are nudged depends on which colleges were popular among previous cohorts and how accurately previous admits' scores reflect colleges' true admissions criteria. This approach increases college selectivity for some students but deters others from attending highly selective or *match* colleges.

Access to admissions information on local four-year colleges also increases four-year college enrollment rates for Black, Hispanic, and low-income students. This is driven by a shift from local community colleges to the state's many small public colleges, indicating that students may have been unaware of these nearby and inexpensive options with high admissions rates. This suggests potential for this sort of information to help close socioeconomic gaps in college enrollment and degree attainment. The current setup of Naviance, however, may also reduce degree attainment

[2]Relevant scattergrams are those where the student is within .5 GPA points and 150 SAT points of the average admit.

and future income by nudging some students to less selective colleges (Chetty et al. 2017; Cohodes & Goodman, 2014; Dillon & Smith, 2018; Goodman, Hurwitz & Smith, 2017; Zimmerman, 2014).

My findings indicate that students prefer to apply to colleges at which they have admissions information, and where they are likely to be admitted. My causal estimates are consistent with the changes I observe over time as scattergrams became available. Students are more likely to be accepted at the colleges to which they apply in the years they can see scattergrams than in the year without them. In addition, the first cohort with access to scattergrams is less likely to apply to *reach* colleges and more likely to attend a *safety* school than students in the previous cohort. These patterns, along with the causal estimates, are consistent with students updating their admissions beliefs when they receive more information, and shifting applications to increase acceptance probabilities. Students may place too much weight, however, on admissibility because some attend less selective colleges when they have this information.

This paper provides some of the first evidence on how information about admissions probabilities, based on GPAs and test scores, influences college application choices. Little empirical work explores how students choose which colleges to apply to when there are thousands of options and when the benefits and costs of colleges appear similar. Pallais (2009) shows that students may use rules of thumb to help simplify this choice and Bond et al. (2018) find that students apply to more selective colleges when their SAT score (and thus admissions probability) unexpectedly increases. I build on this work, and models of the application choice problem by Chade, Lewis and Smith (2014) and Fu (2014), by employing student data, and exogenous shocks to the availability and nature of admissions information, to empirically test how students use admissions information.

I also show that a popular technology can significantly change application choices with personalized information. This is consistent with prior work showing the importance of information provision in college choices, especially, and sometimes only, if it is personalized or accompanied by individual assistance (Barr & Turner, 2018; Bettinger, Long, Oreopoulos & Sanbonmatsu, 2012; Castleman & Page, 2015; Hoxby & Turner, 2015; Hurwitz & Smith, 2017; Luca & Smith, 2013). The information provision studied here is unique in that it is provided by a low-cost technology used by more than 40% of high schoolers, and it is based on students' peers.

Information may have large effects in the present setting because of its framing, focus on peers, and personalized nature. The lines noting the average scores of previously admitted students are very salient and create reference points that are easy for students to understand (Kahneman, 1992; Kahneman, 2003). Little work shows how the framing of information relates to its impact in education contexts (Lavecchia, Liu, & Oreopoulos, 2017). My findings are consistent with work showing that simplifying information has large effects on education choices, but I find some negative consequences from data framing (Hastings & Weinstein, 2008). Students may also respond strongly to the scattergram data because they are based on their peers. Students react to peer norms in other settings and they look to their peers for guidance in the college application process (Akerlof & Kranton, 2002; Bursztyn & Jensen, 2015; Radford, 2013). Information may also matter in this setting because of its personalized nature. Providing individualized guidance and encouragement has increased the effectiveness of other information interventions (Bettinger et al., 2012; Castleman & Page, 2015). My findings suggest that personalizing and disseminating information with technology could be a more cost-effective way to attain the impacts associated with personalized assistance. Many districts pay less than $10 per student for access to this technology.

Despite the rapid rise of education technologies, including many in the college choice space, there is little convincing evidence on how technology can transform education experiences (Escueta, Quan, Nickow & Oreopoulos, 2017; Shellenbarger, 2017; Shulman, 2018). Existing research indicates the potential for technology to improve students' choices and outcomes, but some technologies reduce student performance or exacerbate socioeconomic gaps (Bergman & Chan, 2019; Dettling, Goodman & Smith, 2018; Escueta et al., 2017; Hurwitz & Smith, 2018; Carter, Greenberg, & Walker, 2016). This paper provides some of the first evidence on how technology can help students with one of the most important decisions of their life, and how it can complement the counselor's role, enabling busy workers to more efficiently meet the needs of the individuals they support. It also provides evidence on one of the most widely adopted college choice technologies.

Naviance increases four-year college enrollment for low-income, Black, and Hispanic students when it provides them information about local public colleges where they are likely to be admitted. It also increases the selectivity of colleges attended by students who are shown information on many relevant match and reach colleges. Students who attend high schools with weaker college-going cultures, however, are more likely to be nudged to less selective colleges based on the available scattergrams. The reference points created by the average admit's GPA also deter students from applying to highly selective colleges. For a given school, the extent to which Naviance helps or hinders its students depends on the college-going culture of the school and how counselors implement the technology. I also find significant variation in Naviance's effects across counselors, which highlights the importance of the individuals implementing technologies.

The paper proceeds as follows. Section 2 describes Naviance, the data and setting, and changes over time. Section 3 examines how access to a college's admissions information changes applications and enrollment at that college. Section 4 describes how signals about one's admissions probabilities influence applications and attendance at individual colleges. Section 5 shows how the full set of information and admissibility signals influence the types of colleges to which students apply and attend. The implications and conclusions are discussed in section 6.

2 Naviance and Setting

2.1 Naviance

Naviance is a software that school districts can purchase to help track student progress and prepare students for postsecondary choices. It includes features to track student goals, course schedules, counselor meetings and graduation requirements. Students can also take quizzes to identify relevant careers and colleges, and see career and college statistics. Students can save colleges in which they are interested, and counselors or parents can log in and save colleges to a student's profile. In addition, Naviance can track the college application process, from requesting counselor recommendations and transcripts to submitting materials via an interface linked with the Common Application. Figure A.1 shows an example of the dashboard monitoring these steps.

Naviance provides a similar support package to each district that purchases it, with some variation depending on the district's needs and plans. At a minimum, the package includes a tutorial of basic features, school counselor training, guidance to provide students, and a district liaison. Counselors are encouraged to introduce and provide guidance on Naviance to students and parents in classes or after school sessions. There are also videos on how to use the platform.

One of the main and most novel features of Naviance is its scattergrams. Figure 1 shows an example. These are scatterplots for individual colleges which depict the standardized test scores and GPAs of previous applicants from a student's high school, as well as the admissions decision each applicant received. Lines on the scattergrams indicate the average GPA and SAT (or ACT) scores for all previously accepted students from the user's high school. I refer to these as the "typical acceptee" lines. These lines vary across high schools and over time, since they are updated every time a new cohort's admissions data are added to Naviance.

Students can easily see how they compare to prior applicants and these lines because Naviance displays a red circle on the scattergram marking the current user's scores. Naviance also contains a page summarizing the colleges a student has saved and how the student compares to the typical acceptee at each saved college (Figure A.2). The typical acceptee's scores are green if the current user's score is above the typical acceptee's and red if it is below. The typical acceptee lines are averages, not minimums, so roughly half of the students below them were accepted to the college. This framing, however, may make admissions seem unlikely for students just below the typical acceptee. Some media attention suggest that students may treat the lines as minimums more than averages and become discouraged (Drezner, 2017; Shellenbarger, 2017; Gelger, 2018).

Students can only see a scattergram if the high school has data on at least five applicants to that college in prior cohorts. Some high schools further restrict this to ten prior applicants. During the time studied, school administrators could select a minimum of five or ten in Naviance's settings. This means that students only see admissions data for colleges that were somewhat popular at their high school in the past. This may not be the optimal set of college information to provide students because it could perpetuate suboptimal college choices. It is, however, a simple way to identify colleges that may be a good fit, in terms of location or culture, for students in a school.

The data that students see are noisy indicators of their probability of admission. Many scattergrams only have a few datapoints and the typical acceptee lines may only be based on a couple of admitted students. Many schools and counselors, however, see value in the high school specific nature of the data. Some believe that college admissions consider where a student went to high school and apply different admissions criteria to students from different high schools.[3] This may be because course rigor varies across high schools or because schools use different GPA scales or grading criteria. The scattergrams offer a way for students to compare themselves to students who faced a similar academic environment. Furthermore, students may care more about the experiences of their peers, who are likely to be similar to them, than a national sample of students.

Over time, additional student data are loaded into Naviance which leads to changes in the scattergrams available and the typical acceptee lines. Schools can select how many prior cohorts' data are used to populate the scattergrams. If schools do not limit the cohorts available, the number of available scattergrams will continue to grow and the typical acceptee lines may become more stable and accurate. Student responses to the availability of scattergrams and typical acceptee lines will, however, impact what becomes visible to future cohorts.[4]

2.2 Setting and Data

I study the impact of Naviance in a medium-sized school district in a Mid-Atlantic state for students who graduated high school between 2014 and 2017. The timeline of the treatment, data available, and major steps in the college application process are in Figure 2. The district contains 10-15 high schools and approximately 4,000 high school graduates each year. My main sample consists of nearly 8,000 students who graduated a district high school in 2015 or 2016 and for whom I have essential data.[5] Descriptive statistics are in Tables 1 and A.1. The district is ethnically diverse, with 8% of students in my sample identifying as Hispanic, 20% Black, 17% Asian, and 49% white. 21% of the students received free or reduced-price lunch (FRPL) while in the district.

The district provided data on student demographics, coursework and grades, as well as test scores. I use racial groups and indicators of free and reduced-price lunch receipt as proxies for

[3]This is supported by surveys of college admissions professionals which indicate that the strength of a high school's curriculum is one of the most important factors in an admissions decision (Clinedinst & Koranteng, 2017).

[4]The typical acceptee's GPA creeps up over time due to the reduction in applications for students just below the GPA line. Students also follow the application patterns of their predecessors. Whether this improves or diminishes the quality of college a student attends depends on the types of colleges to which the student's predecessors applied.

[5]This includes year of graduation, high school, and 11th grade weighted GPA. When analyzing the impact of the SAT line on students' choices, students who did not take the SAT are excluded. I exclude the 2017 students from most analyses because I am missing NSC records for them. Students in the district's alternative high school are excluded.

disadvantage since I do not have information on many other factors (e.g. first-generation status) which contribute to student disadvantage and college choice (Easton, Johnson & Sartain, 2017; NCES, 1998; Roderick et al, 2009). These data are linked to National Student Clearinghouse data on postsecondary enrollment and degree completion for students who graduated high school by 2016. The district started collecting college application and admissions data in 2014. Application data are based on requests in Naviance to send student transcripts to colleges. Since most colleges require an official high school transcript, this should capture nearly everywhere students apply.

Admissions decisions are self-reported in a graduating student survey to which approximately 90% of students respond. Any inaccuracies in the self-reported admissions data will appear in the scattergrams. There is likely some under-reporting of acceptances, which will bias the acceptance criteria shown to students. The direction of this bias depends on which students under-report admissions, and this is difficult to identify.[6] While missing admissions data may bias the accuracy of the admissions information students see, it will not bias the estimates of the treatment I study.

The district enters the application and admissions data into Naviance at the end of each school year, along with data on test scores and GPAs, to populate the scattergrams. I use the application data uploaded to Naviance to reconstruct the scattergrams and identify the typical acceptee profile for each college, high school, and year combination. I also use these data to determine when each college-high school combination would have met the minimums of five and ten prior applicants. I cannot determine which high schools used which minimum applicant cutoff, but it appears that some schools use each one. Figure A.3 shows discontinuities at both thresholds.

The district purchased Naviance in 2014. At this point, there were no application data to upload, so high school students could access all features of Naviance except for the scattergrams. In the summer of 2014, application, admissions, and achievement data were uploaded to Naviance. Then, all high schoolers could see scattergrams based on students who graduated from their school in 2014. I focus on the college choices of the 2015 and 2016 graduating cohorts. These students were starting 11th and 12th grade when scattergrams first became visible. During the

[6] Many students do not report rejections, so the district treats non-responses as rejections. There appear to be a few students who over-report their acceptances, but this is less common than under-reporting. Appendix C contains a more thorough description of the data and the accuracy of students' self-reported experiences.

2015 school year, the 11th graders were getting ready to take the SAT and may have used the scattergram data to determine the SAT scores for which they should aim. The 12th graders were choosing where to apply to college and may have used the scattergrams to help with these choices. The 12th graders submitted applications by the winter of 2015, and received their admissions decisions by April 2015. In April and May of 2015, these students chose where to attend among the colleges to which they had been accepted, and most of them enrolled in college a few months later.

In June 2015, data on the class of 2015 were uploaded to Naviance and the scattergrams were updated to reflect the experiences of the 2014 and 2015 cohorts. This made new scattergrams available, since more colleges met the minimum data requirements, and existing scattergrams received new data points, which shifted the typical acceptee lines. Thus, the rising 12th graders, who would graduate in 2016, could see these updated scattergrams during the summer and fall when they were choosing where to apply and submitting their applications.

Login records are available for the class of 2017. They indicate the number of times each student's account was used to log onto Naviance in each grade during high school.[7] I cannot tell which scattergrams a student views, but students appear to use Naviance a lot. Usage was most frequent in 12th grade; the average student logged onto Naviance 23 times in 12th grade year and 43 times over the course of high school.[8] Figure A.4 shows that usage rates are highest for white and Asian students and those who never received free or reduced-price lunch.

Counselors were responsible for implementing Naviance. They received introductory materials and training from Naviance, similar to what other districts receive. The district counseling office also provided guidance to schools' counseling departments about when and how often to log into the platform with students. Counselors set up information sessions for parents and students, and logged on with students during school. They also provided specific suggestions about how to use the platform. In general, counselors had autonomy over the advice they provided.[9]

The school district is high performing compared to other districts regionally and nationally. The average student in my sample applied to five colleges and was accepted to about half of them.

[7]These may include parent logins since parents did not have accounts separate from their students.

[8]On average, students logged in 3 times in 9th grade, 5 in 10th grade, 11 in 11th grade, and 23 times in 12th grade.

[9]More implementation details are in Appendix B.

93% of the district's high school students graduate and 84% of students in my sample attended college in the fall following graduation, compared to national rates of 83% and 66%, respectively (NCES, 2015). Additionally, 71% of students in my sample who attended college started at a four-year college, compared to 64% nationally. This is consistent with the school district's low poverty rates. Despite high college enrollment rates, 27% of students who enroll in college attend a *safety* college, so there is room to improve the quality of colleges that students attend.

Given these outcomes and demographics, students in my sample probably have more information about college than the average student. This means there may be less room to influence college enrollment, but potentially more room to influence their application portfolios. Students and parents in this district may be more eager to consume information about college admissions or more inclined to apply to the highly selective colleges at which admissions information may be most relevant. For these reasons, it is unclear if my results will understate or overstate the average impact of admissions information on the college choices of U.S. high school students.

2.3 Changes Over Time

Students in this district first gained access to Naviance's scattergrams in summer 2015. In 2015, the scattergrams were for a roughly even mix of private, out-of-state public and in-state public colleges. In 2016, several private colleges and some out-of-state public colleges received scattergrams, shifting the mix to nearly 50% private and only 18% in-state public colleges. In both years, approximately 70% of the scattergrams were for highly selective colleges. The average student in my sample could view 47 scattergrams. I cannot tell which scattergrams a student actually viewed. Students may be guided to particular scattergrams by counselors and parents, or by quizzes in Naviance which suggest colleges based on a student's entered preferences.

Table 1 compares application and attendance patterns in the years that students could access the scattergrams (2015 and 2016) to those in the year before scattergrams were available (2014). There is a small shift in student characteristics over time but there is no change in the fraction of students attending four-year colleges.[10] In addition, there is no significant difference in the

[10]There is an increase in the share of free or reduced-price lunch students over this time. These students have lower college enrollment rates than higher income students, so we may have expected to see a decrease in college enrollment

number of colleges to which students apply; however, students are significantly more likely to be accepted at the colleges to which they apply in the years they can see scattergrams than in 2014.

This is consistent with students using the admissions information to choose colleges where they are more likely to be admitted. Students also apply to fewer *reach* colleges in the first year with scattergrams and they shift applications to colleges at which they are further above the average admitted student (in the district) than in the previous year (Figure A.5). In addition, Panel (C) of Table 1 indicates that students are 2 percentage points more likely to attend a *safety* college in the first year with scattergrams than in the previous year. These are colleges where students are likely to be admitted, but also where their achievement level exceeds the majority of other students. This may explain why college persistence rates are slightly lower in 2015 than 2014.

These patterns suggest that students may become intimidated by admissions information and reduce applications (and attendance) at colleges where they perceive their admissions probability to be low. The changes are consistent with students using Naviance to identify colleges at which they are likely to be accepted. This is good for admissions outcomes, but may deter students from attending the most selective college they are qualified to attend. This could prevent them from realizing the benefits associated with more selective colleges, including a higher graduation probability and higher earnings (Chetty et al., 2017; Dillon & Smith, 2018; Goodman & Cohodes, 2014). In the following sections, I examine the causal mechanisms driving these patterns by showing how access to scattergrams and their admissibility signals impact college choices.

3 Access to Admissions Information

First, I examine how gaining access to a college's admissions information influences where students apply to and attend college. Access to a college's scattergram may act as a nudge towards that college, perhaps because it increases awareness of the college or because it makes colleges with information seem like less risky choices than other colleges. The scattergrams also contain information about a college's popularity among students' peers. Students may update applications based on this, especially if they take popularity as a signal that the college is a good fit.

over this time. The lack of such a decrease could be due to the scattergrams increasing enrollment for this population.

3.1 Empirical Approach

Admissions data are shown on scattergrams for colleges with at least five or ten prior applicants from a high school. Each high school determines if five or ten is the appropriate minimum. I estimate the causal impact of access to a scattergram using a regression discontinuity design around these minimums. I compare application and attendance rates for colleges with just fewer than five or ten prior applicants to those which just met the criteria. I do not know which high schools use which threshold, so I stack my data and simultaneously estimate the discontinuities at both thresholds. I calculate a college's distance (in applications) from each cutoff and include an observation for each student, college, and threshold combination. The true impact of gaining access to a scattergram is twice my estimate because only one threshold is relevant to each scattergram.[11]

The discontinuities at both thresholds can be seen in Figure A.3, and the stacked version is in Figure 3. These figures show that application probabilities are linearly increasing in the number of applications a college previously received, with clear discontinuities at the visibility thresholds. This motivates the following local linear specification to estimate the impact of scattergram visibility on the probability that a student applies to or attends a college.

$$Y_{i,k} = \alpha_0 + \alpha_1 Visible_{j,k,t} + \alpha_2 Apps_{j,k,t} + \alpha_3 Visible \times Apps_{j,k,t} + \psi_i + \phi_{k,t} + \epsilon_{i,k,t} \tag{1}$$

Here, i indicates the individual, j the high school, k the college, and t the year. $Apps_{j,k,t}$ represents college k's distance, in applications (received from high school j between 2014 and year $t-1$), from the relevant application threshold. $Visible_{j,k,t}$ is a dummy variable indicating whether the number of applications exceeds the threshold. The interaction term $Visible \times Apps_{j,k,t}$ enables the slope of the regression lines to vary above and below the threshold. Y_{ik} is an indicator for whether student i applied to or attended college k. Observations are student-college-threshold combinations. I cluster standard errors at the student level and include fixed effects for each college by year and student. For each high school, I define the set of potential scattergrams, K_j,

[11] Crossing the visibility threshold at five increases the probability of having access to a scattergram from zero to some positive number, P. At ten it changes from P to 1. I do not know what P is and cannot estimate it in my data. However, I do not need to know this parameter to stack the data if I assume homogeneous treatment effects at the thresholds. The TOT effect is twice what I estimate since the probability of being treated at the five and ten thresholds sums to one.

as the colleges which received at least one application between 2014 and 2016 from high school j. This set varies across high schools, but is constant within a high school over time.

In some cases I estimate equation 1 separately by student distances from the average scores. I calculate the distance of students' 11th grade GPAs and maximum SATs from the typical acceptee for each college in K. For scattergrams below the visibility threshold, I impute what students would have seen. I define *near* the typical acceptee as within .5 GPA points and 150 SAT points. This definition matches the optimal bandwidth used in section 4 and it balances tradeoffs between sample size and the concentration of the visibility effects among students closest to the averages.

I focus on colleges within four applications of the visibility threshold.[12] This is the maximum feasible bandwidth that is the same for both thresholds, and on each side, without including colleges with no prior applications.[13] Variation in the number of prior applications comes from the popularity of a college and years over which application data were collected. Application data are based on transcript requests and they cannot easily be manipulated.[14] Any differences in the colleges on either side of the threshold should also be captured by the college by year fixed effects. I can employ these fixed effects because colleges have to cross the thresholds for each high school.

The colleges near the thresholds of five and ten prior applicants are not the most popular ones in this district. In terms of where students apply, in-state public colleges are under-represented and private colleges are over-represented. The regression discontinuity approach only enables me to estimate a local average treatment effect for colleges near the thresholds. I can, however, use the quasi-random variation in a coIllege's visibility across high schools and over time to examine how scattergram visibility impacts applications at the full set of colleges. For this, I use a specification which includes student fixed effects (ψ_i) and college by year fixed effects ($\phi_{k,t}$).

$$Y_{i,k} = \beta_0 + \gamma_1 Visible_{j,k,t} + \psi_i + \phi_{k,t} + \epsilon_{i,k,t} \tag{2}$$

γ_1 indicates the average impact of scattergram visibility (for all colleges with scattergrams in this

[12] Colleges with five to eight prior applicants appear twice per student in my estimates because they are in both cutoffs' bandwidths. Table A.2 shows robustness to randomly selecting one threshold per student-college combination.

[13] Colleges with zero prior applications do not fit the linear trend. Since the number of applications is discrete and I have relatively few groups, I cannot use traditional methods to calculate the optimal bandwidth.

[14] Appendix D shows there is no evidence of manipulation to make scattergrams available.

district) on applications or attendance (Y). $Visible_{j,k,t}$ is an indicator for whether college k's scattergram is visible in high school j in year t. Standard errors are clustered by student.

3.2 Results

Students are significantly more likely to apply to colleges with visible admissions information than colleges which just miss the visibility cutoffs. Panel (A) of Figure 3 shows a discontinuity in application probabilities at the point where a college crosses a visibility threshold. The x-axis shows how far a college is (in terms of applications) from the visibility threshold and the y-axis shows the fraction of students who apply to the colleges which are x distance from a threshold. Table 2 reports that application rates jump by .27 percentage points, from 1.37 percentage points to 1.64 percentage points, when a college crosses a visibility threshold.[15] Thus, the presence of admissions data increases the probability of applying to a college by at least 20%. The true effect is twice the point estimate (.54 pp) because scattergram visibility only changes at half the thresholds I use.[16] Table A.3 shows that the main results are robust to several alternate specifications.[17]

The dashed lines in Figure 3 compare the discontinuities for students who are near and far from the typical acceptee lines. They show that the discontinuity in application rates is much larger for students near the lines than those who are far from them. Table 2 column 6 reports that, among students near the typical acceptee's SAT and GPA, scattergram visibility increases application probabilities by .56 percentage points. The visibility effect increases as student scores become more similar to the typical acceptee's (Table 2 and Figure A.6).[18] Thus, information seems to have the largest impact on the application choices of the students for whom it is most relevant.

Table 3 shows that, among students who are near the typical acceptee, gaining access to a scat-

[15] The probability of applying to any one of these colleges is low because there are many scattergrams.

[16] My estimates are similar when I randomly select one threshold to keep for each student-college combination (Table A.2).This avoids double-counting student-college combinations.

[17] They are similar when I expand or shrink the bandwidth, use a triangular kernel specification, and when I cluster the standard errors by level of treatment (school by year by college), or when I use the approach described by Kolesár & Rothe (2018) for regression discontinuity designs with discrete running variables. In addition, the results are not driven by the serial correlation, since scattergram visibility has the largest impact in the first year available. (There is serial correlation in the running variable over time, since the number of prior applications can only increase over time.)

[18] My results are similar when I look at student proximity to alternate versions of the typical acceptee lines (Table A.4). I focus on proximity to the weighted GPA and SAT 2400 because these measures contain the most information. Table A.5 also shows how the visibility effects vary by distance from the average GPA and SAT lines.

tergram has the largest impacts on students who received free or reduced-price lunch (FRPL) and Black or Hispanic students. Scattergram visibility increases applications by 40% (1.2pp) among students who received FRPL and 36% (1.2pp) for Black and Hispanic students. These students are the most likely to lack information about college (Hoxby & Avery, 2013).[19]

Scattergram visibility also has larger impacts for in-state public colleges, increasing application rates by 62% (1pp). Students may view scattergrams for in-state public colleges more than other colleges, because they are nearby and inexpensive, or because they are more likely to have heard about these colleges. Thus, large effects at these colleges could be due to students viewing their information at higher rates, or because it is easier to influence applications at colleges which are inexpensive and nearby. The district is located in a state with many small in-state public colleges, so students may have been unaware of these options before they saw scattergrams.

The application effects translate into effects on attendance for some subgroups. The dashed lines in Panel (B) of Figure 3 show a .1 percentage point discontinuity, in attendance rates, for students near the average admit. There does not appear to be a discontinuity for students who are far from the lines or the pooled sample. Table 4 shows that once I add college by year (and student) fixed effects, this drops to an insignificant .01 percentage points. This is may be from limited power; the college fixed effects absorb a lot of the variation in outcomes.[20] I may also find weaker effects for attendance than applications because students have to be admitted to a college before they can attend it, and students can apply to many colleges but they can only attend one.

Table 4 indicates that visibility has a significant impact on attendance rates for Black and Hispanic students, as well as at in-state public colleges. Students are .28 percentage points (or 127%) more likely to attend an in-state public college if its scattergram is visible. Black and Hispanic students who are similar to the typical acceptee are .47 percentage points (196%) more likely to attend a college if they can see its scattergram. There is also a marginally significant attendance effect for students receiving free or reduced-price lunch who are near the lines. These are the same students whose application choices are most influenced by access to the admissions data.[21]

[19] Income and race are correlated with other unobservable factors that influence college awareness.

[20] This reduction is driven by the college fixed effects. With student fixed effects alone, the discontinuity is a significant .1 percentage point. The graph does not contain fixed effects. Estimates without fixed effects are in Table A.3.

[21] These students also apply to fewer colleges than their higher income and white/Asian peers. Since you can only

Dividing the attendance estimates by the application estimates indicates that 22% of Black and Hispanic students induced to apply to a college by a scattergram go on to attend that college. This is 38% for students near the typical acceptee lines, probably because they are more likely to be admitted to the college (and not many more selective colleges) than students far from the lines. In addition, 29% of students induced to apply to an in-state public college by a scattergram attend it.

The previous results are local average treatment effects for colleges near the visibility thresholds. Table 5 shows that, for the full set of colleges, scattergram visibility increases applications by .9 percentage points. This is more than three times the effect for colleges near the visibility thresholds. For students near the typical acceptee lines, visibility increases applications by 1 percentage point (approximately double the LATE). I also detect significant effects on attendance for the full set of scattergrams (Table A.6). These estimates indicate that admissions information has large effects on where students apply and attend, with larger impacts for the more popular colleges.

Finally, student responses to scattergrams vary based on the counselor to which the student is assigned.[22] This suggests there may be variation across counselors in the guidance they provide around how to use Naviance and college options. Students may also find scattergrams more beneficial if their counselor provides little support in the college choice process.

4 Role of Admissibility Signals

Next, I examine how students use the admissibility signals in Naviance to choose where to apply to college. On average, students can see 47 scattergrams. This is a lot of information to sort through and students are far from the typical acceptee's scores on many scattergrams. To better understand how students sort through the scattergrams and use admissions information, I study student reactions to the two clear signals the scattergrams provide about admissions probabilities.

Scattergrams show students (1) how similar their GPAs and SATs are to previous admits, and (2) whether their scores are above or below average. The first signal tells students something about their admissions probability, and whether they are qualified for a more selective college. The

attend one college (immediately after high school), a FRPL or Black/Hispanic student's application is more likely to translate into attendance than another student's application.

[22]Students are assigned to counselors based on their last name, so selective sorting should not drive these patterns.

second signal has little bearing on a student's admission probability, conditional on being near the line, because students just above and below a noisy average should have similar admissions rates. Students may, however, use the lines as heuristics or reference points because of their saliency and the complex nature of college choices. I find evidence consistent with this hypothesis and with students updating their application portfolios to increase their perceived admissions probabilities.

4.1 Empirical Approach

Students can easily see how they compare to the typical acceptee because Naviance marks the user's position on a scattergram with a red circle, and the college dashboard color codes whether a user's scores is above average (Figure A.2). For each college, the typical acceptee lines vary across the high schools and years the scattergram is available.[23] This generates quasi-random variation in a student's distance from the perceived admissions criteria. I use this variation to identify the causal effect of one's perceived admissions probability on the decision to apply to a college.

Naviance users can choose which types of GPAs and test scores populate the scattergrams. For simplicity, I focus on one orientation of the scattergram. I report results for the weighted GPA and SAT M+V+W (2400) averages because they are more informative than the unweighted GPAs and SAT M+V (1600).[24] There is more variation in the scores on the larger scales, the weighted GPA includes information about the rigor of students' courses, and the 2400 SAT score includes writing scores. Results for the unweighted GPAs and 1600 scale SAT scores are similar.

First, I examine whether students just above the typical acceptee lines have different responses to the availability of admissions information than those just below them. I focus on students within .1 GPA points or 50 SAT points of the typical acceptee, so that the students are all similar. Within these bandwidths, I estimate α_1 in equation 1 separately for students above and below the lines. I do this separately for the SAT and GPA lines (and find no evidence of a joint effect).

Second, I estimate the impact of students' perceived admissions probabilities, as captured by their distances from the average admit, on applications and attendance. I use the quasi-random variation in the average admit's scores across schools and years to identify the causal effects of the

[23]They are also fairly noisy signals because they are typically only based on a few admitted students.

[24]Users could also view ACT scores but few students in the district took the ACT so there was much less data on it.

perceived admissions criteria. Finally, I look at all available scattergrams and estimate the effect of being just below the average GPA or SAT (relative to just above it) using a regression discontinuity.

I estimate the impact of the average scores, and one's distance from them, using the specification in equation 3. This specification includes college by year and high school fixed effects (δ_{tk} and ψ_s), as well as controls for student demographics ($Demographics_i$) and academic achievement ($AcadAchieve_i$). I allow for application probabilities to change discontinuously when a student moves below the typical acceptee's score to account for the potential effect of this signal on student outcomes. This amounts to a regression discontinuity design around a typical acceptee's score, where the coefficient β_1 indicates the extent to which being below the score has a causal impact on students' applications.[25] β_2 indicates how the probability of applying to a college changes as a student's GPA or SAT moves further above the typical acceptee's, and β_3 indicates how this probability changes as a student's score moves further below the typical acceptee's.

$$\begin{aligned} Y_{ik} = \beta_0 + &\beta_1 Below_{ik} + \beta_2 ScoreDist_{ik} * Above_{ik} + \beta_3 ScoreDist_{ik} * Below_{ik} \\ &+ \beta_4 Demographics_i + \beta_5 AcadAchieve_i + \delta_{tk} + \psi_s + \epsilon_{i,k} \end{aligned} \tag{3}$$

Observations are student-scattergram combinations, where k indicates the college associated with the scattergram and i the individual. $Below_{i,k}$ is an indicator for whether the student is below the typical acceptee's score for college k and $Above_{i,k}$ is an indicator for being above it. $ScoreDist_{i,k}$ represents the distance of student i's GPA or SAT from the typical acceptee for college k. $Y_{i,k}$ is an indicator for whether student i applies to or attends college k. Standard errors are clustered by student. I separately estimate the impact of the GPA and SAT lines because student responses are driven by the GPA line. Table A.7 describes the results when I jointly estimate the impacts of these lines (following Papay, Murnane & Willet (2015) and Robins & Reardon (2012)).

I focus on colleges which received at least ten applications in prior years since their scattergram will appear regardless of the minimum rule the high school is using.[26] The optimal bandwidths are .5 GPA points and 150 SAT points, which are consistent with the definition of near described

[25] I focus on the impact of being below a line, rather than above it, because the placebo test in Figure 5 suggests that the line is reducing aspirations for students below it, rather than increasing them for students above it.

[26] The results are similar but muted if I treat five as the universal minimum.

in the previous section (Calonico, Cattaneo, & Titiunik, 2014). Columns (6) and (7) in Table A.1 describe the observations that fall in these bandwidths. The average student is within the GPA bandwidth for 18 scattergrams and the SAT bandwidth for 11 scattergrams.[27]

4.2 Results

Section 3 shows that access to admissions information has the largest effect on students whose scores are most similar to the average admit. This effect is driven by both the GPA and SAT lines (Figure 4). Students within .1 GPA points or 50 SAT points of the typical acceptee lines are more likely to increase applications due to scattergram visibility than students who are not. In addition, students whose GPAs are just above the average admit's increase applications more in response to visibility than students who are just below it (Table A.5). This suggests that students are most likely to respond to information when it signals something positive about their admissibility, and students may use the average scores as heuristics to help them determine where to apply.

The navy lines in Panel (A) of Figure 5 show that students are most likely to apply to colleges at which their GPA matches or slightly exceeds the average GPA of previous admits. Application probabilities decrease with a student's distance from the GPA line. Students may reduce applications as they move further below the average because their perceived admissions probability declines. Students moving further above the average GPA may decrease applications because the information signals that they can gain admission to more selective colleges.

Panel (A) of Figure 5 also shows a significant gap in application rates at the point where a student's GPA crosses above the average GPA. This is consistent with students interpreting the line, or GPA color-coding, as signals about their admissibility. They may also use these as heuristics to help them determine where to apply. The gray lines in Figure 5 are based on students who graduated in 2014 and could not see any scattergrams. Comparing the gray and navy lines, it appears that the GPA line reduces aspirations for students just below it rather than increasing them for students above the line. This motivates the focus on the negative effect of being below the line.[28]

[27] Appendix D shows no manipulation of the running variables in the regression discontinuity specifications.

[28] The peak at zero for students who could not see the scattergrams is partly mechanical because the typical acceptee lines are based on their application patterns. These applicants (in 2014) must be similar to the average admit in 2014 because the average admit is based on the 2014 applicants. The reduction in application probabilities over time for a

Row (1) of Table 6 shows that students just below the typical acceptee's GPA are 1.1 percentage points (8%) less likely to apply to the college than students just above it. Rows (2) and (3) indicate that moving .1 GPA point away from the average GPA decreases application rates by about one percentage point. Panel (B) of Figure 5 shows a similar pattern for SAT scores, but there is no discontinuity at the average SAT line.[29] This may be because there are many sources of information about SAT admissions, including within Naviance. If students find information inconsistent with what they see on scattergrams they may not place much weight on Naviance's SAT signals.

Figure 6 shows that admissibility signals are most important for application decisions at highly selective colleges. Table 6 reports that students just below the average GPA for a highly selective college are 1.9 percentage points (15%) less likely to apply than students just above it. There is no discontinuity for less selective colleges. Admissibility signals may be most relevant to highly selective college decisions because admissions probabilities are much lower at these colleges, or because students in this district have access to more scattergrams for highly selective colleges than less selective ones.[30] If students only see a few less selective schools, the decision of which to apply to may be relatively simple. In contrast, choosing among 15-20 highly selective colleges may be a daunting task, leading students to rely on heuristics to narrow their choice set.

This is consistent with the larger impacts I find for students who could see more scattergrams.[31] The admissibility signals in Naviance also have the largest effects on students who logged onto Naviance the most (Table A.10). White and Asian students (as well as non-FRPL) students were the most likely to be frequent Naviance users, so it is unclear if these large effects are driven by looking at more scattergrams (more often) or other characteristics of these students.

Overall, these estimates indicate that application choices are sensitive to what the typical ac-

---

college is due to mean reversion and to students spreading out their applications over a larger set of colleges.

[29]The results are robust to a triangular kernel specification as well as to larger and smaller bandwidths. In addition, they are similar with a donut specification, which excludes students whose Naviance dots are on top of the typical acceptee line (Tables A.7, A.8 and A.9). The results are also similar when I look at alternate orientations of the scattergrams, including unweighted GPAs or SAT scores on the 1600 scale (Table A.7).

[30]This is because high-achieving students apply to more colleges than low-achieving students, and high-achieving students disproportionately apply to highly selective colleges.

[31]Higher income and white/Asian students can see the most scattergrams (because of the high schools they are concentrated in) and they are the most responsive to the averages. The class of 2017 is also more responsive than earlier cohorts, perhaps because they could see more scattergrams. This is indicates that students' focus on the averages may be a growing concern as districts use Naviance for longer.

ceptee profiles signal about a student's probability of admission. Application and attendance choices respond to the probability of admission conveyed in the student's distance to the lines and in some cases whether they are above or below the line.[32] Student responses to the lines also vary based on their assigned counselor (Figure A.7). Some counselors appear to mitigate the responses to the line, while others exacerbate them. This indicates potential for school personnel to help students synthesize admissions information and craft an application portfolio.

5 Cumulative Effects of Admissions Information

So far, I have shown how information or admissibility signals for a particular college influences applications and enrollment at that college. Now, I show how the full set of available information and signals influences application portfolios, college attendance, and college selectivity. In my sample, the average student could access 47 scattergrams and was near the typical acceptee for 21. I define the scattergrams for which students are near the average admit as the relevant ones, since these influence applications the most. Then, I explore how this set of information influences college choices, which is useful for understanding the cumulative effects of Naviance.

5.1 Empirical Approach

For each student, I calculate the number of relevant scattergrams she could access, the number that were *reach*, *match*, and *safety* colleges, and how many were in-state public colleges. Relevant scattergrams are defined as those where the student's SAT and GPA are within 150 and .5 points, respectively, of the average admit. This is consistent with the definition of near from section 3. *Reach* colleges are defined as those where the student's maximum SAT score is below the 25th percentile of accepted students' SATs, and *safety* colleges are those where her SAT score is above the 75th percentile. Match colleges are those where her SAT is in the inter-quartile range.[33]

[32]The concentration of responses among highly selective colleges, combined with low admissions rates at these colleges, may explain why I find no significant effect of being below a college's GPA line on attendance at that college. Low attendance probabilities for these colleges contribute to limited power to detect effects on attendance. I do, however, find that students are most likely to attend colleges where they are similar to previous admits (Tables 6 and 7).

[33]This is based on the inter-quartile range of accepted students' SATs from IPEDS in 2015. I use this measure because it is simple to calculate for all students. The measures used by Hoxby & Avery (2013) and Dillon & Smith (2018) are

To estimate the cumulative effects, I use within school variation in how many relevant *reach*, *match*, or *safety* college scattergrams a student could view based on the colleges meeting the visibility thresholds and the average scores. There are two sources of identifying variation within a high school. First, two students with identical scores in different cohorts will see different sets of scattergrams and different average scores on the scattergrams available in both years. Second, classmates will see the same set of scattergrams, but the set of relevant and *reach*, *match*, or *safety* ones will vary according to the student's SAT and GPA. Most of my identifying variation comes from variation over time in what similar students see because I control for academic achievement.

I use this variation to measure how the set of scattergrams influences where students apply and attend. I regress Y_i, a characteristic of a student's application portfolio or college attended, on a characteristic of the visible scattergrams, SGs_i. Scattergram characteristics, SGs_i, include the number available, the number of *reach*, *match*, or *safety* scattergrams, and the number for in-state public colleges. I flexibly control for academic achievement, demographics, school fixed effects, and year fixed effects. $\Gamma_1$indicates how gaining access to an additional relevant scattergram of type SGs_i (such as a *match* college) impacts outcome Y_i, such as attendance at a *match* college.

$$Y_i = \Gamma_0 + \Gamma_1 SGs_i + \Gamma_2 Demographics_i + \Gamma_3 AcadAchieve_i + \delta_t + \psi_j + \epsilon_i \tag{4}$$

5.2 Results

Students' application portfolios reflect the set of colleges with visible and relevant scattergrams. Table 8 indicates that the extent to which students apply to and attend *reach*, *match*, or *safety* colleges is related to the number of relevant scattergrams they see for each of these types of colleges. For example, students who could see more relevant scattergrams for *reach* colleges were more likely to apply to and attend *reach* colleges. Low-income and minority students who see more *match* scattergrams are also more likely to attend a match college and persist in college.[34]

The effects on persistence are small and noisy, but in directions consistent with prior research on college match (Dillon & Smith, 2018; Cohodes & Goodman, 2014). The coefficients tend to be

more complicated to calculate, especially since I have a non-representative sample of applicants to each college.

[34] The persistence estimate is only marginally significant (Table A.11.

positive for students nudged to attend match or reach schools and negative for those nudged towards safety schools. They are also positive and marginally significant for in-state public colleges, probably because seeing more of these scattergrams increases attendance at match colleges. The persistence estimates may be small due to limited power or the nature of the effects.[35]

Among Black, Hispanic, and low-income students, every additional relevant scattergram they saw for an in-state public college increased their probability of attending a four-year college by 2.3 percentage points (Table A.11. This effect is driven by the many smaller and less-selective in-state public colleges near the district. Students may have been unaware of these options before Naviance, so that learning about nearby and inexpensive options, other than the local community college and state flagship, shifted attendance from the local community college to these schools.[36]

The set of relevant scattergrams does not impact the number of applications or the probability of being accepted to a college. This is consistent with students applying to the same number of colleges in the years with and without scattergrams. Thus, scattergrams lead to substitutions in applications across colleges, not changes in the number of colleges to which students apply. Some of this substitution is driven by the switches among *reach*, *match*, and *safety* schools, as shown in Table 8. Students also shift applications from medium popular colleges, such as neighboring states' flagship universities, to less popular colleges in the first year with scattergrams.[37] This is consistent with scattergrams broadening the set of schools to which students apply.

The constant application rate also indicates that students who do not apply to a college because they are just below the GPA line switch their application to another college. I find no evidence of students shifting applications to the college's closest competitor or a college of similar selectivity. Instead, students appear to shift applications to less selective colleges (Figure A.8). Figure 7 shows

[35]Persistence data are only available for the 2015 cohort. In addition, changes to persistence come through changes to where students attend college. Given the magnitude of the changes in attendance, the small and mostly insignificant effects on persistence are not surprising. For example, if the set of scattergrams available increases attendance at match colleges by 3 percentage points (which is about the largest effect I find), and if those schools have persistence rates that are 30 percentage points higher than students' counterfactuals, the expected effect on persistence would be about 0.009.

[36]I find no effect on the overall rate of college attendance which suggests that the scattergrams shift students from two-year to four-year colleges. The lack of change in district-wide four-year attendance rates in the year before and after scattergrams were available may be due to the increase in the share of Black and FRPL students in the district over this time. These students have lower four-year attendance rates, so an increase in their representation could have led to a decrease in the district's college-going rate if not for the scattergram's availability increasing their attendance rates.

[37]See Appendix C for a description of this and other definitions used in this section.

that students below the typical acceptee lines at highly selective colleges are less likely to attend an elite college than students above the lines. Students also substitute enrollment away from private colleges to in-state public colleges when below the GPA line at a private college.

Overall, these results indicate that the admissions information conveyed on the scattergrams is improving some students' college choices, but deterring others from applying to highly selective colleges and attending the most selective college for which they are qualified. The net effects of Naviance's scattergrams depend on the set of relevant scattergrams to which a student had access and the magnitude of the typical acceptee deterrence relative to the positive effects of access to a college's information. Among colleges near the visibility threshold, Figure 4 shows that the visibility effect is larger than the deterrant effect of the average GPA. It is unclear if this holds for more popular colleges or in districts with more available scattergrams.

6 Discussion and Conclusion

This paper shows that a technology, such as Naviance, is capable of providing personalized college admissions information to many students in a way that significantly alters college choices. Providing access to data on schoolmates' college admissions experiences increases applications to that college. Application effects are largest for students who are most likely to lack information about the college admissions process, and this translates into an effect on where they attend college. Providing low-income and minority students information about local and inexpensive options also increases their four-year college enrollment. Thus, this type of information and technology has the potential to help to close socioeconomic gaps in college enrollment and impact students' labor market outcomes (Zimmerman, 2014).

The overall effects of Naviance's admissions information varies across students and schools based on which colleges and admissions criteria are visible. Students increase applications most when the information conveys a positive signal about their admissibility and fit at the college. Thus, application portfolios reflect the set of colleges with visible scattergrams and average admissions scores near the student's. Whether this is a good set of colleges to nudge students towards depends on the types of colleges to which their schoolmates previously applied and how accurately previous admits' scores reflect the true admissions criteria. Students in high schools with strong college-going cultures are more likely to be nudged to highly selective colleges while those in schools with suboptimal college choices among older cohorts will be nudged to repeat the suboptimal choices of their peers.[38] In the future, it may be valuable to more carefully curate the colleges for which students receive information. In cases where insufficient data on prior applicants from a high school exist to make a scattergram visible, districts could pool data across schools; this may also help to improve the accuracy of the admissions criteria that students see.

While the typical acceptee lines have some negative consequences, their capacity to make admissions information very salient may drive some of the positive information effects. On net, the positive effect of providing information has a larger impact on application and enrollment choices at a college than the negative effect of the GPA line.[39] Future work could explore how to positively harness the power of salient information while minimizing suboptimal responses.[40]

Counselors or Naviance staff could also do more to help students accurately interpret the scattergram data. Naviance usage, and the impacts of the admissions information, vary based on the counselor assigned to a student. In some cases, Naviance may be a substitute for the advice provided by counselors, while in other settings it may be a complement to the counselor's role, enabling them to more efficiently serve students.

Interesting avenues of future research would be to examine how features aside from the scattergrams impact college choices and how counselors' implementation influences its effectiveness. The impact of this technology may also change as more cohorts of data are added, making more scattergrams available and increasing the stability of the typical acceptee lines. In the three years I study, the impact of an individual scattergram's visibility shrinks as more scattergrams become available (Table A.14); however, the importance of the lines grows as students have more scattergrams to sort through.[41] Furthermore, responses to the typical acceptee's GPA lead the average

[38]Many schools have few students attending four-year colleges, let alone highly selective ones, so the the scattergram tool may not improve college choices in these places (Radford, 2013; Hoxby & Avery, 2013).

[39]Figure 4 shows a positive impact of scattergram access for students below the typical acceptee line.

[40]Naviance could include adding an inter-quartile range to the graphs or adding a gradient of shading around the lines to depict how admissions probabilities change throughout the scattergram. They could also stop making a user's score red when it is just below average, perhaps turning them yellow, and only turning them red at a lower threshold.

[41]Districts are able to choose how many cohorts of data students can see. It is possible that students will discount the information more if many older cohorts are included since trends over time will not be captured.

GPA line to increase over time, which can reduce the accuracy of the information students see.[42]

Finally, my results may understate the true effect of this type of admissions information on the average student since I do not know which threshold applied at each high school. In addition, access to this type of information may have larger impacts in districts where fewer students attend college or where students have less information about college. Students in this school district have high college attendance rates compared to the national average. Given that over forty percent of U.S. high school students are using Naviance, and many of them are less advantaged than those in my sample, this technology has the potential to influence national trends in college choices.

More broadly, this paper shows that information can have large effects on where students apply to college and that a low-cost technology can effectively deliver personalized information. The framing and personalization of information in this context may explain why I find larger effects than some prior studies. This sort of technological personalization can also be more cost effective than personalized assistance and it can be implemented quickly at a large scale.[43] Students may pay attention to the information in Naviance because it is based on their schoolmates, and thus likely to be more relevant than general information. This is consistent with other work showing that students care about peer norms and college choices (Bursztyn & Jensen, 2015; Radford, 2013). This paper, however, shows that nudging towards social norms may not be optimal if the norms are suboptimal. In addition, data framing may lead to adverse reaction, so designers of information interventions should carefully consider potential responses.

Finally, this paper indicates that information about admissibility is an important piece of the application choice problem. Students may, however, place too much weight on their admissibility. Given the high returns to many highly selective colleges, and the low cost of applying to them, it is probably not optimal for students to respond so strongly to admissions signals.

[42]On average, they increase by 0.008 GPA points per year (p=0.002). If they continue to increase, the positive effect of scattergram visibility may be overtaken by the negative effect of the typical acceptee's GPA, so that the net effects of visibility are no longer positive. The SAT lines get lower over time (4 points per year).

[43]Cost data is unavailable this district, but a few other districts pay less than ten dollars per student for Naviance.

7 References

Akerlof, George A., and Rachel E. Kranton. 2002. Identity and schooling: Some lessons for the economics of education. *Journal of Economic Literature* 40 (4): 1167-1201.

Avery, Christopher, and Thomas J. Kane. 2004. Student perceptions of college opportunities. The Boston COACH program. In *College choices: The economics of where to go, when to go, and how to pay for it*, pp. 355-394. University of Chicago Press.

Barr, Andrew, and Sarah Turner. 2018. A Letter and Encouragement: Does Information Increase Postsecondary Enrollment of UI Recipients?. *American Economic Journal: Economic Policy* 10 (3): 42-68.

Bergman, Peter, and Eric Chan. Forthcoming. Leveraging technology to engage parents at scale: Evidence from a randomized controlled trial. *Journal of Human Resources*.

Bettinger, Eric P., Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu. 2012. The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment. *The Quarterly Journal of Economics* 127, (3): 1205-1242.

Bond, Timothy N., George Bulman, Xiaoxiao Li, and Jonathan Smith. 2018. Updating human capital decisions: Evidence from SAT score shocks and college applications. *Journal of Labor Economics* 36 (3): 807-839.

Bursztyn, Leonardo, and Robert Jensen. 2015. How does peer pressure affect educational investments?. *The Quarterly Journal of Economics* 130, (3): 1329-1367.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2014. Robust nonparametric confidence intervals for regressiondiscontinuity designs. *Econometrica* 82, (6): 2295-2326.

Carter, Susan Payne, Kyle Greenberg, and Michael S. Walker. 2017. The impact of computer usage on academic performance: Evidence from a randomized trial at the United States Military Academy. *Economics of Education Review* 56: 118-132.

Castleman, Benjamin L., and Lindsay C. Page. 2015. Summer nudging: Can personalized text messages and peer mentor outreach increase college going among low-income high school graduates?. *Journal of Economic Behavior & Organization* 115: 144-160.

Castleman, Benjamin L., and Lindsay C. Page. 2016. Freshman year financial aid nudges: An experiment to increase FAFSA renewal and college persistence. *Journal of Human Resources* 51, (2): 389-415.

Chade, Hector, Gregory Lewis, and Lones Smith. 2014. Student portfolios and the college admissions problem. *Review of Economic Studies* 81, (3): 971-1002.

Chetty, Raj, John N. Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan. 2017. Mobility report cards: The role of colleges in intergenerational mobility. *National Bureau of Economic Research* Working paper no. w23618.

Clinedinst, Melissa, and Anna-Maria Koranteng. 2017. 2017 State of College Admission. *National Association for College Admisssion Counseling*. Retrieved from https://www.nacacnet.org/globalassets/documents/publications/research/soca17final.pdf

Cohodes, Sarah R., and Joshua S. Goodman. 2014. Merit aid, college quality, and college completion: Massachusetts' Adams scholarship as an in-kind subsidy. *American Economic Journal: Applied Economics* 6, (4) : 251-85.

Dettling, Lisa J., Sarena Goodman, and Jonathan Smith. 2018. Every little bit counts: The impact of high-speed internet on the transition to college. *Review of Economics and Statistics* 100, (2): 260-273.

Dillon, Eleanor Wiske, and Jeffrey A. Smith. 2018. The consequences of academic match between students and colleges. *National Bureau of Economic Research* Working paper no. w25069.

Drezner, D.W. 2017. Perspective: The extremely useful piece of software for high school students that I would like to kill with fire: Why Naviance is great and drives me crazy. *The Washington Post*. September 20. Retrieved from:
https://www.washingtonpost.com/news/posteverything/wp/2017/09/20/the-extremely-useful-piece-of-software-for-high-school-students-that-i-would-like

Easton, John Q., Esperanza Johnson, and Lauren Sartain. 2017. *The predictive power of ninth-grade GPA.* Chicago, IL: University of Chicago Consortium on School Research.

Escueta, Maya, Vincent Quan, Andre Joshua Nickow, and Philip Oreopoulos. 2017 Education technology: an evidence-based review. *National Bureau of Economic Research* Working paper no. w23744.

Fu, Chao. 2014. Equilibrium tuition, applications, admissions, and enrollment in the college market. *Journal of Political Economy* 122, (2): 225-281.

Gelger, Will. 2018. What are My Chances? High Schoolers, Scattergrams, and the College Admissions Process. *Forbes*. February 20. Retrieved from:
https://www.forbes.com/sites/noodleeducation/2018/02/20/scattergrams-and-college-admissions/#7decedde7af8

Goodman, Joshua, Oded Gurantz, and Jonathan Smith. Forthcoming. Take Two! SAT Retaking and College Enrollment Gaps. *American Economic Journal: Economic Policy*.

Goodman, Joshua, Michael Hurwitz, and Jonathan Smith. 2017. Access to 4-year public colleges and degree completion. *Journal of Labor Economics* 35, (3): 829-867.

Hastings, Justine, Christopher A. Neilson, and Seth D. Zimmerman. 2015. The effects of earnings disclosure on college enrollment decisions. *National Bureau of Economic Research*, Working paper no. w21300.

Hastings, Justine S., and Jeffrey M. Weinstein. 2008. Information, school choice, and academic achievement: Evidence from two experiments. *The Quarterly Journal of Economics* 123, (4): 1373-1414.

Hoxby, Caroline, and Christopher Avery. 2013. The missing one-offs: The hidden supply of high-achieving, low-income students. *Brookings Papers on Economic Activity* 2013, (1): 1-65.

Hoxby, Caroline M., and Sarah Turner. What high-achieving low-income students know about college. *American Economic Review* 105, no. 5 (2015): 514-17.

Hurwitz, Michael, and Jessica Howell. Estimating causal impacts of school counselors with regression discontinuity designs. *Journal of Counseling & Development* 92, no. 3 (2014): 316-327.

Hurwitz, Michael, and Jonathan Smith. 2018. Student responsiveness to earnings data in the College Scorecard. *Economic Inquiry* 56, (2): 1220-1243.

Kahneman, Daniel. 1992. Reference points, anchors, norms, and mixed feelings. *Organizational behavior and human decision processes* 51, (2): 296-312.

Kahneman, Daniel. 2003. Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review* 93, (5): 1449-1475.

Kolesár, Michal, and Christoph Rothe. 2018. Inference in regression discontinuity designs with a discrete running variable. *American Economic Review* 108, (8): 2277-2304.

Lavecchia, Adam M., Heidi Liu, and Philip Oreopoulos. 2016. Behavioral economics of education: Progress and possibilities. In the *Handbook of the Economics of Education*, vol. 5, pp. 1-74. Elsevier.

Luca, Michael, and Jonathan Smith. 2013. Salience in quality disclosure: evidence from the US News college rankings. *Journal of Economics & Management Strategy* 22, (1): 58-77.

Mulhern, Christine. 2019. Beyond teachers: Estimating individual guidance counselors' effects on educational attainment. *Working Paper*.

National Center for Educational Statistics. 1998. First-generation students: Undergraduates whose parents never enrolled in postsecondary education. Washington, DC: U.S. Department of Education.

Pallais, Amanda. 2015. Small differences that matter: Mistakes in applying to college. *Journal of Labor Economics*, 33 (2): 493-520.

Papay, John P., Richard J. Murnane, and John B. Willett. 2014. High-school exit examinations and the schooling decisions of teenagers: Evidence from regression-discontinuity approaches. *Journal of Research on Educational Effectiveness* 7 (1): 1-27.

Radford, Alexandria Walton. 2013. *Top student, top school?: How social class shapes where valedictorians go to college.* University of Chicago Press.

Reardon, Sean F., and Joseph P. Robinson. 2012. Regression discontinuity designs with multiple rating-score variables. *Journal of Research on Educational Effectiveness* 5, (1): 83-104.

Roderick, Melissa, Jenny Nagaoka, Vanessa Coca, and Eliza Moeller. *From high school to the future: Potholes on the road to college.* Research Report. Consortium on Chicago School Research. 1313 East 60th Street, Chicago, IL 60637, 2008.

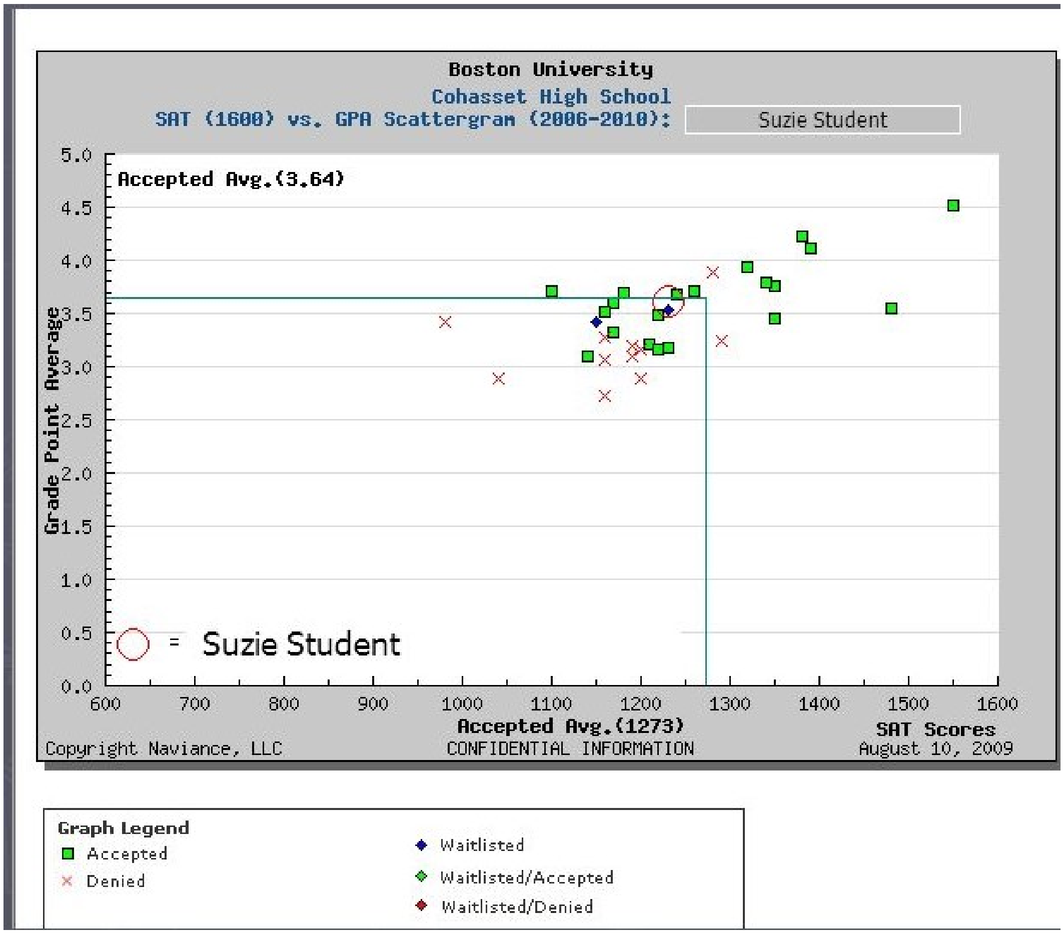
Shellenbarger, Sue. 2017. College-search quandry? There's an app for that: Dozens of digital tools aim to help students find their dream school. *Wall Street Journal*. September 19. Retrieved from:
https://www.wsj.com/articles/college-search-quandary-theres-an-app-for-that-1505832484

Shulman, Robyn D. 2018. "EdTech investments rise To A historical $9.5 Billion: What your startup needs to know." *Forbes.* January 26. Retrieved from:
https://www.forbes.com/sites/robynshulman/2018/01/26/edtech-investments-rise-to-a-historical-9-5-billion-what-your-startup-needs-to-know

Zimmerman, Seth D. 2014. The returns to college admission for academically marginal students. *Journal of Labor Economics* 32, (4): 711-754

8 Tables and Figures

Figure 1: Example Scattergram

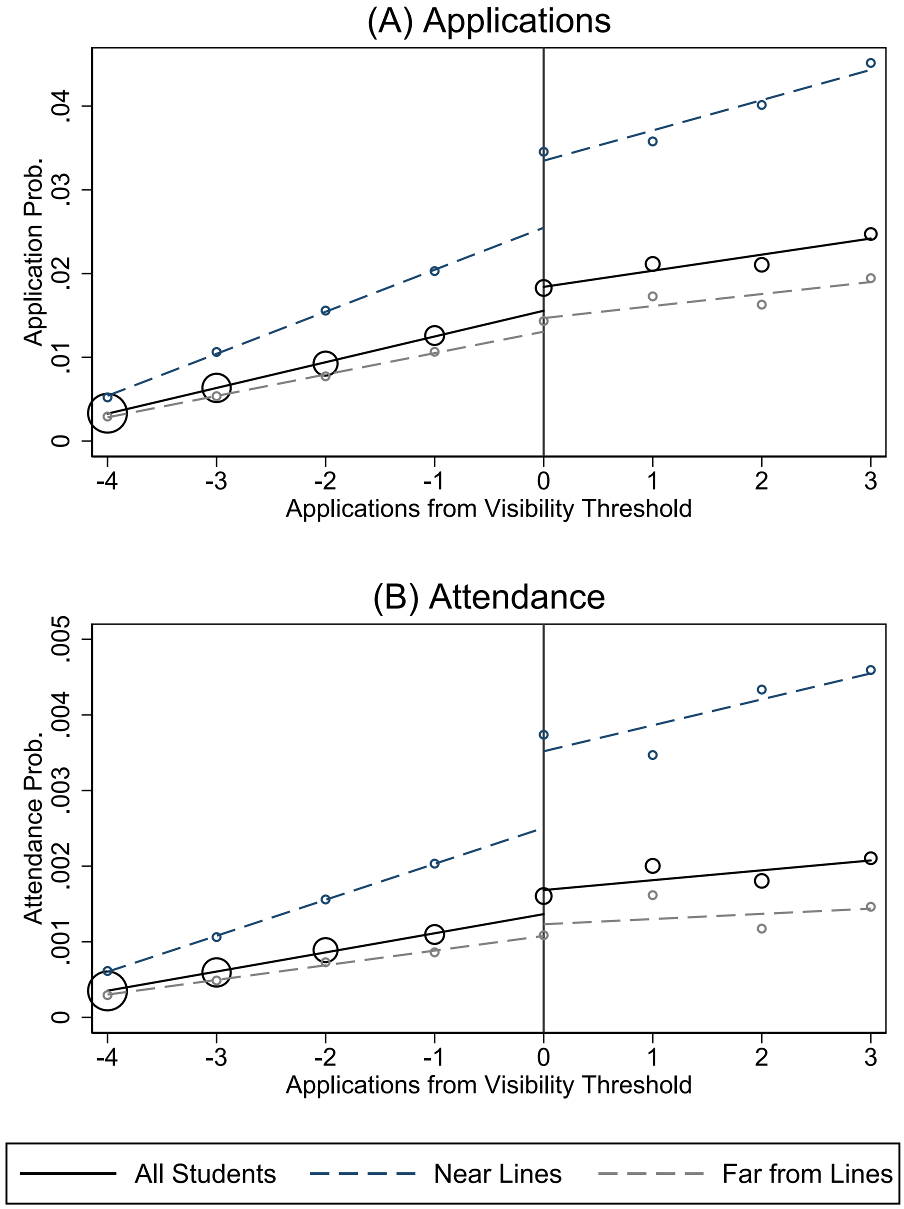


Notes: Photo credit: Naviance LLC. This is a fictional example of a scattergram. The red circle represents the GPA and SAT score of the student currently viewing the scattergram. The blue lines represent the average GPA and SAT score for students from the same high school. Naviance updated the scattergram format in 2017, but this new version was not available to most students in the study sample while they were applying to college.

Figure 2: Timeline

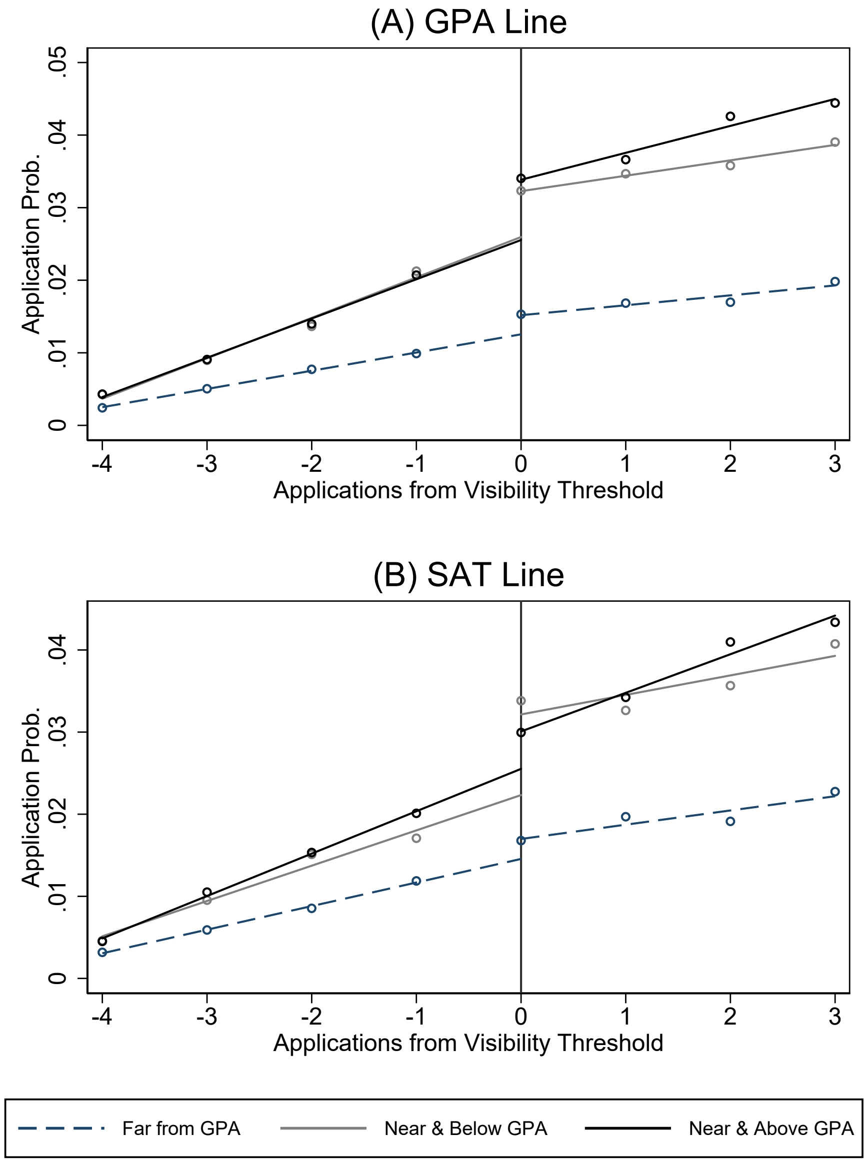
	2014: Naviance purchased (no data for scattergrams)	**2015:** Scattergrams Available (based on 2014 data)	**2016:** Additional Scattergrams and Data Available	**2017:** Additional Scattergrams and Data Available	*Data Available*
Class of **2014**	**Grade 12:** Applying to College	**Entering College/ Labor Force**			
Class of **2015**	**Grade 11:** Taking SAT	**Grade 12:** Applying to College	**Entering College/ Labor Force**		*Applications + NSC Enrollment Records*
Class of **2016**	**Grade 10**	**Grade 11:** Taking SAT	**Grade 12:** Applying to College	**Entering College/ Labor Force**	
Class of **2017**	**Grade 9**	**Grade 10**	**Grade 11:** Taking SAT	**Grade 12:** Applying to College	*Applications + Naviance Login Records*

Figure 3: Impact of Scattergram Visibility on College Applications and Attendance



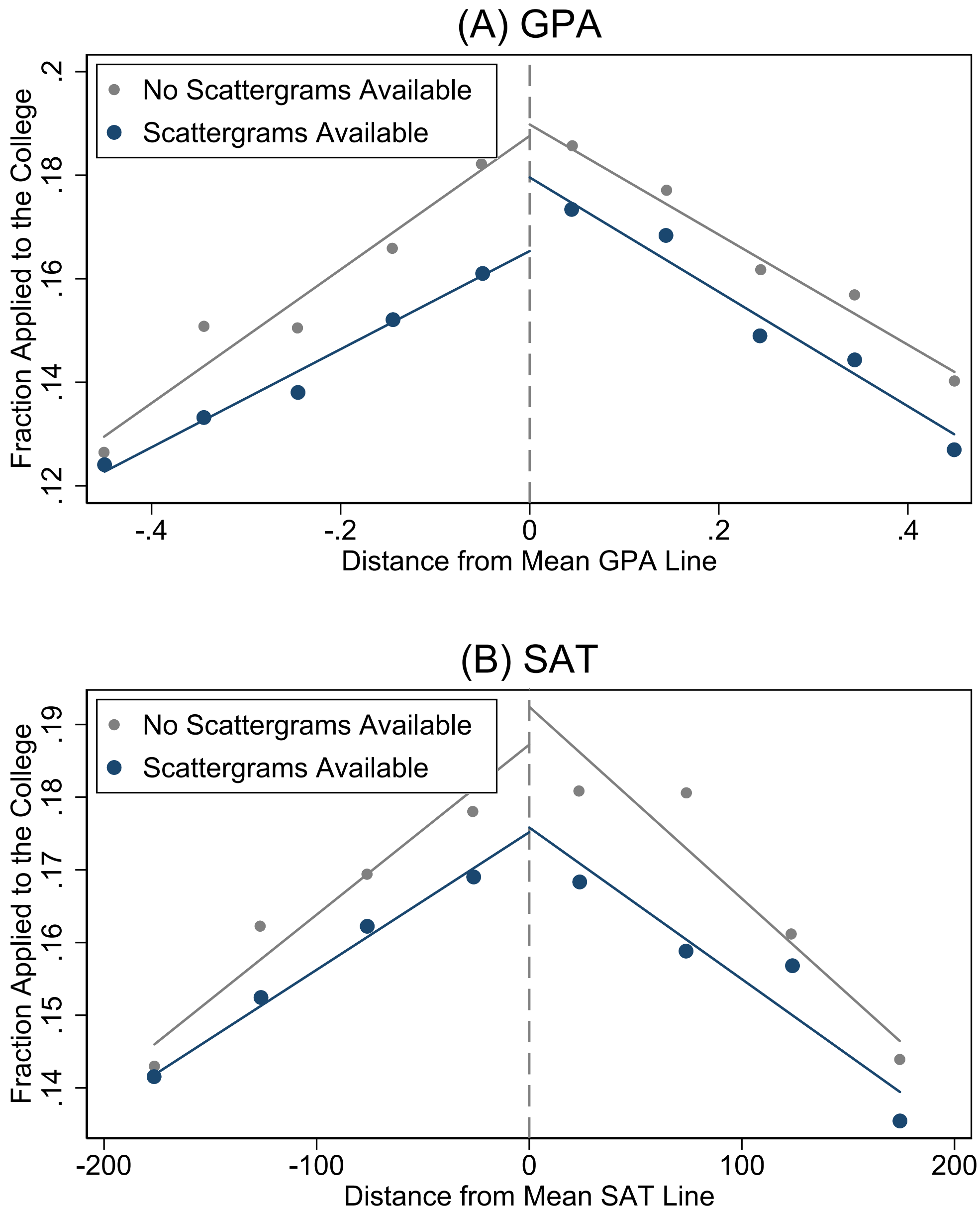
Notes: The figures above show how the probability of applying to (A) or attending (B) a college changes when a college crosses a scattergram visibility threshold, and how this varies based on the student's similarity to previously accepted students (as measured by proximity to the typical acceptee lines). A college's scattergram becomes visible to students after it receives five or ten applications from the student's high school. (I do not know which threshold each high school uses.) The X-axis shows how far a college was, in terms of applications, from each of these minimum applicant thresholds (in 2015 and 2016). Since I use both thresholds, college-high school combinations with 5 to 8 applications in the previous year are included twice in this graph for the same student. Observations are student-college-threshold combinations. The dots on the y-axis represents the fraction of students who applied to (or attended) a college with previous applications x distance from the threshold. The black solid lines are based on all students in the sample. The sizes of the black circles represent the number of observations associated with each bin on the x-axis. The dashed lines break this sample into students who are (or would have been) near and far from the typical acceptee lines. The navy dashed line is based on student-college combinations where students are within .5 GPA points and 150 SAT points of the typical acceptee lines, and the gray dashed lines include the remaining student-college combinations. I computed hypothetical typical acceptee lines for colleges which did not meet the cutoff for a scattergram based on the prior applications, and used these to compute near and far for student-college combinations to the left of the RD threshold. Students to the left of the RD line would not have seen these lines. This is based on weighted 11th grade GPAs and SAT scores on the 2400 scale.

Figure 4: Impact of Scattergram Visibility on Applications by Proximity to Typical Acceptee Lines



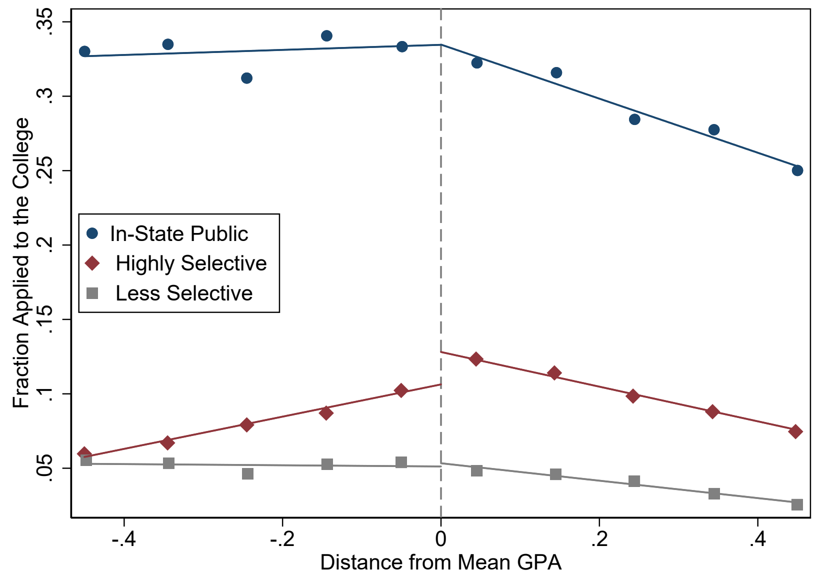
Notes: The figures above show how the probability of a student applying to a college changes when a college crosses a scattergram visibility threshold, and how this varies based on the proximity of the student to the typical acceptee lines. Panel (A) is based on the weighted GPA lines and near is defined as within .1 GPA points. Panel (B) is based on the SAT 2400 scale and near is defined as within 50 SAT points. I computed hypothetical typical acceptee lines for colleges which did not meet the cutoff for a scattergram based on the prior applications and used these to compute near, far, above and below, for student-college combinations to the left of the RD threshold. Students to the left of the RD threshold would not have seen these lines. Observations are student-college-threshold combinations. I used distances to both thresholds (five and ten) where relevant. The X-axis shows how far a college was, in terms of applications, from each of these minimum applicant thresholds (in 2015 and 2016). The dots on the y-axis represents the fraction of students who applied to a college with previous applications x distance from the threshold. The pattern for students who are far from the lines and above them is very similar to that for students who are far from the lines and below them.

Figure 5: Application Probability by Distance from Typical Acceptee Lines



Notes: The figures above show how application rates varied based on a student's position on a scattergram relative to the typical acceptee's GPA (A) and SAT (B). The navy lines are based on students from the years in which scattergrams were available (2015-2016) and the gray lines are based on students in 2014 who could not see any scattergrams. Panel (A) compares the fraction of students applying to a given college with the distance of their GPA from the average weighted GPA of previously admitted students at their high school. Weighted GPAs from 11th grade are used to determine the distance from the mean weighted GPA line depicted on the scattergram when the student is in 12th grade. Panel (B) compares the fraction of students applying with the distance of their SAT from the average SAT of admitted students at their high school. A student with the same SAT as the average admitted student will have a distance of zero. Students' maximum SAT scores on the 2400 scale are used to determine the distance from the mean SAT line on the scattergram. Observations are student-college combinations, and the college in this pair must have received at least 10 applications from the student's high school in 2014 for the observation to be included in this graph. This is the set of scattergrams to which students in 2015 would have certainly had access. The 2014 (no scattergrams) lines are based on a student's distance from the average accepted student in 2014, however these students could not see the average. The peak, for these students, at 0 is partly mechanical because the averages are based on their own choices. For panel (A), the data are binned in intervals of 0.1 from the threshold at zero, and in panel (B) they are binned in 50-point intervals. The y-axis represents the fraction of students in each bin who applied at the college. A bin includes multiple scattergrams and it may include the same students multiple times (but for different scattergrams). The fitted lines come from a local linear regression discontinuity model with a bandwidth of .5 GPA points or 150 SAT points.

Figure 6: Application Probability by Distance from Mean GPA Lines and College Type



Notes: This figure compares the fraction of students who applied to a college with the distance of the student's weighted 11th grade GPA from the typical acceptee line she could see and the type of college. Observations are student-college combinations, and the college in this pair must have received at least ten previous applications from the student's high school to be included in this graph. The data are binned in intervals of 0.1 from the threshold at zero. The fitted lines come from a local linear regression discontinuity model with a bandwidth of 0.5. Colleges are broken into highly selective and less selective categories based on Barron's selectivity ratings. The in-state public colleges are excluded from the selectivity groups so that each student-college combination appears at most once in this figure.

Figure 7: Impact of Individual Scattergrams on Elite College Attendance

Notes: The figure above plots the average impact of a college's typical acceptee GPA line on whether a student attends an elite college. Each dot represents the average impact of an individual college's line (across all the high schools). Elite colleges are the public and private colleges defined as "Elite" by Barron's *Profiles of American Colleges*. The x-axis represents the average location of the college's weighted GPA line, across all high schools in the district.

Table 1: Summary Statistics

		Year		
	Sample	2014	2015	2016
	(1)	(2)	(3)	(4)
(A) Demographics				
White/Asian	0.65	0.68	0.66*	0.65**
Black/Hispanic	0.28	0.26	0.27	0.29***
Free/Reduced Lunch	0.21	0.19	0.21**	0.22**
(B) Academics				
GPA (11th gr. weighted)	3.41	3.41	3.42	3.40
SAT(M+V+W)	1689	1698	1695	1683
Attend 4-yr Coll	0.60	0.59	0.59	0.60
Attend 2-yr Coll	0.24	0.23	0.24	0.25
Persist in Coll	0.78	0.79	0.78	.
(C) Applications				
Number of Apps	5.15	5.21	5.17	5.13
Num Reach Apps	1.53	1.59	1.48*	1.59
Num Match Apps	2.31	2.40	2.34	2.29*
Num Safety Apps	1.29	1.30	1.31	1.27
Acceptances	2.51	2.35	2.55***	2.47***
(D) Attendance				
Reach	0.19	0.18	0.18	0.19
Match	0.54	0.56	0.55	0.54
Safety	0.27	0.26	0.28*	0.27
(E) Scattergrams				
Total	47	0	33	62
In GPA Bandwidth	18	0	12	22
In SAT Bandwidth	11	0	8	14
Relevant	21	0	15	26
N	7,647	3,758	3,733	3,914

Notes: Column 1 contains the full sample of students (who are in the 2015 and 2016 cohorts). They all appear in the scattergram introduction regressions. Column (2) contains all students who graduated from the district in 2014. These students could not see any scattergrams. Columns (3) and (4) contain students who graduated in 2015 and 2016, respectively. They could see the scattergrams and column (1) is a weighted average of these columns. The stars indicate the statistical significance from a t-test for a difference in means between students in 2014 and those in 2015 or 2016, who could see scattergrams. (*$p<.10$ **$p<.05$ *** $p<.01$). Free/reduced lunch is an indicator for students who ever received free or reduced-price lunch while enrolled in the district. Students who indicate two or more races are excluded from the race categories in Panel (A). GPA refers to 11th grade weighted GPA and SAT refers to the maximum SAT on the old 2400 scale. New SAT scores have been converted to the old 2400 scale. Scattergrams refers to the minimum number of scattergrams to which a student had access based on her graduation year and high school. It is the number of colleges with at least 10 prior applicants. If a college was using the minimum of five applicants, more scattergrams would have been visible. Attend 4-yr college is an indicator for whether the student attended a four-year college within six months of graduating high school. Attend 2-yr is similarly defined but for two-year colleges. Reach schools are colleges at which the student's maximum SAT score is below the 25th percentile of accepted students' SATs, as reported to IPEDS in 2015 by the college. Match schools are colleges at which the student's maximum SAT score is within the interquartile range of accepted students' SATs. Safety schools are colleges at which the student's maximum SAT score is above the 75th percentile of accepted students' SATs.

Table 2: Impact of Scattergrams on Applications and Attendance by Proximity to Typical Acceptee

		Near GPA		Near SAT		Near Both	Near
	All	.5	.1	150	50	.5 & 150	Neither
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(A) Applied							
Visible	0.0027***	0.0040***	0.0060**	0.0038***	0.0048**	0.0056***	0.0024***
	(0.0004)	(0.0010)	(0.0025)	(0.0013)	(0.0023)	(0.0017)	(0.0004)
CCM	0.0137	0.0228	0.0268	0.0252	0.0267	0.0278	0.0083
(B) Attended							
Visible	0.0001	0.0004	-0.0005	0.0000	0.0003	0.0001	0.0001
	(0.0001)	(0.0003)	(0.0007)	(0.0004)	(0.0007)	(0.0005)	(0.0001)
CCM	0.0011	0.0024	0.0031	0.0028	0.0030	0.0034	0.0004
N	2,565,375	666,731	132,319	432,073	153,384	272,995	1,739,515

Notes: Heteroskedasticity robust standard errors clustered by student are in parentheses. (*p<.10 **p<.05 *** p<.01). All regressions include fixed effects for each student and college by year. Observations are all student-college-threshold combinations for which the college was within four applications of the threshold at five or ten and the college received at least one application from the student's high school between 2014 and 2016. Student-college combinations are included twice for colleges with five to eight prior applications since they fall in the bandwidth for both thresholds. GPAS are weighted and are are on a five point scale. The SAT scores are on the 2400 scale. CCM refers to the mean application or attendance probability predicted at a college at the threshold if its scattergram had not been made visible.

Table 3: Heterogeneity in Impacts of Scattergram Visibility

		Free/Reduced Lunch		White or	Black or	In-St. Public	Other Colleges	
	All	Never	Ever	Asian	Hispanic	Colleges	High Sel.	Less Sel.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) All Students								
Visible	0.0027***	0.0026***	0.0016*	0.0025***	0.0027***	0.0098***	0.0027***	0.0021***
	(0.0004)	(0.0005)	(0.0009)	(0.0005)	(0.0008)	(0.0030)	(0.0007)	(0.0006)
CCM	0.0137	0.0146	0.0121	0.0136	0.0144	0.0185	0.0159	0.0106
N	2,565,375	2,031,177	534,198	1,696,273	708,692	63,947	1,001,540	1,499,888
(B) Near Lines								
Visible	0.0056***	0.0048***	0.0115**	0.0035*	0.0124***	-0.0085	0.0056*	0.0056***
	(0.0017)	(0.0018)	(0.0052)	(0.0019)	(0.0046)	(0.0257)	(0.0029)	(0.0020)
CCM	0.0278	0.0279	0.0287	0.0270	0.0340	0.0782	0.0375	0.0178
N	272,995	242,939	29,994	210,107	46,767	2,803	105,607	162,455

Notes: Heteroskedasticity robust standard errors clustered by student are in parentheses. (*p<.10 **p<.05 *** p<.01). All regressions include fixed effects for each student and college by year. Observations are all student-college-threshold combinations for which the college was within four applications of the threshold at five or ten and the college received at least one application from the student's high school between 2014 and 2016. CCM refers to the mean application probability predicted at a college at the threshold if its scattergram had not been made visible. Near is defined as within .5 GPA points or 150 SAT points. This is based on weighted GPA points and SAT points on the old 2400 scale. Column (2) is based on students who never received free or reduced-price lunch from the district and column (3) contains students who received it at least once. Students who indicate two or more races or report a race that is not white, Black, Asian, or Hispanic are excluded from columns (4) and (5). In-state public colleges are excluded from the highly and less selective college categories in columns (7) and (8). Selectivity ratings are based on Barron's 2009 selectivity index, and where missing, selectivity rankings from IPEDS in 2002.

Table 4: Impact of Scattergram Visibility on Attendance

		Free/Reduced Lunch		White or	Black or	In-St. Public	Other Colleges	
	All	Never	Ever	Asian	Hispanic	Colleges	High Sel.	Less Sel.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) All Students								
Visible	0.0001	0.0001	0.0003	-0.0001	0.0006**	0.0028**	0.0001	0.0002
	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.0014)	(0.0002)	(0.0002)
CCM	0.0011	0.0013	0.0008	0.0012	0.0011	0.0022	0.0009	0.0013
N	2,565,375	2,031,177	534,198	1,696,273	708,692	63,947	1,001,540	1,499,888
(B) Near Lines								
Visible	0.0001	-0.0004	0.0026*	-0.0009	0.0047***	0.0010	0.0001	0.0001
	(0.0005)	(0.0006)	(0.0015)	(0.0006)	(0.0015)	(0.0087)	(0.0008)	(0.0007)
CCM	0.0034	0.0037	0.0023	0.0037	0.0024	0.0192	0.0033	0.0030
N	272,995	242,939	29,994	210,107	46,767	2,803	105,607	162,455

Notes: Heteroskedasticity robust standard errors clustered by student are in parentheses. (*p<.10 **p<.05 *** p<.01). Regressions include fixed effects for each student and college by year. Observations are student-college-threshold combinations. CCM refers to the mean attendance probability predicted at a college at the threshold if its scattergram had not been made visible. Near is defined as within .5 GPA points or 150 SAT points. This is based on weighted GPA points and SAT points on the old 2400 scale. Column (2) is based on students who never received free or reduced-price lunch from the district and column (3) contains students who received it at least once. Students who indicate two or more races or report a race that is not white, Black, Asian, or Hispanic are excluded from columns (4) and (5). In-state public colleges are excluded from the highly and less selective college categories in columns (7) and (8). Selectivity ratings are based on Barron's 2009 selectivity index, and where missing, selectivity rankings from IPEDS in 2002.

Table 5: Impact of All Scattergrams on Applications based on College Fixed Effects

		Free/Reduced Lunch		White or	Black or	In-St. Public	Other Colleges	
	All	Never	Ever	Asian	Hispanic	Colleges	High Sel.	Less Sel.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) All Students								
Visible	0.0092***	0.0084***	0.0040***	0.0075***	0.0053***	0.0225***	0.0082***	0.0057***
	(0.0004)	(0.0004)	(0.0006)	(0.0005)	(0.0006)	(0.0022)	(0.0005)	(0.0004)
CCM	0.0079	0.0086	0.0058	0.0081	0.0075	0.0096	0.0094	0.0062
N	8,914,720	7,018,780	1,895,940	5,844,978	2,503,108	300,304	2,939,362	5,062,656
(B) Near Lines								
Visible	0.0104***	0.0097***	0.0109***	0.0089***	0.0080***	0.0109	0.0097***	0.0071***
	(0.0009)	(0.0009)	(0.0027)	(0.0010)	(0.0024)	(0.0100)	(0.0013)	(0.0011)
CCM	0.0180	0.0181	0.0174	0.0172	0.0207	0.0228	0.0251	0.0119
N	583,508	520,768	62,740	451,196	98,194	27,742	233,076	303,676

Notes: Heteroskedasticity robust standard errors clustered by student are in parentheses. (*p<.10 **p<.05 *** p<.01). All regressions include fixed effects for each student and college by year. Observations are all student-college-threshold combinations for which the college received at least one application from the student's high school between 2014 and 2016. Near is defined as within .5 GPA points or 150 SAT points. This is based on weighted GPA points and SAT points on the old 2400 scale. Column (2) is based on students who never received free or reduced-price lunch from the district. Column (3) contains all students who received it at least once while enrolled in the district. Students who indicate two or more races or report a race that is not white, Black, Asian, or Hispanic are excluded from columns (4) and (5). The in-state public colleges are excluded from the highly and less selective college categories in columns (7) and (8). CCM refers to the mean application probability predicted at a college without a scattergram. Selectivity ratings are based on Barron's 2009 selectivity index. Where this is missing, selectivity rankings from IPEDS in 2002 are used.

Table 6: Impact of Mean GPAs on Applications

		Free/Reduced Lunch		White or	Black or	In-St. Public	Other Colleges	
	All	Never	Ever	Asian	Hispanic	Colleges	High Sel.	Less Sel.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Below GPA	-0.011***	-0.013***	0.015	-0.013***	0.014	0.013	-0.019***	0.001
	(0.004)	(0.004)	(0.011)	(0.004)	(0.009)	(0.012)	(0.005)	(0.005)
Dist Above GPA	-0.103***	-0.110***	-0.090***	-0.121***	-0.020	-0.026	-0.121***	-0.016
	(0.011)	(0.011)	(0.029)	(0.012)	(0.027)	(0.040)	(0.019)	(0.013)
Dist Below GPA	-0.085***	-0.078***	-0.118***	-0.081***	-0.112***	-0.169***	-0.075***	-0.033**
	(0.010)	(0.010)	(0.024)	(0.012)	(0.020)	(0.043)	(0.015)	(0.016)
CCM	0.139	0.139	0.143	0.141	0.131	0.322	0.123	0.048
N	110,013	99,304	11,628	85,875	26,522	19,081	43,719	39,620

Notes: Heteroskedasticity robust standard errors clustered by student are in parentheses. (*p<.10 **p<.05 *** p<.01). College by year and high school fixed effects are included, as well as controls for 11th grade GPA, maximum SAT score, gender, sepcial education, and dummy variables for race categories and ever receiving free or reduced-price lunch. Optimal bandwidths for each regression are calculated as described in Calonico, Cattaneo and Titiunik (2014). All estimates are for weighted GPAs on a five point scale. The outcome is applying to the college associated with the scattergram treating the student. N refers to the number of student-scattergram combinations on which the regression is based. Column (2) is based on students who never received free or reduced-price lunch from the district and column (3) contains all students who received it. Students who indicate two or more races or report a race that is not white, Black, Asian, or Hispanic are excluded from columns (4) and (5). In-state public colleges are excluded from the highly and less selective college categories in columns (7) and (8). Selectivity ratings are based on Barron's 2009 selectivity index, and where missing, selectivity rankings from IPEDS in 2002. CCM refers to the mean application probability for students with GPAs just above the typical acceptee's.

Table 7: Impact of Mean SATs on Applications

		Free/Reduced Lunch		White or	Black or	In-St. Public	Other Colleges	
	All	Never	Ever	Asian	Hispanic	Colleges	High Sel.	Less Sel.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Below SAT	0.0040	0.0045	-0.0033	0.0032	0.0143	0.0111	0.0005	0.0108**
	(0.0038)	(0.0040)	(0.0112)	(0.0043)	(0.0100)	(0.0110)	(0.0059)	(0.0047)
Dist Above SAT	-0.0002***	-0.0002***	-0.0003***	-0.0002***	-0.0002*	-0.0003***	-0.0001*	0.0000
	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0001)	(0.0001)	(0.0001)	(0.0000)
Dist Below SAT	-0.0001***	-0.0001***	-0.0001	-0.0001***	-0.0002**	-0.0002**	-0.0001	-0.0001***
	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0001)	(0.0001)	(0.0001)	(0.0000)
CCM	0.131	0.129	0.157	0.133	0.119	0.313	0.110	0.050
N	97,226	92,766	11,294	105,567	16,012	15,082	49,691	29,159

Notes: Heteroskedasticity robust standard errors clustered by student are in parentheses. (*p<.10 **p<.05 *** p<.01). College by year and high school fixed effects are included, as well as controls for 11th grade GPA, maximum SAT score, gender, special education, and dummy variables for race and ever receiving free or reduced-price lunch. Optimal bandwidths for each regression are calculated as described in Calonico, Cattaneo and Titiunik (2014). All estimates are for SAT scores on the 2400 scale. New scores have been converted to old ones where relevant.The outcome is applying to the college associated with the scattergram treating the student. N refers to the number of student-scattergram combinations on which the regression is based. Column (2) is based on students who never received free or reduced-price lunch from the district and column (3) contains all students who received it. Students who indicate two or more races or report a race that is not white, Black, Asian, or Hispanic are excluded from columns (4) and (5). In-state public colleges are excluded from the highly and less selective college categories in columns (7) and (8). Selectivity ratings are based on Barron's 2009 selectivity index, and where missing, selectivity rankings from IPEDS in 2002. CCM refers to the mean application probability for students with SATs just above the typical acceptee's.

Table 8: Cumulative Impact of Scattergrams

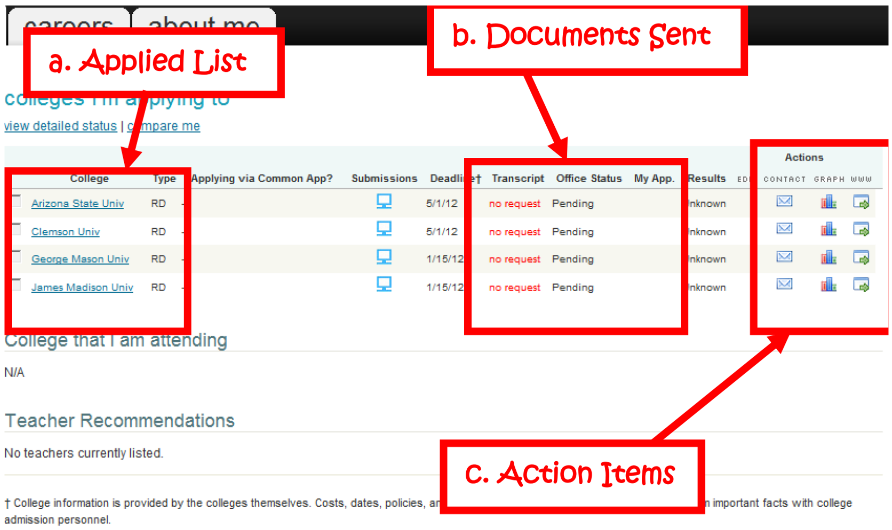
	Applications				Acceptances	Attend College				
	Num.	Reach	Match	Safety		Reach	Match	Safety	Four-yr	Persist
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Total SGs	-0.010	-0.028***	0.029***	-0.011***	0.001	-0.003***	0.005***	-0.003***	-0.001	0.001
	(0.008)	(0.004)	(0.005)	(0.004)	(0.005)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Reach SGs	0.006	0.128***	0.028	-0.150***	0.014	0.019***	0.010**	-0.032***	-0.002	0.004
	(0.037)	(0.020)	(0.023)	(0.017)	(0.022)	(0.003)	(0.005)	(0.003)	(0.004)	(0.005)
Match SGs	-0.006	-0.029***	0.072***	-0.051***	0.002	-0.003***	0.010***	-0.009***	-0.001	0.002
	(0.012)	(0.006)	(0.007)	(0.005)	(0.007)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Safety SGs	-0.028*	-0.078***	-0.035***	0.083***	-0.003	-0.011***	-0.001	0.011***	-0.000	-0.001
	(0.016)	(0.009)	(0.010)	(0.007)	(0.009)	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)
In-St Public SGs	0.036	-0.089***	0.164***	-0.042*	0.056*	-0.014***	0.034***	-0.011**	0.008	0.012*
	(0.053)	(0.029)	(0.033)	(0.024)	(0.031)	(0.005)	(0.007)	(0.005)	(0.005)	(0.007)
N	5,176	5,176	5,176	5,176	5,176	5,176	5,176	5,176	5,176	2,466

Notes: Heteroskedasticity robust standard errors are in parentheses. (*p<.10 **p<.05 *** p<.01). High school and year fixed effects are included. I control for academic achievement using fixed effects for 50 point intervals of maximum SAT scores, and .1 point intervals of students' weighted 11th grade GPAs. Controls include demographic indicators for race (white, asian, black or hispanic), free-or-reduced price lunch, special education, and gender. There is one observation per student. Persistence refers to persistence into a second year of college. These data are only available for students who graduated high school in 2015. Reach schools are colleges at which the student's maximum SAT score is below the 25th percentile of accepted students' SATs, as reported in IPEDS in 2015. Match schools are colleges at which the student's maximum SAT score is within the interquartile range of accepted students' SATs. Safety schools are colleges at which the student's maximum SAT score is above the 75th percentile of accepted students' SATs. Acceptances are self-reported but I corrected the self reports if a student attended a college where an acceptance decision was not reported. I assume a student must have been accepted to a college if she attends the college.

Appendix

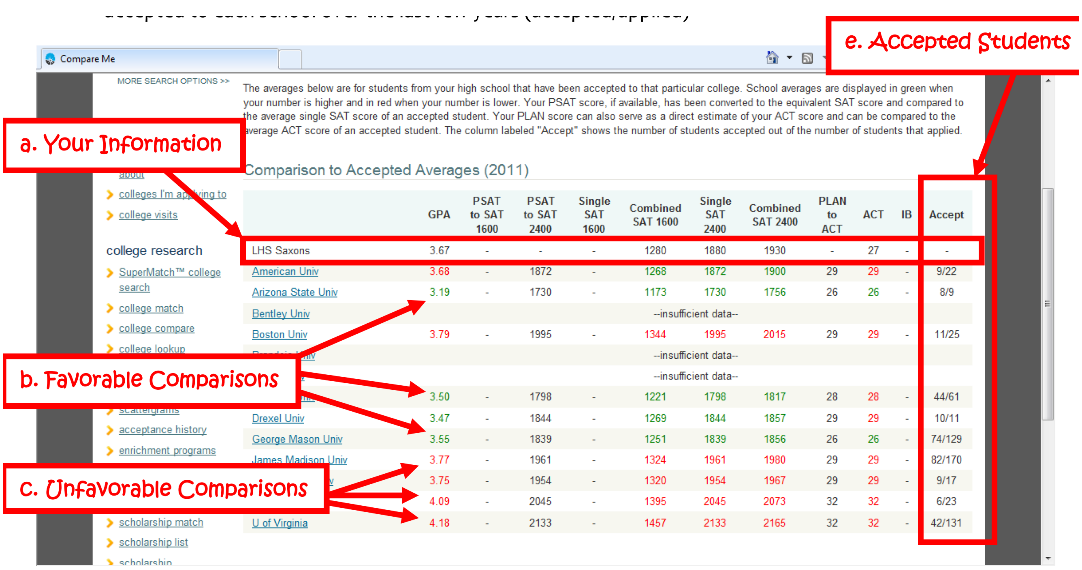
A Additional Tables and Figures

Figure A.1: Example of College Dashboard on Naviance



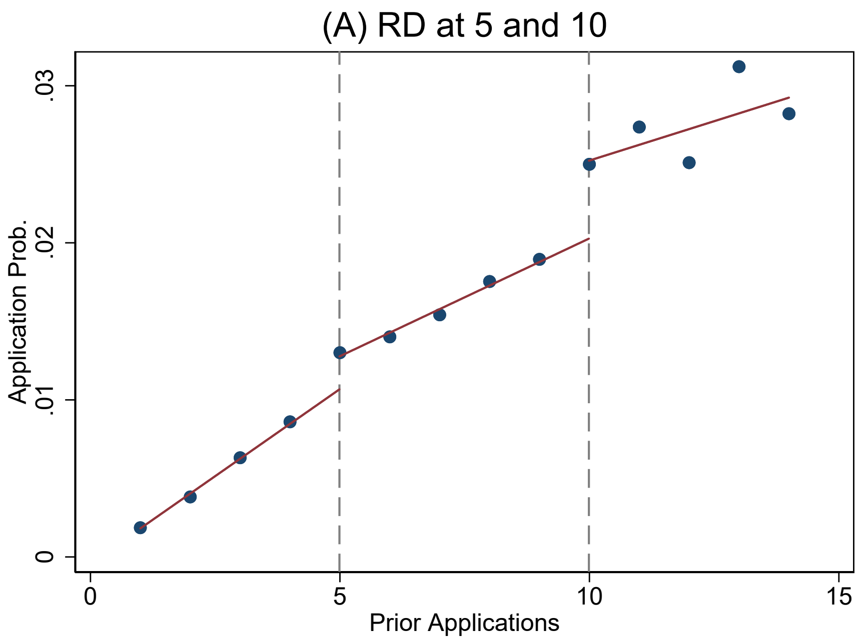
Notes: This is an example of the college dashboard on Naviance. (This is not from the district studied). The red boxes were added as notes by the high school posting instructions on how to use Naviance. Source: Langley High School and Naviance.

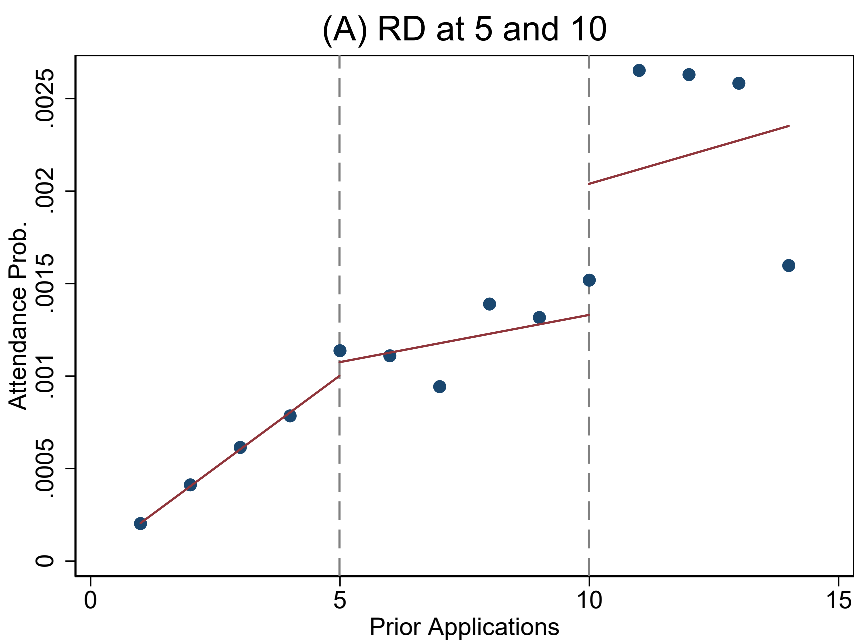
Figure A.2: Example of College Comparisons on Naviance



Notes: This is an example of the how students can compare colleges on Naviance. (This is not from the district studied). The red boxes were added as notes by the high school posting instructions on how to use Naviance. Source: Langley High School and Naviance.

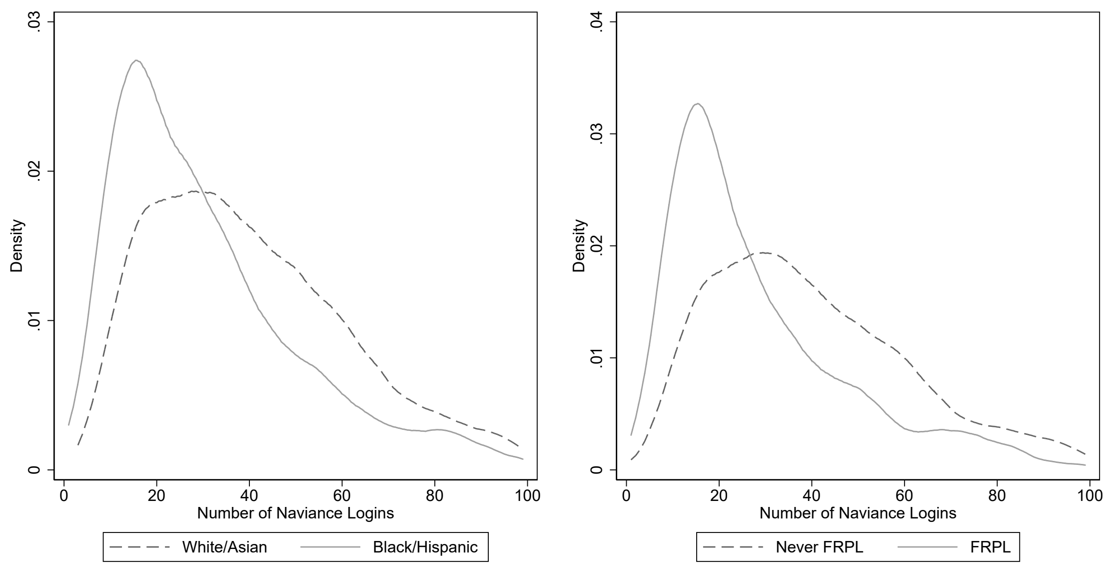
Figure A.3: Discontinuities at 5 and 10 Prior Applications





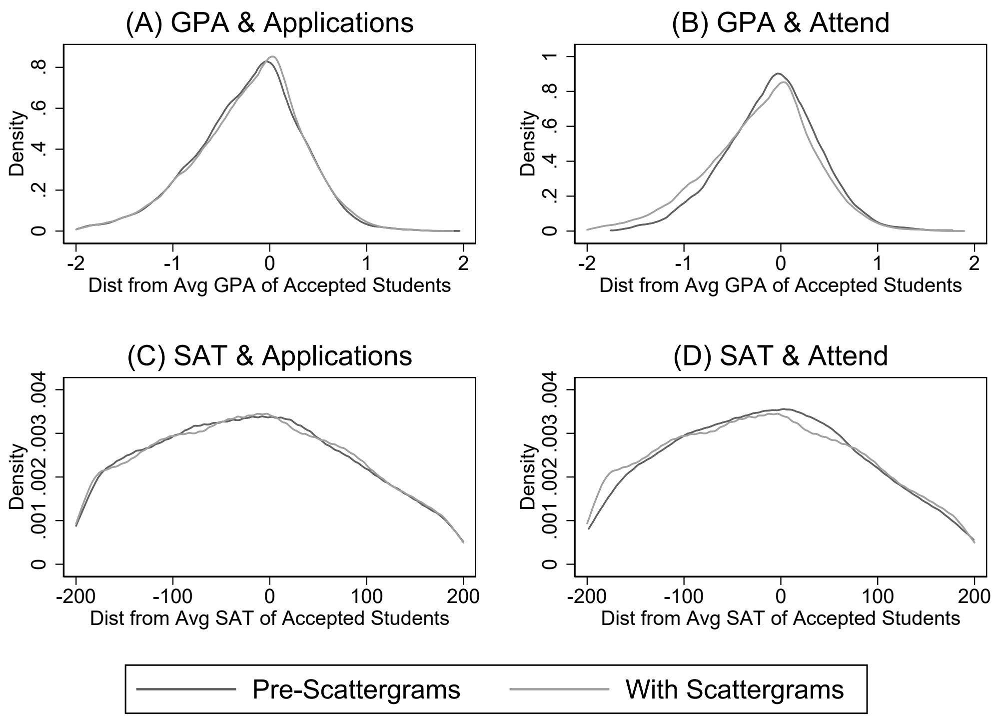
Notes: The figures above show how the probability of a student applying to (A) or attending a college (B) changes based on the number of applications a college has received from the student's high school prior to the student's year of graduation (2015 or 2016). Each dot on the x-axis represents the exact number of applications sent from the high school. On the y-axis, the dot indicates the average fraction of students who applied to or attended each of the colleges with the associated number of prior applications. The fitted lines are from a local linear regression discontinuity model. The graph includes all student-college combinations for which at least one and fewer than fourteen applications from the student's high school were sent to the college since 2014 and prior to the student's graduation year. Scattergrams become visible when a college has received five or ten applications. High schools choose which visibility threshold to use but I could not determine which threshold applied to which high school.

Figure A.4: Density of Naviance Login Rates for the Class of 2017



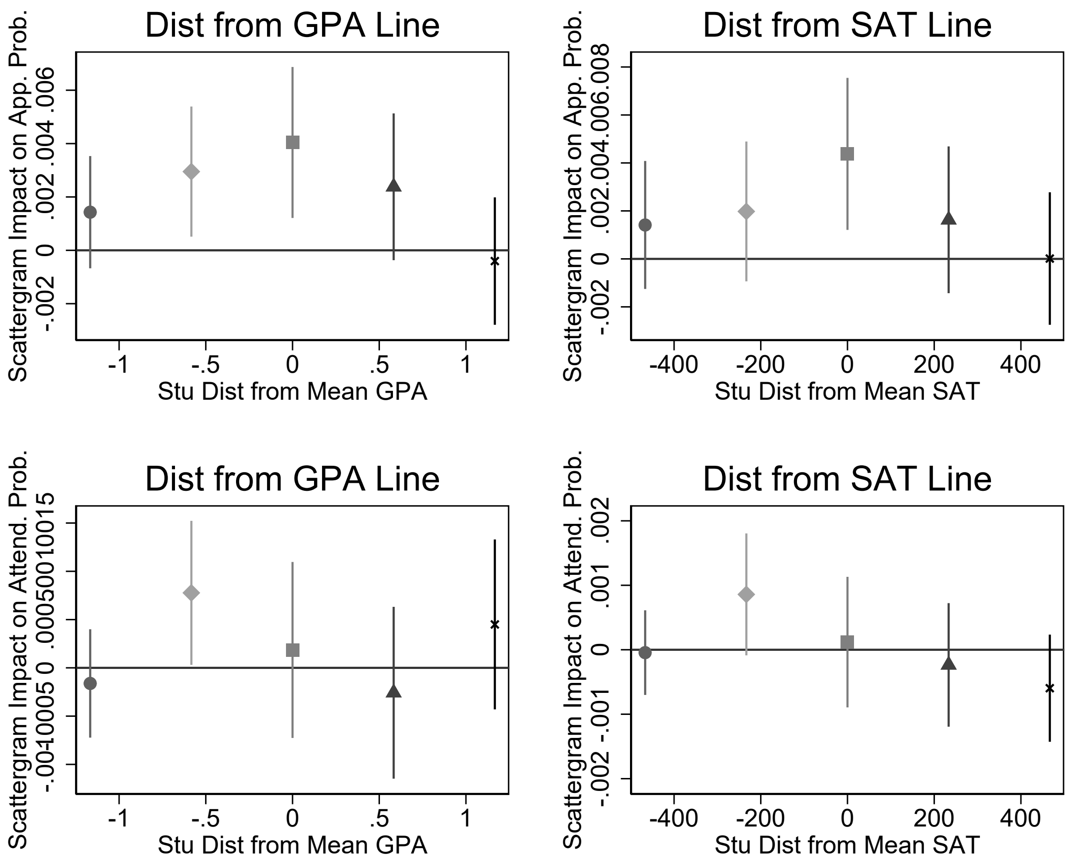
Notes: The figures above indicate the densities for the number of times students logged onto Naviance. These login rates are only available for the class of 2017 but they include all logins since the fall of 2014. (These include all logins through the student's account and thus may capture parents or other individuals logging onto Naviance.) The panel on the left compares the login rates of Black and Hispanic students to those for white and Asian students. The panel on the right compares login rates for students who never received free or reduced-price lunch to students who received it at least one year while enrolled in the district.

Figure A.5: Application and Attendance Density by Distance from Mean GPA or SAT



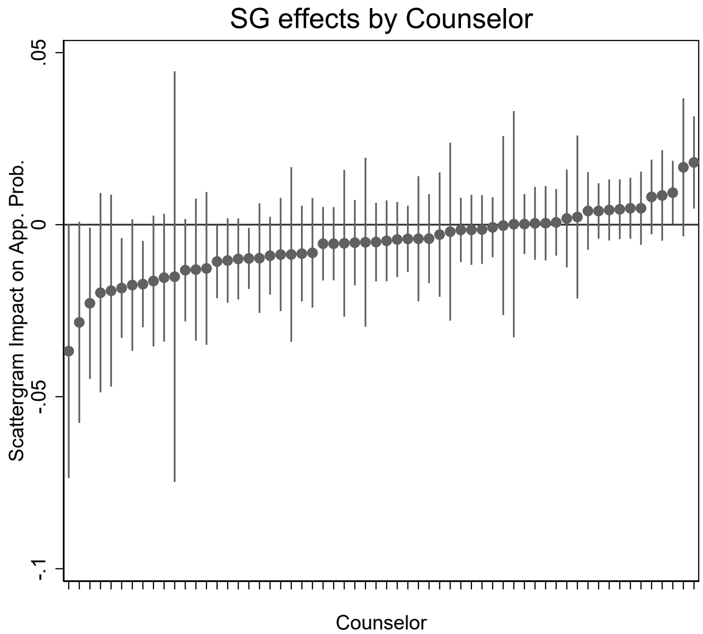
Notes: The figures above show how the types of colleges to which students applied or attended shifted when scattergrams became available. In particular, they show the densities of applications (A and C) and attendance (B and D) as a function of the student's GPA or SAT distance from the average GPA or SAT, of all admitted students in the district, at the college to which they apply or attend. I use the district-wide averages because these may be a more accurate measure of the college's admissions criteria than the school averages, especially for colleges with only a few admitted students from a high school. This also enables me to calculate averages for colleges with only a few or no students admitted at some high schools. The figures are based on weighted GPAs and SAT scores on the 2400 scale. The "Pre-Scattergrams "line is based on the students graduating high school in 2014 and the "With Scattergrams" line is based on students graduating in 2015 and 2016. Students who graduated in 2014 could not see scattergrams.

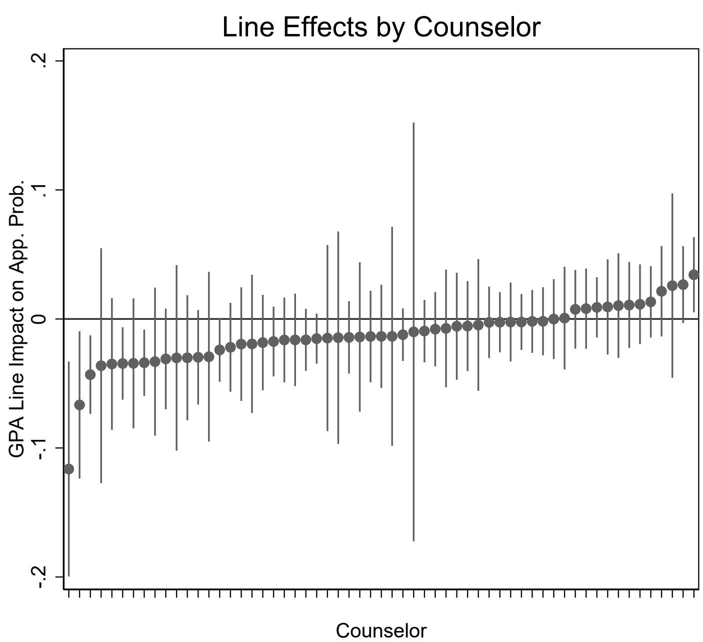
Figure A.6: Magnitude of Scattergram Impact by Proximity to Typical Acceptee



Notes: The above figures show how the magnitudes of the discontinuities in application or attendance probabilities at the visibility thresholds vary based on how similar the students were to the typical acceptee's from their high school. I construct these by separately estimating discontinuities for students who are of varying distances from the typical acceptee's GPA and SAT. For the GPA, I bin students in .5 GPA intervals, starting with students who are within .25 GPA points of typical acceptee. For the SAT I use bins of 150. This is based on weighted GPAs and SAT scores on the 2400 scale. For students who could not see a scattergram, I calculate how far a student would have been from the typical acceptee line based on prior applications. The middle dot in panel (A) indicates, for students whose GPA was within .25 GPA points of the typical acceptee's, how much more likely they were to apply to a college if they could see its scattergram compared to similar students who could not see the scattergram. The bars indicate the standard errors of the discontinuity estimates (where standard errors are clustered by student). These estimates are based on regressions which include student and college by year fixed effects.

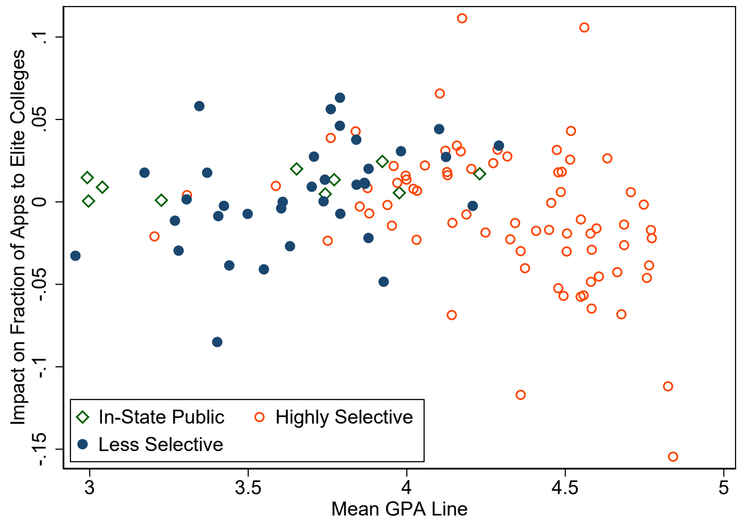
Figure A.7: Effect Sizes by Counselor





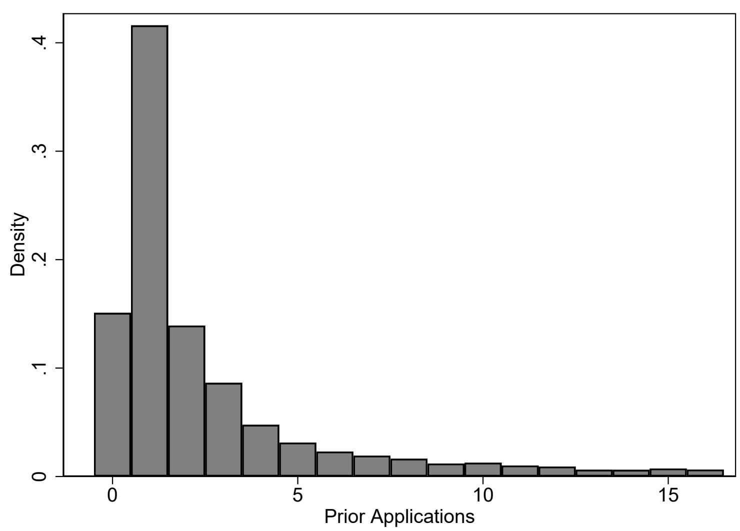
Notes:The figures above show the interactions between counselor effects and scattergram effects. Panel (A) shows how the effect of gaining access to a scattergram varies across counselors. Panel (B) shows how the impact of being below the GPA line varies across counselors. Counselors are assigned by last name so this variation may be due to differences in counseling practices rather than sorting into counselors.

Figure A.8: Impact of Individual Scattergrams on Elite College Applications



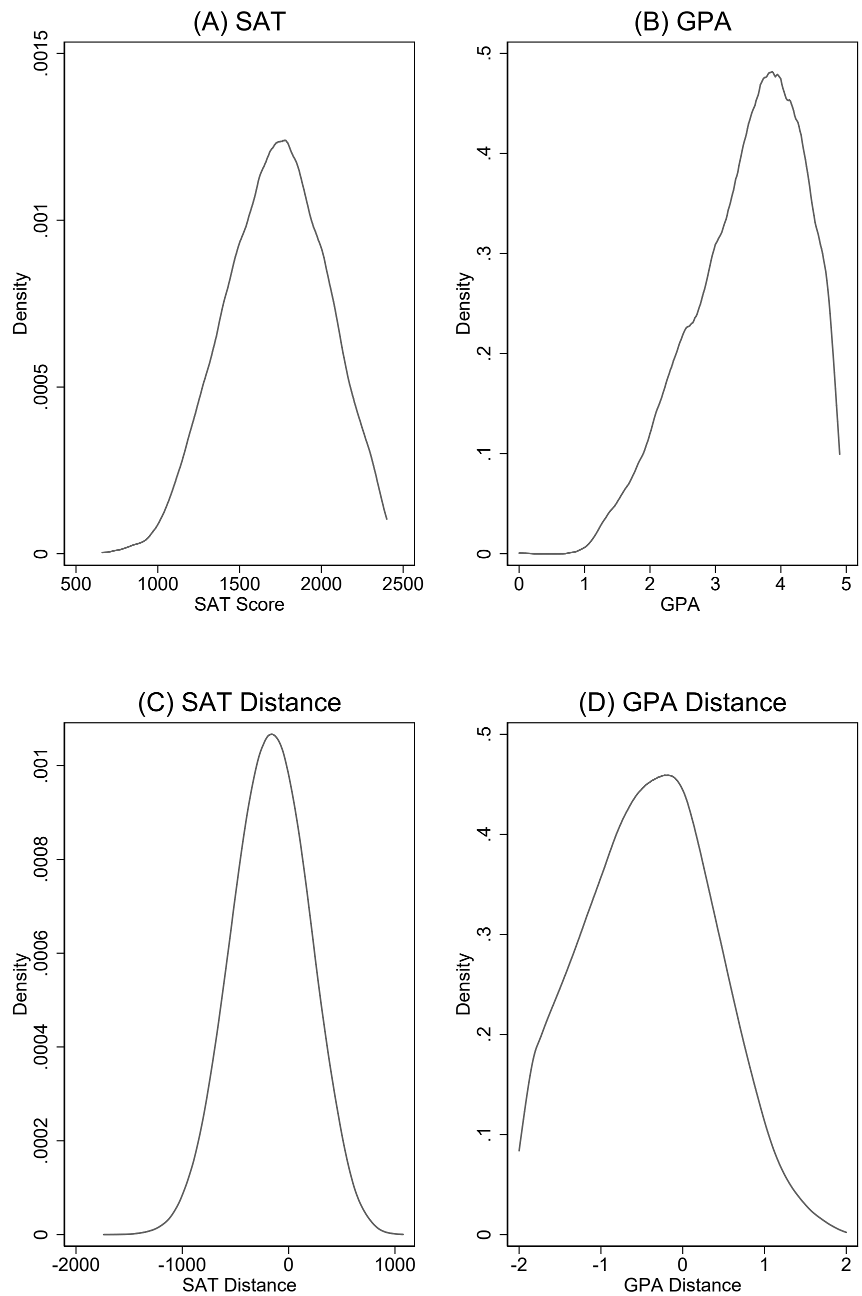
Notes: The figure above plots the average impact of a college's typical acceptee GPA line on the fraction of elite colleges to which a student applies. Each dot represents the average impact of an individual college's line (across all the high schools). Elite colleges are the public and private colleges defined as "Elite" by Barron's *Profiles of American Colleges*. The x-axis represents the average location of the college's weighted GPA line, across all high schools in the district.

Figure A.9: Density of Prior Applications



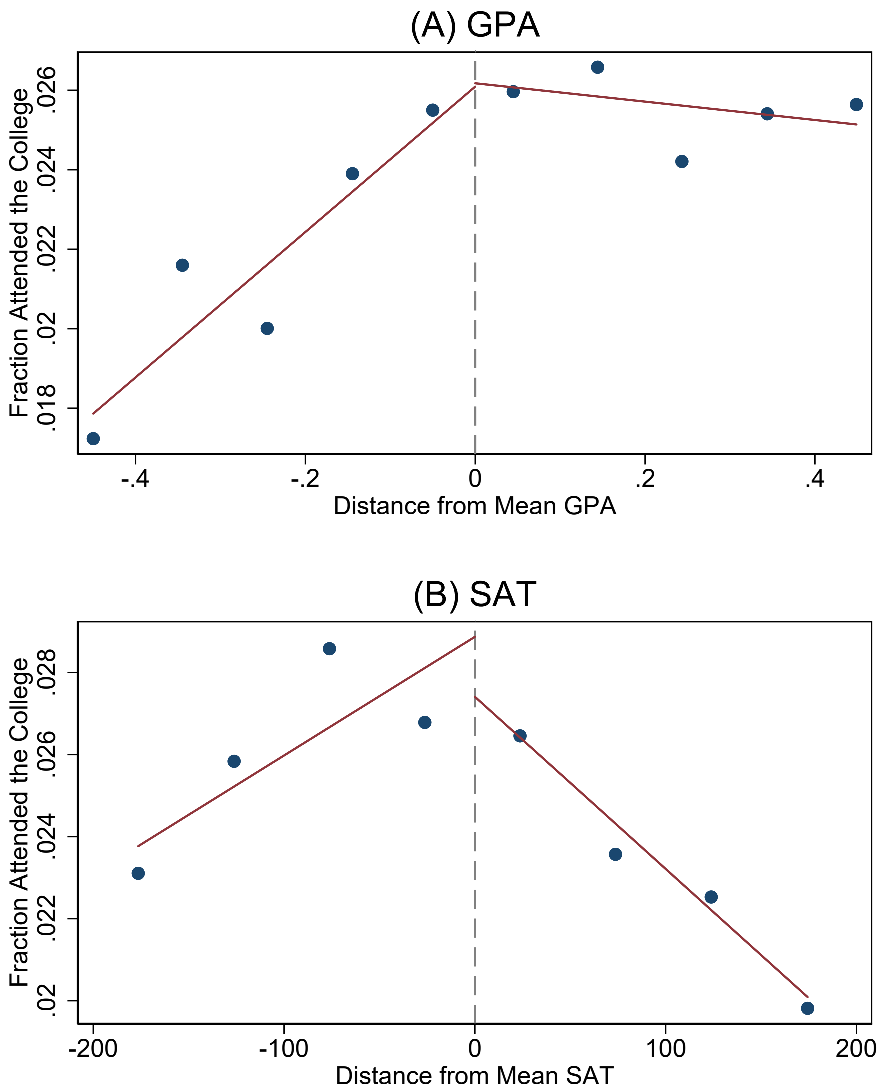
Notes: The figure above depicts the density of prior applications received by colleges. Prior applications refers to the cumulative number of applications received by a college from a high school since 2014 but prior to the current year. For each high school, it includes the set of colleges which received an application from that high school between 2014 and 2016.

Figure A.10: Densities of SAT scores and GPAs



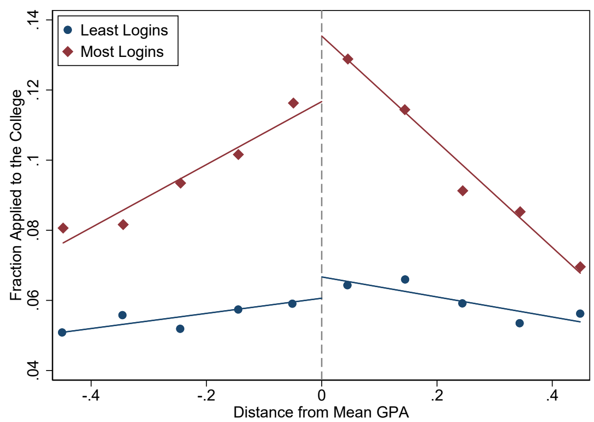
Notes: The top row shows the densities for SAT scores (the old version on the 2400 scale) in panel (A) and weighted 11th grade GPAs in panel (B). The bottom row shows the densities for the distance of the student's SAT score (old 2400 version) and weighted GPA from the typical acceptee's SAT or GPA on the scattergram. Scores on the new version of the SAT have been converted to the old version equivalent score using the scale provided by the College Board. Maximum SAT scores are used in these figures. The figures are based on student-scattergram combinations since the same student has a different distance value for each scattergram. Thus, students may appear multiple times in each figure. There is no statistically distinguishable evidence of heaping on either side of the mean SAT or GPA lines.

Figure A.11: Attendance Probability by Distance from Typical Acceptee Lines



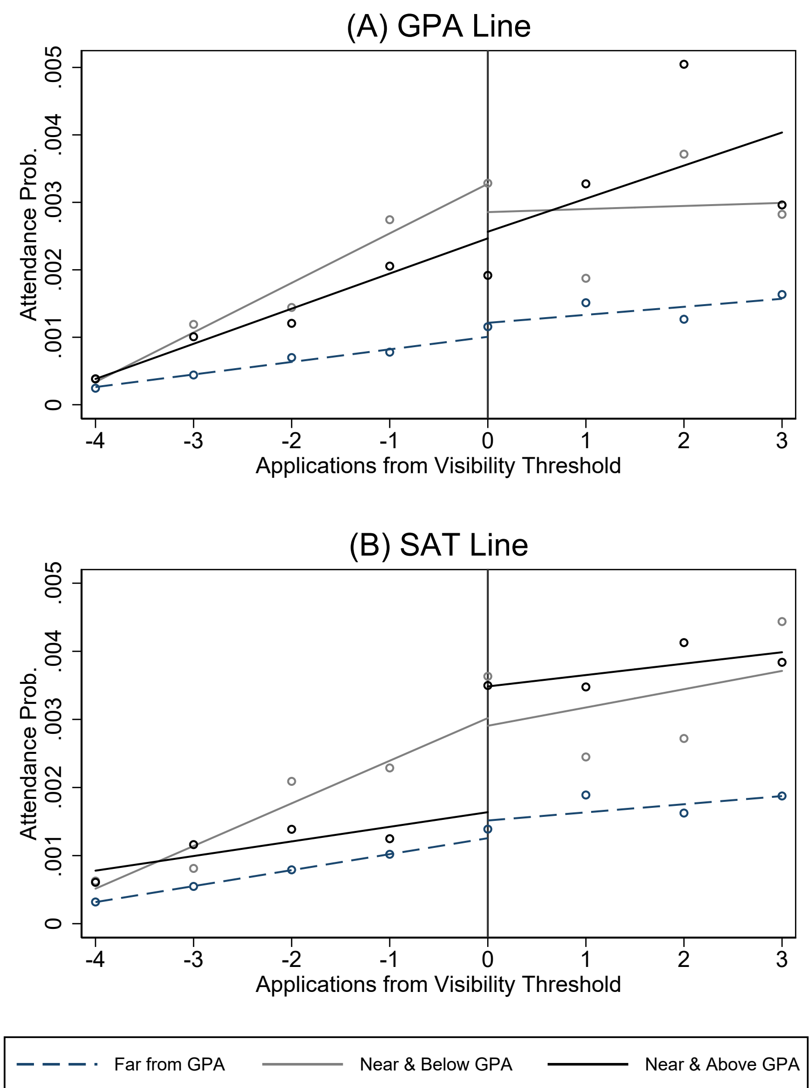
Notes: The figures above show how attendance rates varied based on a student's position on a scattergram relative to the typical acceptee's GPA (in panel A) and SAT (in panel B). For panel (A), Weighted GPAs from 11th grade are used to determine the distance from the mean weighted GPA line depicted on the scattergram when the student is in 12th grade. The data are binned in intervals of 0.1 from the threshold at zero. For panel (B), students' maximum SAT scores on the old 2400 scale are used. The data are binned in intervals of 50 from the threshold at zero. The fitted lines come from a local linear regression discontinuity model with a bandwidth of 0.5 for panel (A) and a bandwidth of 200 for panel (B). The y-axis represents the fraction of students in each bin who attended the college (in 2015 or 2016). A bin includes multiple scattergrams (and colleges) and it may include the same students multiple times (but for different scattergrams). Observations are student-scattergram combinations (where scattergrams are based on colleges which received at least 10 prior applications).

Figure A.12: Application Probability by Distance from Mean GPA and Naviance Logins



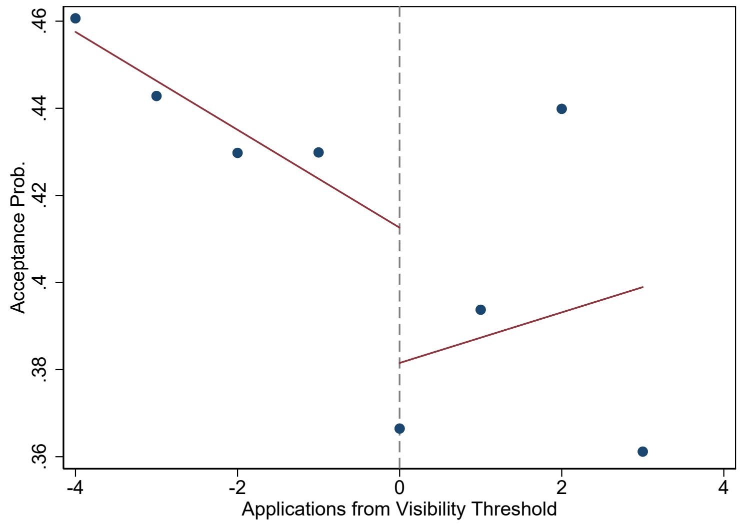
Notes: The figure above shows how application rates varied for the class of 2017 based on a student's position on a scattergram relative to the typical acceptee's GPA. Login data are only available for students who graduated from the district in 2017. The red line indicates students who logged onto Naviance the most (top 50%) since 2014. The blue line is based on students who logged on the least (bottom 50%). Logins count any time the student's account is used. Weighted GPAs from 11th grade are used to determine the distance from the mean weighted GPA line depicted on the scattergram when the student is in 12th grade. The data are binned in intervals of 0.1 from the threshold at zero. The y-axis represents the fraction of students in each bin who applied to the college. The fitted lines come from a local linear regression discontinuity model with a bandwidth of 0.5. Observations are student-scattergram combinations (where scattergrams are based on colleges which received at least 10 prior applications).

Figure A.13: Impact of Scattergram Visibility on Attendance by Proximity to Typical Acceptee



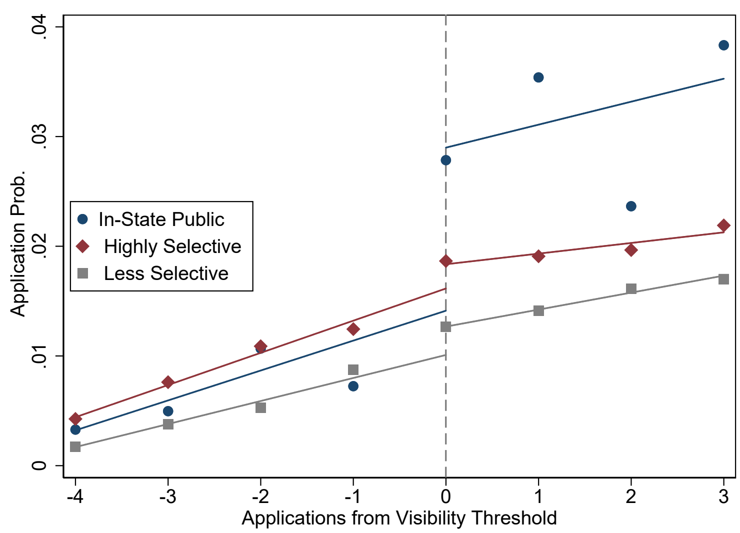
Notes: The figures above show how the probability of a student attending a college changes when a college crosses a scattergram visibility threshold, and how this varies based on the proximity of the student to the typical acceptee lines. Panel (A) is based on the weighted GPA lines and near is defined as within .1 GPA points. Panel (B) is based on the SAT 2400 scale and near is defined as within 50 SAT points. I computed hypothetical typical acceptee lines for colleges which did not meet the cutoff for a scattergram based on the prior applications and used these to compute near, far, above and below, for student-college combinations to the left of the RD threshold. Students to the left of the RD threshold would not have seen these lines. Observations are student-college-threshold combinations. I used distances to both thresholds (five and ten) where relevant. The X-axis shows how far a college was, in terms of applications, from each of these minimum applicant thresholds (in 2015 and 2016). The dots on the y-axis represents the fraction of students who attended a college with previous applications x distance from the threshold.

Figure A.14: Impact of Scattergram Visibility on Acceptance (Conditional on Applying)



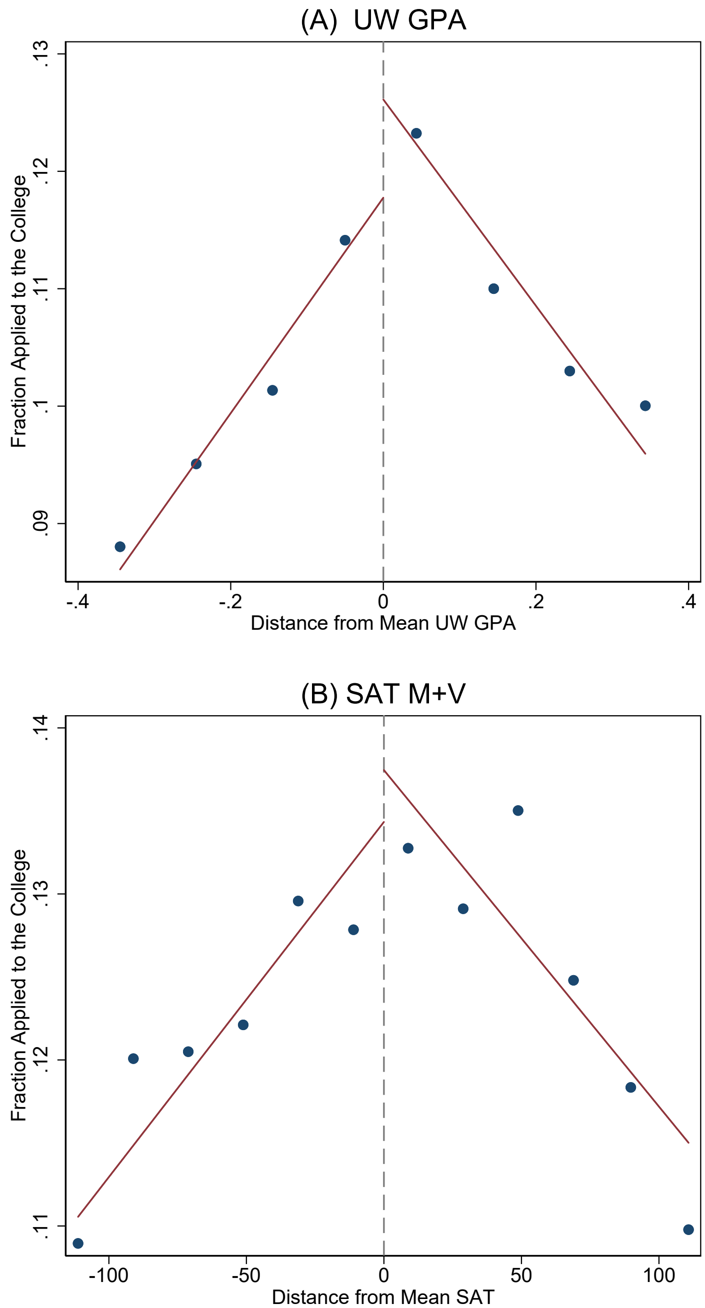
Notes: This figure shows how the probability of a student being accepted to a college, conditional on applying, changes when a college crosses a scattergram visibility threshold. A college's scattergram becomes visible to students after it receives five or ten applications from the student's high school. (I do not know which threshold each high school uses.) The X-axis shows how far a college was, in terms of applications, from each of these minimum applicant thresholds (in 2015 and 2016). Since I use both thresholds, college-high school combinations with 5 to 8 applications in the previous year are included twice in this graph for the same student. Observations are student-college-threshold combinations. The dots on the y-axis represents the fraction of students who were accepted to the college, conditional on applying. The fitted lines are from a local linear regression discontinuity model with a bandwidth of 4 applications.

Figure A.15: Impact of Scattergram Visibility on Application Probability by College Type



Notes: This figure compares the fraction of students who attended a college with the distance of the student's weighted 11th grade GPA from the typical acceptee line she could see and the type of college. Observations are student-college combinations, and the college in this pair must have received at least ten previous applications from the student's high school to be included in this graph. The data are binned in intervals of 0.1 from the threshold at zero. The fitted lines come from a local linear regression discontinuity model with a bandwidth of 0.5. Colleges are broken into highly selective and less selective categories based on Barron's selectivity ratings. The in-state public colleges are excluded from the selectivity groups so that each student-college combination appears at most once in this figure.

Figure A.16: Application Probabilities by Distance from Other Lines



Notes: The figures above show how attendance rates varied based on a student's position on a scattergram relative to the typical acceptee's GPA (in panel A) and SAT (in panel B). For panel (A), unweighted GPAs from 11th grade are used to determine the distance from the mean unweighted GPA line depicted on the scattergram when the student is in 12th grade. The data are binned in intervals of 0.1 from the threshold at zero. For panel (B), students' maximum SAT scores on the old 1600 scale are used. The data are binned in intervals of 20 from the threshold at zero. The fitted lines come from a local linear regression discontinuity model with a bandwidth of 0.5 for panel (A) and a bandwidth of 100 for panel (B). The y-axis represents the fraction of students in each bin who attended the college (in 2015 or 2016). A bin includes multiple scattergrams (and colleges) and it may include the same students multiple times (but for different scattergrams). Observations are student-scattergram combinations (where scattergrams are based on colleges which received at least 10 prior applications).

Table A.1: Additional Summary Statistics

	All	Free/Reduced Lunch: Never	Free/Reduced Lunch: Yes	White or Asian	Black or Hispanic	Weighted By Scattergram Obs.: GPA BW	Weighted By Scattergram Obs.: SAT BW
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(A) Demographics							
White	0.49	0.58	0.16	0.75	0.00	0.57	0.57
Asian	0.17	0.17	0.16	0.25	0.00	0.22	0.22
Black	0.20	0.14	0.45	0.00	0.72	0.11	0.11
Hispanic	0.08	0.05	0.18	0.00	0.28	0.05	0.05
Free/Reduced Lunch	0.21	0.00	1.00	0.10	0.47	0.10	0.09
(B) Academics							
GPA (11th gr. weighted)	3.41	3.58	2.79	3.64	2.88	3.97	3.89
SAT(M+V+W)	1689	1740	1444	1765	1477	1821	1836
Attend 4-yr Coll	0.60	0.67	0.32	0.68	0.42	0.81	0.82
Attend 2-yr Coll	0.24	0.20	0.38	0.21	0.32	0.11	0.11
Persist in Coll	0.78	0.83	0.60	0.84	0.65	0.90	0.91
(C) Applications							
Number of Apps	5.15	5.50	3.89	5.41	4.60	6.51	6.70
Num Reach Apps	1.53	1.40	2.10	1.23	2.34	1.40	1.26
Num Match Apps	2.31	2.53	1.38	2.55	1.73	3.01	3.25
Num Safety Apps	1.29	1.48	0.50	1.55	0.63	1.84	1.91
Highly Selective	3.98	4.30	2.63	4.33	3.07	5.23	5.25
Acceptances	2.51	2.78	1.53	2.77	1.92	3.36	3.43
(D) Attendance							
Reach	0.19	0.17	0.31	0.15	0.33	0.26	0.09
Match	0.54	0.54	0.54	0.55	0.51	0.43	0.63
Safety	0.27	0.29	0.16	0.30	0.16	0.31	0.29
Highly Selective	0.56	0.60	0.32	0.62	0.37	0.71	0.69
(E) Scattergrams							
Total	47	50	38	51	39	59	59
In GPA Bandwidth	18	20	8	21	10	32	30
In SAT Bandwidth	11	13	5	13	6	19	22
Relevant	21	23	10	25	12	36	35
N	7,647	6,004	1,643	5,005	2,156	134,023	85,451

Notes: Column 1 contains the full sample of students. They all appear in the scattergram introduction regressions. The number of times they appear depends on the number of colleges which received an application from their high school between 2014 and 2017 and how many of these colleges fell within the bandwidth. Column (2) contains all students who never received free or reduced-price lunch while enrolled in the district, while column (3) contains students who received it. Students who indicate two or more races or report a race that is not white, Black, Asian, or Hispanic are excluded from columns (4) and (5). In columns (6) and (7) there is one observation for each student-scattergram combination for which the student the student is near the GPA line (column 6) or SAT line (column 7). I define near to the GPA line as students' whose weighted GPAs are within .5 GPA points of the typical acceptee's weighted GPA. I define near to the SAT line as students' whose SAT scores (on the 2400 scale) are within 150 points of the typical acceptee's SAT score. Free/reduced lunch is an indicator for students who ever received free or reduced-price lunch while enrolled in the district. Students who indicate two or more races are excluded from the race categories in Panel (A). GPA refers to 11th grade weighted GPA and SAT refers to the maximum SAT on the old 2400 scale. New SAT scores have been converted to the old 2400 scale. Scattergrams refers to the minimum number of scattergrams to which a student had access based on her graduation year and high school. It is the number of colleges with at least 10 prior applicants. If a college was using the minimum of five applicants, more scattergrams would have been visible. Attend 4-yr college is an indicator for whether the student attended a four-year college within six months of graduating high school. Attend 2-yr is similarly defined but for two-year colleges. Reach schools are colleges at which the student's maximum SAT score is below the 25th percentile of accepted students' SATs, as reported to IPEDS in 2015 by the college. Match schools are colleges at which the student's maximum SAT score is within the interquartile range of accepted students' SATs. Safety schools are colleges at which the student's maximum SAT score is above the 75th percentile of accepted students' SATs. Selectivity ratings are based on Barron's 2009 selectivity index. Where this is missing, selectivity rankings from IPEDS in 2002 are used.

Table A.2: Impact with One Observation per Student-College and Randomly Selected Threshold

		Near GPA		Near SAT		Near Both	Near
	All	.5	.1	150	50	.5 & 150	Neither
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(A) Applied							
Visible	0.0031***	0.0045***	0.0061**	0.0049***	0.0061***	0.0069***	0.0025***
	(0.0004)	(0.0010)	(0.0024)	(0.0013)	(0.0022)	(0.0016)	(0.0004)
CCM	0.0136	0.0228	0.0271	0.0246	0.0264	0.0272	0.0083
(B) Attended							
Visible	0.0002	0.0005	-0.0005	0.0002	0.0005	0.0003	0.0001
	(0.0001)	(0.0003)	(0.0007)	(0.0004)	(0.0007)	(0.0005)	(0.0001)
CCM	0.0011	0.0023	0.0030	0.0027	0.0030	0.0033	0.0004
N	2,394,811	608,959	120,755	394,676	140,159	247,144	1,638,275

Notes: Heteroskedasticity robust standard errors clustered by student are in parentheses. (*p<.10 **p<.05 *** p<.01). All regressions include fixed effects for each student and college by year. Observations are all student-college combinations for which the college was within four applications of the threshold at five or ten and the college received at least one application from the student's high school between 2014 and 2016. When a student-college combination is in the bandwidth for two thresholds, I randomly select one threshold observation to keep. This happens for colleges with five to eight prior applications since they fall in the bandwidth for both thresholds. W GPA refers to weighted GPAs, which are on a five point scale, and these SAT scores are on the (old) 2400 scale. CCM refers to the mean application or attendance probability predicted at a college at the threshold if its scattergram had not been made visible.

Table A.3: Bandwidth and Fixed Effects Comparisons for Scattergram Impacts

	Main	SEs	Fixed Effects			Bandwidths			Triangular
	FE: Stu, CollxYr BW: 4	Kolesar & Rothe	None	Coll HS Yr	Coll Student	0-20	1-14	+/- 3	Kernel
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(A) Applied									
Visible	0.0027***	0.0027*	0.0029***	0.0028***	0.0028***	0.0019***	0.0018***	0.0027***	0.0031 ***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0003)	(0.0004)	(0.0005)	(0.0006)
Visible & Near Lines	0.0056***	0.0056*	0.0080***	0.0068***	0.0058***	0.0022	0.0040**	0.0071***	0.0089***
	(0.0017)	(0.0014)	(0.0015)	(0.0016)	(0.0016)	(0.0014)	(0.0016)	(0.0020)	(0.0019)
(B) Attended									
Visible	0.0001	0.0002	0.0001	0.0002	0.0002	0.0002**	0.0001	0.0001	0.0002
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Visible & Near Lines	0.0001	0.0010**	0.0010**	0.0005	0.0005	0.0003	0.0004	0.0002	0.0011*
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0006)	(0.0006)
N	2,565,375	2,565,375	2,565,375	2,565,375	2,565,375	4,513,998	2,660,037	1,236,352	4,521,645
N Near Lines	272,995	272,995	273,097	273,096	272,998	294,843	287,923	166,597	294,945

Notes: Heteroskedasticity robust standard errors clustered by student are in parentheses. (*p<.10 **p<.05 *** p<.01). Observations are student-college-threshold combinations for all colleges which received at least one application from the student's high school between 2014 and 2016. Near lines is defined as within .5 GPA points of the weighted GPA line and 150 SAT points of the SAT M+V+W line. All regressions in columns (1)-(5) are based on a bandwidth 4 applications. All regressions in columns (1)-(2) and (6)-(8) include student and college by year fixed effects. For columns (6) - (8), a college is in the bandwidth (x) if the number of applications it received in the prior years is in the noted range. Column (9) contains the result from a triangular kernel specification with a bandwidth of 4.

Table A.4: Impact of Scattergrams by Proximity to Other Typical Acceptee Lines

		Near UW GPA		Near SAT M+V		Near Both	Near
	All	.5	.1	150	50	.5 & 150	Neither
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(A) Applied							
Visible	0.0027***	0.0035***	0.0055***	0.0025**	0.0024	0.0030**	0.0024***
	(0.0004)	(0.0008)	(0.0020)	(0.0011)	(0.0018)	(0.0013)	(0.0005)
CCM	0.0137	0.0193	0.0249	0.0240	0.0268	0.0261	0.0080
(B) Attended							
Visible	0.0001	0.0002	0.0004	0.0003	-0.0001	0.0005	0.0001
	(0.0001)	(0.0003)	(0.0006)	(0.0003)	(0.0006)	(0.0004)	(0.0001)
CCM	0.0011	0.0020	0.0024	0.0024	0.0029	0.0027	0.0003
N	2,565,375	908,296	191,989	623,977	227,209	436,252	1,469,312

Notes: Heteroskedasticity robust standard errors clustered by student are in parentheses. (*p<.10 **p<.05 *** p<.01). All regressions include fixed effects for each student and college by year. Observations are all student-college-threshold combinations for which the college was within four applications of the threshold at five or ten and the college received at least one application from the student's high school between 2014 and 2016. Student-college combinations are included twice for colleges with five to eight prior applications since they fall in the bandwidth for both thresholds. UW GPA refers to unweighted GPAs, which are on a four point scale, and these SAT scores are on the (old) 1600 scale. CCM refers to the mean application or attendance probability predicted at a college at the threshold if its scattergram had not been made visible.

Table A.5: Scattergram Impacts on Applications and Attendance by Distance to GPA and SAT

	W GPA				SAT M+V+W		
	All (1)	BW 1 (2)	BW .5 (3)	BW .1 (4)	BW 300 (5)	BW 150 (6)	BW 50 (7)
(A) Apply							
Visible	0.0027***	0.0034***	0.0040***	0.0060**	0.0032***	0.0038***	0.0048**
	(0.0004)	(0.0007)	(0.0010)	(0.0025)	(0.0009)	(0.0013)	(0.0023)
Visible & Above	0.0021**	0.0024**	0.0019	0.0070**	0.0032**	0.0035*	0.0038
	(0.0009)	(0.0011)	(0.0015)	(0.0035)	(0.0013)	(0.0018)	(0.0032)
Visible & Below	0.0022***	0.0035***	0.0053***	0.0052	0.0026**	0.0037**	0.0071**
	(0.0005)	(0.0009)	(0.0014)	(0.0037)	(0.0013)	(0.0018)	(0.0033)
CCM	0.0158	0.0183	0.0237	0.0268	0.0213	0.0242	0.0262
(B) Attend							
Visible	0.0001	0.0003	0.0004	-0.0005	0.0003	0.0000	0.0003
	(0.0001)	(0.0002)	(0.0003)	(0.0007)	(0.0003)	(0.0004)	(0.0007)
Visible & Above	0.0000	0.0000	-0.0003	-0.0007	-0.0002	-0.0007	0.0010
	(0.0003)	(0.0003)	(0.0005)	(0.0010)	(0.0004)	(0.0006)	(0.0009)
Visible & Below	0.0001	0.0005*	0.0011**	-0.0007	0.0007*	0.0007	-0.0004
	(0.0001)	(0.0003)	(0.0004)	(0.0012)	(0.0004)	(0.0006)	(0.0012)
CCM	0.0019	0.0022	0.0031	0.0030	0.0025	0.0033	0.0023
N	2,565,375	1,167,906	666,731	132,319	800,508	432,073	153,384
N Above	671,907	525,023	320,169	68,653	380,179	214,080	80,384
N Below	1,186,729	642,597	346,223	62,910	420,202	217,793	72,589

Notes: Heteroskedasticity robust standard errors clustered by student are in parentheses. (*p<.10 **p<.05 *** p<.01). All regressions include fixed effects for each student and college by year. Observations are all student-college-threshold combinations for which the college was within four applications of the threshold at five or ten and the college received at least one application from the student's high school between 2014 and 2016. Student-college combinations are included twice for colleges with five to eight prior applications since they fall in the bandwidth for both thresholds. CCM refers to the mean application or attendance probability predicted at a college at the threshold if its scattergram had not been made visible. Near is defined as within .5 GPA points or 150 SAT points (based on weighted GPAs and SAT scores on the (old) 2400 scale.

Table A.6: Average Treatment Effects of Scattergram Visibility on Attendance

		Free/Reduced Lunch		White or	Black or	In-St. Public	Other Colleges	
	All	Never	Ever	Asian	Hispanic	Colleges	High Sel.	Less Sel.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) All Students								
Visible	0.0007***	0.0007***	0.0002*	0.0005***	0.0006***	0.0042***	0.0004***	0.0008***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0009)	(0.0001)	(0.0001)
CCM	0.0007	0.0008	0.0004	0.0007	0.0006	0.0016	0.0007	0.0007
N	8,914,720	7,018,780	1,895,940	5,844,978	2,503,108	300,304	2,939,362	5,062,656
(B) Near Lines								
Visible	0.0011***	0.0008***	0.0026**	0.0006**	0.0027***	-0.0004	0.0009**	0.0012***
	(0.0003)	(0.0003)	(0.0010)	(0.0003)	(0.0008)	(0.0067)	(0.0004)	(0.0004)
CCM	0.0018	0.0018	0.0011	0.0017	0.0020	0.0039	0.0019	0.0016
N	583,508	520,768	62,740	451,196	98,194	27,742	233,076	303,676

Notes: Heteroskedasticity robust standard errors clustered by student are in parentheses. (*p<.10 **p<.05 *** p<.01). All regressions include fixed effects for each student and college by year. Observations are all student-college-threshold combinations. Panel (A) includes LATE estimates based on the RDD for colleges within four applications of the threshold at five or ten. Panel (B) includes all colleges and relies on college by year and student fixed effects for identification. CCM refers to the mean attendance probability predicted at a college at the threshold if its scattergram had not been made visible. Near is defined as within .5 GPA points or 150 SAT points. This is based on weighted GPA points and SAT points on the old 2400 scale. Column (2) is based on students who never received free or reduced-price lunch from the district. Column (3) contains all students who received it at least once while enrolled in the district. Students who indicate two or more races or report a race that is not white, Black, Asian, or Hispanic are excluded from columns (4) and (5). The in-state public colleges are excluded from the highly and less selective college categories in columns (7) and (8). Selectivity ratings are based on Barron's 2009 selectivity index. Where this is missing, selectivity rankings from IPEDS in 2002 are used.

Table A.7: Results using Alternative Definitions of Typical Acceptee Lines

	GPA		SAT			Both GPA (Wtd) & SAT(M+V+W)		
	Weighted	Unweighted	M+V+W (2400)	M+V (1600)	Any Line	Below Both	Below at Least One	Below Just One
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) Applied								
Below Line	-0.0107***	-0.0103***	0.0040	-0.0051	-0.0049	0.0022	0.0029	-0.0092***
	(0.0035)	(0.0034)	(0.0038)	(0.0037)	(0.0041)	(0.0042)	(0.0027)	(0.0027)
Dist Above	-0.1034***	-0.1641***	-0.0002***	-0.0004***				
	(0.0106)	(0.0157)	(0.0000)	(0.0000)				
Dist Below	-0.0846***	-0.0349**	-0.0001***	-0.0000				
	(0.0098)	(0.0156)	(0.0000)	(0.0000)				
N	123,429	131,271	101,188	98,970	71,342	71,342	71,342	192,382
(B) Attended								
Below Line	0.0003	-0.0005	0.0016	-0.0029	-0.0015	0.0016	0.0012	0.0007
	(0.0016)	(0.0014)	(0.0018)	(0.0019)	(0.0018)	(0.0018)	(0.0011)	(0.0011)
Dist Above	-0.0091*	-0.0140**	-0.0001***	-0.0001***				
	(0.0048)	(0.0056)	(0.0000)	(0.0000)				
Dist Below	-0.0246***	-0.0129**	-0.0000**	0.0000				
	(0.0042)	(0.0052)	(0.0000)	(0.0000)				
N	123,429	131,271	101,188	98,970	71,342	71,342	71,342	192,382

Notes: Heteroskedasticity robust standard errors clustered by student are in parentheses. (*p<.10 **p<.05 *** p<.01). College by year and high school fixed effects are included. The optimal bandwidths are calculated as described in Calonico, Cattaneo and Titiunik (2014). New SAT scores have been converted to the old scale. All columns include controls for 11th grade GPA, maximum SAT score, gender, special education and dummy variables for race and ever receiving free or reduced-price lunch. Column (5) compares students who are below all the lines (weighted, unweighted, SAT M+V and SAT M+V+W) to students who are above at least one line. Column (6) compares students below both the weighted GPA and SAT line to students who are above at least one line. Column (7) compares students who are below the weighted GPA, SAT M+V+W line or both, to students who are above both lines. Column (8) compares students who are below the weighted GPA or SAT M+V+W line (but not both), to students who are above both lines. N refers to the number of student-scattergram combinations on which the regression is based.

Table A.8: Results for Alternative Specifications Around GPA Line

	Main	GPA Bandwidth		Donut RD		Alternate Specifications		Other Fixed Effects		
		0.4	0.6	+/- .05	+/- .1	Quadr Dist	Triangular Kernel	None No Controls	Student	Scattergram
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(A) Applied										
Below GPA	-0.011***	-0.008**	-0.010***	-0.012***	-0.016***	-0.005	-0.012***	-0.013***	-0.006*	-0.012***
	(0.004)	(0.004)	(0.003)	(0.004)	(0.005)	(0.005)	(0.004))	(0.004)	(0.004)	(0.003)
Dist Above GPA	-0.103***	-0.094***	-0.109***	-0.102***	-0.108***	-0.098***		-0.096***	-0.119***	-0.086***
	(0.011)	(0.014)	(0.009)	(0.012)	(0.014)	(0.035)		(0.010)	(0.010)	(0.009)
Dist Below GPA	-0.085***	-0.088***	-0.093***	-0.081***	-0.079***	-0.143***		-0.080***	-0.089***	-0.086***
	(0.010)	(0.013)	(0.008)	(0.011)	(0.012)	(0.035)		(0.009)	(0.009)	(0.009)
(B) Attended										
Below GPA	0.000	-0.008**	-0.010***	0.001	-0.001	0.000	0.000	0.000	0.001	0.000
	(0.002)	(0.004)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Dist Above GPA	-0.009*	-0.094***	-0.109***	-0.007	-0.008	-0.026		-0.001	0.002	-0.015***
	(0.005)	(0.014)	(0.009)	(0.006)	(0.007)	(0.017)		(0.005)	(0.005)	(0.004)
Dist Below GPA	-0.025***	-0.088***	-0.093***	-0.025***	-0.023***	-0.040**		-0.019***	-0.043***	-0.016***
	(0.004)	(0.013)	(0.008)	(0.005)	(0.005)	(0.016)		(0.004)	(0.004)	(0.004)
N	123,429	100,729	144,229	111,233	97,703	123,429	345,548	131,704	131,437	131,704

Notes: Heteroskedasticity robust standard errors clustered by student are in parentheses. (*p<.10 **p<.05 *** p<.01). The main regression, is in column (1). It is based on a bandwidth of .5 GPA points, college by year and high school fixed effects, as well as controls. Controls include 11th grade GPA, maximum SAT score, gender, special education, and dummy variables for race and ever receiving free or reduced-price lunch. All estimates are for weighted GPAs in 11th grade. N refers to the number of student-scattergram combinations on which the regression is based. The specifications in columns (1-7) include college by year and high school fixed effects and controls for student characteristics. The donut RD columns (4 and 5) exclude student observations in which the student is within .05 or .1 GPA points of the GPA line. A quadratic term is added for GPA distance in column (6). Control variables are excluded from columns (7)-(10).

Table A.9: Results for Alternative Specifications Around SAT Line

	Main	SAT Bandwidth		Donut RD		Alternate Specifications		Other Fixed Effects		
		0.4	0.6	+/- 10	+/- 20	Quadr Dist	Triangular Kernel	None No Controls	Student	Scattergram
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(A) Applied										
Below SAT	0.0040	0.0050	0.0033	0.0043	0.0002	0.0071	-0.006	0.0042	0.0050	0.0045
	(0.0038)	(0.0052)	(0.0036)	(0.0047)	(0.0055)	(0.0058)	(0.005)	(0.0041)	(0.0041)	(0.0038)
Dist Above SAT	-0.0002***	-0.0001**	-0.0002***	-0.0002***	-0.0002***	-0.0001		-0.0001***	-0.0002***	-0.0001***
	(0.0000)	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0001)		(0.0000)	(0.0000)	(0.0000)
Dist Below SAT	-0.0001***	-0.0001	-0.0001***	-0.0001***	-0.0001**	-0.0001		-0.0002***	-0.0002***	-0.0002***
	(0.0000)	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0001)		(0.0000)	(0.0000)	(0.0000)
(B) Attended										
Below SAT	0.0016	-0.0000	0.0025	0.0018	0.0007	0.0005	-0.002	0.0019	0.0027	0.0019
	(0.0018)	(0.0024)	(0.0018)	(0.0023)	(0.0026)	(0.0027)	(0.002)	(0.0020)	(0.0020)	(0.0018)
Dist Above SAT	-0.0001***	-0.0001*	-0.0000***	-0.0000**	-0.0000**	-0.0000		-0.0000***	-0.0000***	-0.0001***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)		(0.0000)	(0.0000)	(0.0000)
Dist Below SAT	-0.0000**	0.0000	-0.0000***	-0.0000	-0.0000	-0.0000		-0.0000**	-0.0001***	-0.0000*
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0001)		(0.0000)	(0.0000)	(0.0000)
N	101,188	57,700	111,554	80,245	74,474	101,188	291,959	101,188	101,004	101,188

Notes: Heteroskedasticity robust standard errors clustered by student are in parentheses. (*p<.10 **p<.05 *** p<.01). The main regression, is in column (1). It is based on a bandwidth of 150, college by year and high school fixed effects, as well as controls. Controls include 11th grade GPA, maximum SAT score, gender, special education, and dummy variables for race and ever receiving free or reduced-price lunch. All estimates are for SAT scores on the 2400 scale. New scores have been converted to the old scale. N refers to the number of student-scattergram combinations on which the regression is based. The specifications in columns (1-7) include college by year and high school fixed effects and controls for student characteristics. The donut RD columns (4 and 5) exclude student observations in which the student is within 10 or 20 points of the SAT line. A quadratic term is added for SAT distance in column (6). Control variables are excluded from columns (7)-(10).

Table A.10: Impact of Mean Lines on Applications by Year and Logins

	2015-2016	2015-2017	2015	2016	2017		
					All	Many Logins	Few Logins
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(A) GPA Line on Apps							
Below GPA	-0.011***	-0.013***	-0.014**	-0.008*	-0.016***	-0.017***	-0.008*
	(0.004)	(0.003)	(0.006)	(0.004)	(0.004)	(0.005)	(0.005)
Dist Above GPA	-0.103***	-0.112***	-0.123***	-0.094***	-0.125***	-0.148***	-0.052***
	(0.011)	(0.008)	(0.018)	(0.012)	(0.011)	(0.013)	(0.014)
Dist Below GPA	-0.085***	-0.074***	-0.093***	-0.082***	-0.051***	-0.065***	-0.022**
	(0.010)	(0.008)	(0.017)	(0.011)	(0.009)	(0.013)	(0.011)
N	123,429	216,655	45,624	84,059	11,2872	77,792	39,406
(B) SAT Line on Apps							
Below SAT	0.0040	-0.0021	-0.0025	0.0092*	-0.0106***	-0.0102**	-0.0099*
	(0.0038)	(0.0025)	(0.0068)	(0.0051)	(0.0036)	(0.0046)	(0.0056)
Dist Above SAT	-0.0002***	-0.0003***	-0.0002***	-0.0001***	-0.0004***	-0.0005***	-0.0003***
	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0001)
Dist Below SAT	-0.0001***	-0.0001***	-0.0000	-0.0002***	-0.0000	-0.0001	0.0000
	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0000)	(0.0001)	(0.0001)
N	101,188	222,325	36,949	52,259	90,598	65,364	26,950

Notes: Heteroskedasticity robust standard errors clustered by student are in parentheses. (*p<.10 **p<.05 *** p<.01). College by year and high school fixed effects are included, as well as controls for 11th grade GPA, maximum SAT score, gender, special education, and dummy variables for race and ever receiving free or reduced-price lunch. The optimal bandwidths, as described in Calonico, Cattaneo and Titiunik (2014), are calculated for each regression. All estimates are for weighted GPAs and SAT scores on the old 2400 scale. New scores have been converted to the old scale. The outcome is applying to the college associated with the scattergram treating the student. N refers to the number of student-scattergram combinations on which the regression is based. Column (1) shows the main results which are based on students who graduated in 2015 and 2016. Login records are only available for students who graduated in 2017. Column (6) is based on students who were in the top 50% in terms of Naviance logins. Column (7) is based on students in the bottom 50%. Students who logged onto Naviance more were in the bandwidth for more scattergrams, which is why the N in column (6) is much larger than the N in column (7).

Table A.11: Cumulative Impact of Scattergrams for Minority and Low-Income Students

	Applications				Acceptances	Attend College				
	Num.	Reach	Match	Safety		Reach	Match	Safety	Four-yr	Persist
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Total SGs	-0.007	-0.026**	0.039***	-0.015**	0.010	-0.004**	0.008***	-0.004***	-0.000	0.005
	(0.019)	(0.013)	(0.010)	(0.006)	(0.010)	(0.002)	(0.002)	(0.001)	(0.002)	(0.004)
Reach SGs	0.002	0.143***	-0.003	-0.130***	0.023	0.023***	-0.009	-0.020***	-0.006	-0.000
	(0.080)	(0.054)	(0.043)	(0.026)	(0.044)	(0.008)	(0.010)	(0.006)	(0.009)	(0.015)
Match SGs	-0.015	-0.036**	0.056***	-0.029***	0.015	-0.004	0.010***	-0.006***	0.001	0.010*
	(0.027)	(0.018)	(0.014)	(0.009)	(0.015)	(0.003)	(0.003)	(0.002)	(0.003)	(0.006)
Safety SGs	-0.019	-0.113***	0.056**	0.042***	-0.001	-0.022***	0.018***	-0.001	-0.004	-0.005
	(0.048)	(0.032)	(0.026)	(0.016)	(0.026)	(0.004)	(0.006)	(0.004)	(0.005)	(0.010)
In-St Public SGs	0.053	-0.049	0.163***	-0.045	0.030	-0.014	0.043***	-0.010	0.023**	0.005
	(0.103)	(0.070)	(0.055)	(0.034)	(0.056)	(0.010)	(0.012)	(0.008)	(0.012)	(0.017)
N	1,409	1,409	1,409	1,409	1,409	1,409	1,409	1,409	1,409	623

Notes: Heteroskedasticity robust standard errors are in parentheses. (*p<.10 **p<.05 *** p<.01). High school and year fixed effects are included. I control for academic achievement using fixed effects for 50 point intervals of maximum SAT scores, and .1 point intervals of students' weighted 11th grade GPAs. Controls include demographic indicators for race (white, asian, black or hispanic), free-or-reduced price lunch, special education, and gender. There is one observation per student. Persistence refers to persistence into a second year of college. These data are only available for students who graduated high school in 2015. Reach schools are colleges at which the student's maximum SAT score is below the 25th percentile of accepted students' SATs, as reported in IPEDS in 2015. Match schools are colleges at which the student's maximum SAT score is within the interquartile range of accepted students' SATs. Safety schools are colleges at which the student's maximum SAT score is above the 75th percentile of accepted students' SATs. Acceptances are self-reported but I corrected the self reports if a student attended a college where an acceptance decision was not reported. I assume a student must have been accepted to a college if she attends the college.

Table A.12: Balance Table for Colleges with and without Scattergrams

	Private	Out-of-State Public	In-State Public	Selectivity Tier	Net Price	Enrollment
	(1)	(2)	(3)	(4)	(5)	(6)
Visible	0.0309	-0.0068	-0.0529	3.4982	-245.2943	-2.2939**
	(0.0357)	(0.0100)	(0.0369)	(7.5569)	(620.4679)	(1.0809)
N	7,956	7,956	7,485	7,485	7,638	7,485

Standard errors clustered by high school and year (combinations) are in parentheses. (*p<.10 **p<.05 *** p<.01). All colleges within four applications of the thresholds at five or ten are included. Selectivity Tier refers to Barron's rankings in 2009. Net Price and Enrollment numbers come from Ipeds in 2015.

Table A.13: Balance Table for Students Above and Below Typical Acceptee Lines

	Below GPA (1)	Below SAT (2)
White/Asian	-0.003	0.001
	(0.004)	(0.004)
Black/Hispanic	0.005	-0.001
	(0.004)	(0.004)
Female	0.004	0.002
	(0.006)	(0.006)
Free or Reduced-Price Lunch	-0.002	0.005
	(0.003)	(0.003)
Special Ed	0.004**	0.000
	(0.002)	(0.001)
N	131,704	101,188

Notes: Heteroskedasticity robust standard errors clustered by student are in parentheses. (*p<.10 **p<.05 *** p<.01). Estimates are from a regression of the indicator for being below the typical acceptee line (for a particular scattergram) on the demographic variable and the distance of one's GPA or SAT from the line. High school and college by year fixed effects are included. The bandwidths are .5 GPA points and 150 SAT points. All estimates are for weighted GPAs and SAT scores on the 2400 scale. New SAT scores have been converted to the old scale. N refers to the number of student-scattergram combinations on which the regression is based.

Table A.14: Scattergram Impacts by Year

	2015 (1)	2016 (2)	2017 (3)	2015-2016 (4)	2015-2017 (5)
(A) Applied					
Visible	0.0037***	0.0026***	0.0009***	0.0031***	0.0022***
	(0.0008)	(0.0005)	(0.0003)	(0.0004)	(0.0003)
Visible & Near Lines	0.0068**	0.0067***	0.0018*	0.0067***	0.0041***
	(0.0031)	(0.0019)	(0.0010)	(0.0017)	(0.0010)
(B) Attended					
Visible	0.0002	0.0003**	0.0003**		
	(0.0002)	(0.0001)	(0.0001)		
Visible & Near Lines	0.0005	0.0006	0.0006		
	(0.0010)	(0.0006)	(0.0005)		
N	1,021,258	1,427,266	1,603,743	2,448,524	4,052,267
N Near Lines	105,559	149,987	246,606	255,546	502,152

Notes: Heteroskedasticity robust standard errors clustered by student are in parentheses. (*p<.10 **p<.05 *** p<.01). All regressions include fixed effects for each student and college by year. Observations are all student-college-threshold combinations for which the college received at least one application from the student's high school between 2014 and 2017. Near is defined as within .5 GPA points or 150 SAT points. This is based on weighted GPA points and SAT points on the old 2400 scale. All regressions are based on a bandwidth of 4 applications.

Table A.15: Impact of Mean SAT and GPA on Attendance

	All	White or Asian	Black or Hispanic	In-St. Public Colleges	Other Colleges High Sel.	Other Colleges Less Sel.
	(1)	(2)	(3)	(4)	(5)	(6)
(A) GPA						
Below GPA	0.000	-0.001	0.001	0.007	-0.000	0.002
	(0.002)	(0.002)	(0.004)	(0.009)	(0.002)	(0.002)
Dist Above GPA	-0.011*	-0.016**	-0.019**	-0.010	-0.001	-0.001
	(0.006)	(0.006)	(0.009)	(0.026)	(0.009)	(0.003)
Dist Below GPA	-0.025***	-0.022***	-0.024***	-0.111***	-0.009	-0.006
	(0.005)	(0.006)	(0.007)	(0.029)	(0.007)	(0.004)
CCM	0.026	0.026	0.028	0.113	0.010	0.005
N	110,013	85,875	26,522	19,081	43,719	39,620
(B) SAT						
Below SAT	0.0013	0.0031*	0.0083*	0.0088	-0.0013	0.0037*
	(0.0019)	(0.0018)	(0.0048)	(0.0095)	(0.0016)	(0.0019)
Dist Above SAT	-0.0001***	-0.0001***	-0.0000	-0.0002**	-0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0000)
Dist Below SAT	-0.0000*	-0.0001***	-0.0000	-0.0001	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0000)
CCM	0.025	0.027	0.019	0.111	0.007	0.005
N	97,226	105,567	16,012	15,082	49,691	29,159

Notes: Heteroskedasticity robust standard errors clustered by student are in parentheses. (*p<.10 **p<.05 *** p<.01). College by year and high school fixed effects are included, as well as controls for 11th grade GPA, maximum SAT score, gender, special education, and dummy variables for race and ever receiving free or reduced-price lunch. Optimal bandwidths for each regression are calculated as described in Calonico, Cattaneo and Titiunik (2014). All estimates are for weighted GPAs, which are on a five point scale, in panel (A) and SAT scores on the 2400 scale in panel (B). New scores have been converted to old ones where relevant. The outcome is attending the college with the relevant scattergram. N refers to the number of student-scattergram combinations on which the regression is based. Students who indicate two or more races or report a race that is not white, Black, Asian, or Hispanic are excluded from columns (2) and (3). The in-state public colleges are excluded from the highly and less selective college categories in columns (5) and (6). Selectivity ratings are based on Barron's 2009 selectivity index. Where this is missing, selectivity rankings from IPEDS in 2002 are used. CCM refers to the mean attendance probability for students with GPAs or SATs just above the typical acceptee's.

B Additional Details on Naviance and the District's Implementation

Naviance is a popular technology used by schools and counselors to help with the college and career counseling process. It is owned by Hobsons and has versions available for all students

Table A.16: Bandwidth Comparisons for Subgroup Results

	All	Free/Reduced Lunch		White or	Black or	In-St. Public	Other Colleges	
		Never	Ever	Asian	Hispanic	Colleges	High Sel.	Less Sel.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) BW: +/- 4								
Visible	0.003***	0.003***	0.002*	0.003***	0.003***	0.013***	0.003***	0.002***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)
N	2,394,811	1,895,046	499,765	1,581,564	663,378	58,730	892,047	1,444,034
(B) BW: +/- 3								
Visible	0.003***	0.003***	0.001	0.003***	0.002	0.007**	0.003***	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)
N	1,122,462	890,641	231,821	745,508	307,209	34,742	518,596	569,124
(C) BW: 1-14								
Visible	0.003***	0.003***	0.001*	0.003***	0.002***	0.012***	0.003***	0.002***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)
N	2,448,524	1,938,859	509,665	1,618,331	676,992	59,569	928,389	1,460,566
(D) BW: 0 - 20								
Visible	0.003***	0.003***	0.002**	0.003***	0.002***	0.012***	0.002***	0.003***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)	(0.002)	(0.001)	(0.001)
N	4,343,434	3,413,033	930,401	2,838,708	1,227,753	103,730	1,410,797	2,828,907

Notes: Heteroskedasticity robust standard errors clustered by student are in parentheses. (*p<.10 **p<.05 *** p<.01). All regressions include fixed effects for each student and college by year. Observations are all student-college-threshold combinations for which the college received at least one application from the student's high school between 2014 and 2016. Near is defined as within .5 GPA points or 150 SAT points. This is based on weighted GPA points and SAT points on the old 2400 scale. Panel (A) contains the main results. Panels (B), (C), and (D) contain alternate bandwidths of prior applications. Column (2) is based on students who never received free or reduced-price lunch from the district. Column (3) contains all students who received it at least once while enrolled in the district. Students who indicate two or more races or report a race that is not white, Black, Asian, or Hispanic are excluded from columns (4) and (5). The in-state public colleges are excluded from the highly and less selective college categories. Selectivity ratings are based on Barron's 2009 selectivity index. Where this is missing, selectivity rankings from IPEDS in 2002 are used.

in grades K-12. I focus on the features used by the high school students and center my research on the scattergrams. It is important to note that this is just one of the many features available in Naviance, though it is perhaps the most novel feature of the platform.

More than forty percent of students in the U.S. use Naviance and information on some of the districts using it, its implementation and features are available on Naviance's website `https://www.naviance.com/`. Naviance provides a similar implementation package to all schools and districts that purchase it, though there are some options for customizing implementation based on the district's needs. The district I studied believed the guidance they received from Naviance to be fairly typical. This guidance included a set of materials and trainings to help them get started as well as a liason whom they could contact with questions. Additional materials and tutorial videos can be found on Naviance's website. Naviance also provides ongoing professional development, and summaries of updates to Naviance's platform, as well as tips, are periodically emailed to counselors.

Naviance is sold directly to schools, not students, and is intended for use with a school counselor. Thus, as in most districts, counselors in this district were responsible for implementing Naviance. Counselors received some training from Naviance and then they were responsible for providing guidance to students and parents around how to use the platform. The district counseling office also provided some guidance to the schools and students. For instance, the district suggested that schools set up time during the school day for students to register with Naviance and explore its main features, such as the quizzes, scattergrams, and how to save colleges.

Counselors set up information sessions for parents and students, and logged on with students during school hours. They also provided students some specific suggestions around how to use the platform. Some counselors encouraged students to start by taking the quizzes or looking at previously popular colleges.

The district office mentioned some concerns about the typical acceptee scores and how they were being used. They wanted counselors to help students interpret these, but it was unclear how much this message permeated through the schools. The district did not explicitly encourage or discourage use of the typical acceptee scores in students' application choices. It appears that different counselors provided different suggestions about how to use Naviance. This is consistent with the lack of a district-wide application strategy. (The district was excited about this project because they wanted to understand how counselors and students used Naviance and how it impacted the advice counselors provided students.)

Counselors and students each have their own accounts. Counselors can use their acccounts to send recommendations and nudges to students. They may also look at scattergrams and other student data to help them determine which colleges to recommend to students. Students use their accounts to access Naviance's features, which include exploring and saving colleges. Some districts allow parents to create their own accounts but in many districts, including the one I study, parents do not have their own accounts. They often used their student's account so it is unclear how much the usage rates are driven by students or parents.

Districts have some control over which information students can see on the scattergrams. During the time I study, schools could set the minimum number of applications needed to make a scattergram visible. These minimums can either be set by the district or by individual schools. The threshold a school selects applies to all colleges but it can be changed at any time. The district I studied informed me that they used the minimums of five and ten and I sat down with a district administrator to examine how they could change the settings. This setting no longer exists in the version of Naviance the district currently uses, so I cannot see what rule the schools are currently using. I also lack power to use the data to detect which cutoffs each high school used.

Districts and schools can also choose whether to limit the number of cohorts that populate the scattergram or use all available data. If the data are not limited, the number of available scattergrams and data points will continue to grow over time. During the period I studied, the district did not limit the cohorts for which data were visible.

This district adds data on a graduating class in the June that the cohort graduates so that students have updated information when searching for colleges over the summer. During their 12th grade, the class of 2015 could access data on the class of 2014. The class of 2016 had access to two sets of scattergrams - one during the 2015 school year and another during 2016 school year. Students logged onto Naviance more during 12th grade than 11th grade, and most application decisions are made in 12th grade, so I focus on the 12th grade scattergrams. Students exploring scattergrams during the fall of their 12th grade will see their 11th grade GPA populated on the scattergram and in the college dashboard.

C Details on the Data and Accuracy of Self-Reported Admissions

Application data are based on requests for students' transcripts to be sent to colleges. This measure may inaccurately count too many colleges in a student's application portfolio if a student decides not to complete the application after requesting a transcript. Transcripts cost a few dollars to send, so it is unlikely that this is happening in many cases. It is possible that this approach misses some applications if students send unofficial transcripts. This is probably not a big concern because most colleges request an official transcript and the transcript request in Naviance triggers a request for a counselor recommendation, which is also necessary at many colleges. In 2015-2016, only 10 students did not submit a transcript request for a four-year college they attended (as indicated by the NSC records) within 6 months of graduating high school.

C.1 Self-Reported Admissions Data

Admissions decisions are self-reported in the senior survey. If the self-reports are inaccurate, or if many students do not report their expereinces, the admissions information students see may be biased. Approximately 90% of students respond to the senior survey, however, many students appear to under-report rejections. For this reason, the district treats non-responses as rejections. Fifteen percent of applications are reported to end in rejections and 53% in admissions. Under-reporting of admissions appears to be less of a problem. 10% of students who apply to at least one college report no admissions. This could be because they were admitted nowhere or because they did not respond to the survey. 9% of students who applied to five or more colleges report no acceptances. This is probably driven by non-response since rather than not being accepted anywhere.

At least 3% of students under-report acceptances. When they complete the survey, 3% of students report plans to attend a college but do not report an acceptance at that college. NSC records indicate that 13% of students attend a college where they did not report an acceptance. Some of the latter discrepancy could be driven by students getting off the waitlist over the summer. In addition, 69% of students' self-reported attendance plans match the NSC records. (I do not use the attendance self-reports in my study.) Districts could use NSC records to update the accuracy of the self-reported data they put into Naviance. This may not be difficult if the district purchases the Naviance feature which links NSC records to student records in Naviance.

Over-reporting of acceptances may also be an issue. 32% of students report acceptances everywhere they apply, but this is largely driven by students who only apply to a few colleges. 5.5% of students who apply to five or more colleges report being accepted everywhere they apply. This may be true or some of it could be driven by students quickly or carelessly responding to the survey.

While missing admissions data may bias the accuracy of the admissions information students see, it will not bias the estimates of the treatment I am studying. Admissions is not one of my main outcomes, and when I use it as an outcome, I correct the admissions self-reports with the attendance self-reports and NSC records. I assume that if a student attends a college she must also have been admitted to it.

C.2 Definitions

I define *safety* colleges as those where the student's SAT is above the 75th percentile of all accepted students, as reported to IPEDS in 2015. Reach colleges are defined as those where the student's SAT is below the 25th percentile. Match colleges are defined as those where a student's SAT score is whithin the inter-quartile range. Elite colleges are the public and private institutions defined as elite by Barron's *Profiles of American Colleges*.

In section 5 I briefly mention medium popular colleges. Medium popular colleges are defined as those where where 5 to 20% of the high school students applied in 2014. Application rates are constant for the most popular colleges, and the least popular colleges experience an increase in applications, likely due to increased awareness from the scattergrams. This seems reasonable if we expect students to always apply to the local public colleges but to substitute across out of state or private colleges based on how much they know about them and their admissions probability. This is consistent with scattergrams broadening the set of schools to which students apply.

I also briefly examine a college's competitors in section 5. College X's closest competitor is defined as the college Y which is most popular among students who applied to college X, relative to its average popularity in the sample.

C.3 A Note on Recreating Scattergrams

Schools have several options for the orientation of the scattergrams, so one needs to confirm the decision rules used by the district or school before using the raw data to reconstruct scattergrams outside of Naviance. Some of these decision rules include the minimum number of applications needed to make it visible, the GPA scale and the number of cohorts with visible data. I visually confirmed that my identification of the typical acceptee lines matched what students observed in 2017 for twenty colleges at one high school. I focus on the version of the scattergram students could see in the fall of their 12th grade. At this time, students could see how their 11th grade GPA compared to the average admit's.

D Additional Robustness Checks

I find no evidence of manipulation of the running variables in any of the regression discontinuity designs I use. Figure A.9 shows no spike in the density of observations with exactly five or ten prior applications. In addition, Table A.12 shows that the colleges just above and below these thresholds are similar in terms of their sector, location, selectivity, price and enrollment.

I find no evidence that students manipulate their SATs to be just above the typical acceptee's (Figure A.10), despite evidence that, in other settings, students retake the SAT until their score is equal to or above minimum admissions thresholds (Goodman, Hurwitz, & Smith, 2017; Goodman, Gurantz, & Smith, forthcoming). The density of SAT scores and the distance of SAT scores from the lines are smooth, as per McCrary (2008). GPAs are typically more difficult to manipulate because they are a combination of many grades over a multi-year period and they are calculated to two decimal points. Manipulation does not appear to be a problem for weighted GPAs (Figure A.10). The density of weighted GPAs, and their distance from the typical acceptee's weighted GPA, are also smooth as per McCrary (2008). Since students view the relevant line after their 11th grade GPA is fixed they cannot manipulate it. While some could view a college's scattergram in 11th grade, the location of the line could have been different so this should not lead to heaping around the 12th grade line. Table A.13 shows no significant differences in observable characteristics for students just above and below the typical acceptee's GPA or SAT.