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

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

April 2019

PRELIMINARY

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, four anonymous referees, Philip Oreopoulos, and seminar participants at Harvard, 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 among more than 4,000 colleges, and traditionally disadvantaged students often lack information about the application process, admissions criteria, and the benefits and costs associated with different types of colleges (Avery & Kane, 2004; Hoxby & Avery, 2013; Hastings, Neilson & Zimmerman, 2015). Improving students' application choices is important because these choices have large impacts on college enrollment, degree attainment and future 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 low-cost technology can change where students apply to and attend college by providing them 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; 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 they are constrained by large caseloads and the high touch nature of their support is not scalable (Hurwitz & Howell, 2014). General or online college resources, such as the *Princeton Review* or the College Scorecard, are more scalable solutions, but they 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 shows students how their academic profiles compare to prior schoolmates' who were admitted or rejected from colleges popular within their high school. This information is conveyed in Naviance's scattergrams, which are scatterplots through which a high school student can see the GPA and SAT (or ACT) scores of prior applicants from her high school to a specific college, as well as the admissions decision each of these applicants received. An example can be seen in Figure 1.

[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

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.

Naviance is an online platform that can be purchased by districts to help with college counseling and students' college choices. In addition to the scattergrams, it contains college and career search tools, descriptive information about colleges, and a portal for contacting counselors and requesting college materials. Schools are encouraged to introduce it to students in 9th or 10th grade so they can explore career options and the scores needed for college admission. Students access it more during 11th grade, when taking college entrance exams, and usage peaks during 12th grade, when students choose where to apply to college, submit applications and enroll in 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 experiences of students who graduated in 2014, and they were updated in 2015 to also include data on the class of 2015.[2] 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 see 47 scattergrams. I examine how access to these scattergrams, and the average acceptance criteria they conveyed, influenced where students applied to college and attended.[3]

This paper contains four main findings. The first three are about how access to a college's admissions information, and what it signals about a student's probability of admission, change applications and enrollment at that college. The fourth is about how the set of admissions information a student can access impacts the student's application portfolio and college attendance.

First, I use a regression discontinuity design to causally show that access to a college's admissions information increases applications and attendance at that college, especially for students with a high probability of admission. A college's scattergram is only visible if the high school

[2]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.

[3]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.

has data on at least five students who previously applied to the college. Some schools further restrict this to colleges with at least ten data points. I use these minimum application cutoffs in a regression discontinuity design to identify the impact of access to admissions data, by comparing application and attendance rates at colleges just above and below the visibility cutoffs.

Students are 20 % more likely to apply to colleges with visible admissions information. Gaining access to a college's admissions information has the largest impact on the students who are most similar to previous admits, as well as Black, Hispanic, and free or reduced-price lunch students, who are most likely to lack this type of information. Black and Hispanic students are 55% more likely to enroll in a college if it is just above a visibility cutoff. I also find larger effects for in-state public colleges, possibly because these are the most commonly viewed scattergrams, or because they are inexpensive and nearby. Students are 53% more likely to apply to an in-state public college if they can see its scattergram and more than twice as likely to enroll in it.

Second, I show how students change their applications based on signals about their probability of admission. For each college, the lines indicating the average GPA and SAT scores of previously admitted students vary significantly across the high schools and 2 years I study. These lines are based on self-reported admissions outcomes and often only a few admitted students, so they offer a noisy signal about a student's admissibility. I use this variation, conditional on college by year fixed effects, to show that students prefer to apply to colleges where they are most similar to previous admits. Students with scores below the average admit are more likely to apply to a college the higher their perceived probability of admission, but students above the average admit are less likely to apply the further they are above the admissions criteria, probably because the signals indicate they can be accepted at a more selective college.

Third, I show that students use the average admissions lines as heuristics to simplify their application choices. I use a regression discontinuity to identify the impact of the average GPA and SAT lines on applications and attendance. Students just below the GPA line are 8% less likely to apply to a college than students just above it. I find no discontinuity at the SAT line, possibly because there are many sources of information on SAT admissions criteria. Students seem to interpret being below the mean GPA as a negative signal and their reactions reduce the selectivity of their application portfolios and college attended. Reactions are largest for students who can see the most scattergrams, indicating that the lines may be used as heuristics to simplify their choices.

Finally, I show that Naviance causes students' application portfolios and attendance choices to reflect the set of colleges with visible and relevant information.[4] The number of relevant *reach*, *match*, and *safety* colleges a student can view depends on quasi-random variation in which colleges crossed the visibility threshold and variation across high schools and time in how accurately the average accepted scores reflect true admissions criteria.[5] Students who see more relevant scattergrams for colleges which are a good academic fit are more likely to attend a *match* college, while those who see more *safety* colleges are more likely to attend a *safety* college. The set of colleges to which students are being nudged depends on which colleges were popular among previous cohorts and how accurately the previous admits' scores reflect colleges' true admissions criteria. This approach improves the quality of where some students attend, but deters others from attending highly selective or *match* colleges. This can impact students' college degree attainment, future employment and earnings (Chetty et al. 2017; Dillon & Smith, 2018; Cohodes & Goodman, 2014).

Admissions information has the most notable effect on Black, Hispanic and low-income students. Every additional relevant scattergram they see, for an in-state public college, causes a 2.3 percentage point increase in four-year college enrollment. This is driven by a shift from local community colleges to the state's many small public colleges, which suggests that students may have been unaware of these nearby and inexpensive options with high admissions rates. It also indicates potential for information of this sort to help close socioeconomic gaps in college enrollment, degree attainment and earnings (Goodman, Hurwitz & Smith, 2017; Zimmerman, 2014).

Access to this type of admissions information may influence college choices for a few reasons. First, access to information for a subset of colleges may act as a nudge towards these colleges, by making students aware of them, or by making them seem like less risky choices than colleges without admissions data. Students may also update their applications based on the prior popularity of

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

[5]The average admit's scores determine relevance, and variation in the number of relevant *match* scattergrams a student sees depends on the accuracy of the lines. If the lines were accurate, almost all relevant scattergrams would be match colleges. They may, however, be inaccurate because they are based on self-reported admissions and often just a few students. I define *reach* (*safety*) colleges as those where a student's SAT is below the 25th percentile (above the 75th percentile) of all admitted students'. *Match* colleges are those where the student's SAT is in the inter-quartile range.

a college, especially if they take it as a signal that the college will be a good fit for them. Their reactions may also be driven by the specific information they receive about their own admissibility, with students updating their application portfolios to increase their admissions probabilities.

My findings indicate that students prefer to apply to colleges at which they have some information about their admissibility and where they are likely to be admitted. My causal estimates are consistent with the changes I observe before and after scattergrams were 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, fewer students apply to *reach* colleges and more students attend *safety* schools in the first year with scattergrams than in the previous year. 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 their admissibility because some attend less selective colleges when they have this information.

This paper provides the first evidence on how admissions information influences students' 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. My paper builds 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 in their application choices.

I also show that a popular technology can create large changes in application choices by providing students 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 GPA and SAT 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).[6] Students may also respond strongly to the scattergram data because they are based on their peers, and students respond strongly to peer norms in other settings (Akerlof & Kranton, 2002; Bursztyn & Jensen, 2015). In addition, information may 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 ten dollars per student for acces to this technology.

Despite the rapid rise of education technologies, including many in the college choice space, there is little convincing evidence on how these technologies can help transform students' education experiences (Escueta, Quan, Nickow & Oreopoulos, 2017; Shellenbarger, 2017; Shulman, 2018).[7] Existing research indicates potential for technology to improve students' choices and outcomes, but some technologies reduce student performance or exacerbate socioeconomic gaps (Bergman & Chan, 2017; 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 these busy workers to more efficiently meet the needs of the individuals they support. It also provides evidence on the effectiveness of one of the most widely adopted college choice technologies.

[6]In particular, the average GPA lines deter some students from attending highly selective colleges.

[7]$9.5 billion were invested in education technology ventures in 2017 (Shulman, 2018).

The paper proceeds as follows. Section 2 contains a description of Naviance, the data and setting, and descriptive statistics on changes over time. Section 3 describes the empirical approach and section 4 describes the results. The implications and conclusions are discussed in section 5.

2 Naviance and Setting

2.1 Naviance

Naviance is a software purchased by school districts which can be used to track student progress and prepare students to make postsecondary choices. In particular, districts can use it to track students' goals, course schedules, counselor meetings and progress towards graduation. Students can also take quizzes to identify careers and colleges which match their interests, see career descriptions, including occupation statistics from the Bureau of Labor Statistics, and view college characteristics, such as size, location, majors, and graduation rate. Students can save colleges in which they are interested, and counselors or parents can log in and save colleges to a student's profile. Naviance can also be used to 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 how it plans to use the software. At a minimum, the package includes a tutorial of the basic features, school counselor training, guidance to provide to students, and a liaison whom the district can contact with questions.[8] Counselors are encouraged to introduce and provide guidance on Naviance to students and parents in classes or after school sessions. Students and parents can also watch tutorials on how to use the platform.

One of the main, and most novel, features of Naviance is its scattergrams. An example can be seen in Figure 1. These are scatterplots which depict the standardized test scores and GPAs of prior applicants from a student's high school to a specific college as well as the admissions decision each applicant received. Lines on the scattergrams indicate the average GPA and SAT

[8] Naviance's website contains some tutorial videos provided to schools and ongoing professional development related to using the platform and understanding new features.

(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.[9]

It is easy for students to see how they compare to prior applicants and these lines because Naviance displays a red circle on the scattergram to mark the scores of the current user. Naviance also contains a page which summarizes the colleges a student has saved and how the student compares to the typical acceptee at each of these colleges. An example of this can be seen in Figure A.2. On this page, the typical acceptee's scores are displayed in green if the current user's score is above the typical acceptee's and red it is below the typical acceptee's. It is important to note that the typical acceptee lines are averages, not minimums, so 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 lines. Some media attention suggests that students may treat the lines as minimums more than averages, and some students become discouraged by them (Drezner, 2017; Shellenbarger, 2017; Gelger, 2018).

Scattergrams for a given college are only displayed if the high school has data on at least five applicants from prior cohorts. Some high schools further restrict this to colleges with at least ten prior applicants. During the time studied, school administrators could set this minimum by logging onto Naviance and toggling between the five or ten option in the scattergram 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.[10]

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.[11] This may

[9]Districts can choose whether to limit the number of cohorts that populate the scattergram or use all available data.

[10]It is, however, a simple way to identify the types of colleges that may be a good fit (in terms of location or culture) for students in a school.

[11]This is supported by surveys of college admissions professionals which indicate that the strength of a high school's

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.[12] 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, however, to the availability of scattergrams and the typical acceptee lines will impact what becomes visible to future cohorts.[13]

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 shown in Figure 2.

The district contains 10-15 high schools and approximately 4,000 students graduate from the district each year. I limit my main sample to the nearly 8,000 students who graduated from the district in 2015 or 2016 and for whom I have the essential data.[14] Descriptive statistics for the main sample are available in Table 1. Overall, the district is ethnically diverse. In my sample, 8% of students identify as Hispanic, 20% Black, 17% Asian, and 49% white. 21% of the students received free or reduced-price lunch (FPRL) at one point while enrolled in the district.

The district provided data on students' demographics, coursework and grades, as well as standardized test scores. These were connected to data from the National Student Clearinghouse

curriculum is one of the most important factors in an admissions decision (Clinedinst & Koranteng, 2017).

[12] One option is to use all the available data.

[13] I find that 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.

[14] 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. All students enrolled in the district's alternative high school are excluded from the analyses.

on students' postsecondary enrollment and degree completion for students who graduated high school in 2016 or earlier. The district started collecting data on students' college applications and admissions decisions in 2014. Application data are based on students' requests in Naviance for their transcripts to be sent to a college. Since students cannot apply to most colleges without sending an official high school transcript, this should capture nearly everywhere they applied.[15]

Admissions decisions are self-reported in a survey given to graduating students. Any inaccuracies in the self-reported admissions data will appear in the scattergrams. The survey response rate is approximately 90%. Many students do not report rejections, so the district treats non-responses as rejections.[16] This may lead to some under-reporting of acceptances, which will bias the acceptance criteria shown to students.[17] The direction of this bias will depend on which students are under-reporting admissions. There also appear to be a few students who over-report their acceptances, but this is less common than under-reporting.[18] The school district enters the application and admissions data into Naviance at the end of each school year, along with data on standardized test scores and GPAs. These data are used to populate the scattergrams.[19]

I use the same application data uploaded to Naviance to reconstruct the scattergrams and identify the typical acceptee profile for each college, high school, and year combination.[20] I also use

[15]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.

[16]Fifteen percent of applications are reported to end in rejections and 53% in admissions. 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.

[17]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. Districts could use NSC records to update the accuracy of the self-reported data they put into Naviance.

[18]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.

[19]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.

[20]I visually confirmed that my identification of the typical acceptee lines matched what students observed in 2017 for

these data to determine when each college-high school combination would have met the minimums of five and ten prior applicants. I am unable to determine which high schools in the district used which minimum applicant cutoff, but it appears that some schools are using each one.[21]

The district purchased Naviance in 2014. At this point, no application data were available to upload to the platform, so high school students could access all features of Naviance except for the scattergrams. In the summer of 2014, student application, admissions, and achievement data were uploaded to Naviance. Then, all high school students could see scattergrams based on the students who graduated from their high school in 2014. I focus on the students who graduated in 2015 and 2016, who were about to enter 11th and 12th grade. During the 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, the graduating students reported their admissions outcomes on the senior survey. These data were linked to the application records and added to Naviance in June 2015. These records were then combined with the data on the class of 2014 and used to update the scattergrams so that they reflected the experiences of the graduates of 2014 and 2015. 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. 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 data are available for the class of 2017. These students were in 9th grade when they could first access Naviance and 10th grade when they could first see scattergrams. They could access

twenty colleges at one high school.

[21] I lack power to detect the cutoffs for individual high schools in the data. Figure A.6 shows discontinuities at both of these thresholds. The district informed me that these were the cutoffs and I sat down with a district administrator and examined how they could change the settings. This setting no longer exists in Naviance so I cannot see what rule schools are currently using.

three different versions of the scattergrams, since they were updated when each of the previous cohort's data were added. They spent little time on Naviance during 9th and 10th grade, and most of their logins occurred during 12th grade. The login records indicate the number of times each student's account was used to log onto Naviance in each grade during their high school career.[22] I cannot tell which scattergrams a student views, but students appear to use Naviance a lot. The average student logged into Naviance 23 times in her senior year and 43 times over the course of high school.[23] Figure A.4 shows that white and Asian students use Naviance more than Black and Hispanic students. In addition, students who never received free or reduced-price lunch logged onto Naviance more than those who ever received it (while enrolled in this district).

Counselors in this district were responsible for implementing Naviance. Naviance provided introductory materials and training for the counselors.[24] The district counseling office also provided some guidance to the school departments about when and how often to log into the platform with students.[25] 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.[26] In general, counselors had autonomy over the advice they provided students and where they recommended that students apply.[27]

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. 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

[22]These may include parent logins since parents did not have accounts separate from their students. Counselors have their own accounts so if they want to show a student something or save something to the student's profile they do not need to use the student's account.

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

[24]Naviance has a general set of materials that they disseminate to schools purchasing the software, so this district received an introduction similar to what other districts would have received. Some materials can be found on Naviance's website and summaries of updates to Naviance's platform, as well as tips, are periodically emailed to counselors.

[25]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.

[26]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 lines 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 lines in students' application choices. It seems like different counselors provided different suggestions about how to use Naviance.

[27]There is no 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.

(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 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.[28]

Given these outcomes and poverty rates, 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 the summer of 2015. In 2015, the scattergrams were for roughly an 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.

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.[29] In addition, there is no significant difference in the 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.

[28] 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.

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

This pattern 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 than in the previous year (Figure A.3).[30] 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'.

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 more likely to be accepted. This is good for admissions outcomes, but may deter students from attending the highest quality college they are qualified to attend. This could prevent them from realizing the benefits associated with more selective colleges, including a higher probability of graduating and higher earnings (Chetty et al., 2017; Dillon & Smith, 2018; Goodman & Cohodes, 2014). Next, I examine the causal mechanisms driving these patterns by showing how access to colleges' scattergrams and their admissibility signals influence applications and attendance.

3 Empirical Approach

3.1 Access to Admissions Information

First, I examine how gaining access to a college's admissions information influences where students apply to and attend college. Scattergrams are displayed for all colleges with at least five or ten prior applicants from a student's high school. Each high school determines whether five or ten is the appropriate minimum number of data points.[31] I estimate the causal impact of having access to a scattergram using a regression discontinuity design around these minimum applicant requirements. I compare application and attendance probabilities for colleges with just fewer than

[30] Here I use the average admitted student in the district to define a more accurate set of admissions criteria. This also enables me to calculate averages for colleges with only a few or no students admitted at some high schools.

[31] The threshold a high school selects applies to all colleges. High school administrators can change the threshold at any time.

five or ten prior applicants to those which just met these criteria.

I do not know which high schools use which threshold, so I stack my data and simultaneously estimate the discontinuities at both thresholds.[32] I calculate a college's distance (in applications) from each threshold and include an observation for each student, college and threshold combination. Since I estimate the discontinuity at both thresholds, but only one is relevant to each scattergram, the true impact of gaining access to a scattergram is twice what I estimate.[33]

The discontinuities at both thresholds can be seen in Figure A.6, and the stacked version is shown 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 introductions on the probability that a student applies to or attends a college.

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

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.[34] For each high school, I define the set of potential scattergrams, K_j, 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.

[32] I do not have enough power to detect which threshold each school is using in each year.

[33] Crossing the visibility thresholds of five and ten increases the probability that a student saw a scattergram. The probability of having access to a scattergram changes from zero to some positive number, P, at five, and 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.

[34] I cluster at the student level because there are multiple observations per student. The standard errors are similar when I cluster them by the level of treatment (school by year by college) or using the approach described by Kolesár & Rothe (2018) for regression discontinuity designs with discrete running variables. These estimates are in table A.2.

I focus on colleges within four applications of the visibility threshold.[35] This is the maximum bandwidth I can use that is the same for both thresholds, and on each side, without including colleges with no prior applications (which do not fit the linear trend).[36] The results are robust to expanding or shrinking the bandwidth as well as to a triangular kernel specification (Table A.2).

I find no evidence of schools manipulating the number of prior applications to make scattergrams available.[37] 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. Variation in the number of prior applications comes from the popularity of a college and the number of years over which application data has been collected.[38] Application data are based on students' transcript requests. The district uploads these data to Naviance, so individual schools cannot easily manipulate them.

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

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

[35]Colleges with five to eight prior applications from a school will appear twice for the same student in my estimates of 1 because they are in the bandwidth for both thresholds.

[36]There is mean reversion (towards zero) for the number of applications a college receives from a high school. However, colleges which received zero prior applications on average receive more applications in the following year (since they can only go up from zero). This leads to a different pattern for colleges with zero prior applications than the other colleges. Since the number of applications is discrete and I have relatively few groups I cannot use traditional methods to calculate the optimal bandwidth.

[37]Figure A.5 shows no spike in the density of observations with exactly five or ten prior applications. In addition, Table A.1 shows that the colleges just above and below these thresholds are similar in terms of their sector, location, selectivity, price and enrollment.

[38]I show the results for the first year (2015) in some cases to address concerns about serial correlation in the running variable over time. Serial correlation refers to the fact that the number of prior applications can only increase over time, so a college with a scattergram can never lose one. The running variable is correlated over time because the number of prior applications, as of 2016, is the sum of the applications in 2015 and the prior applications as of 2015.

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

3.1.1 Variation in Information Effects by Admissibility Signals

Next, I examine how the impact of gaining access to admissions information varies based on its relevance and what it signals about a students' admissibility. On average, students can see 47 scattergrams. This is a lot of information for students to sort through and students will be far from the typical acceptee lines on many of the scattergrams they can view. To see how students' responses vary with the relevance of the information, I repeat the regression discontinuity analysis just described, but focus on how the magnitude of the discontinuity varies with students' proximity to the typical acceptee. For each student and college combination, I calculate the distance of the student's 11th grade GPA and maximum SAT score from the typical acceptee's.[39] I impute what students would have seen based on prior applications when a scattergram was not visible.

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) lines because they are more informative than the unweighted GPAs and SAT M+V (1600).[40] 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 students' writing scores. Results for the unweighted GPAs and 1600 scale SAT scores are very similar and are included in the Appendix.

I define *near* the typical acceptee as being within .5 GPA points and 150 SAT points of the average lines. I use this definition because it matches the optimal bandwidth described in section 3.2 and it balances tradeoffs between sample size and the concentration of the visibility effects among students who are closest to the typical acceptee lines. I estimate the coefficient $\alpha_1$in equation 1 only for students who are near the typical acceptee, and then separately for students who are not

[39]I use 11th grade GPAs because they are what a student would see when looking at the scattergram in the summer before or fall of her 12th grade, when she was choosing where to apply to college. 12th grade GPAs would not be known until after the student applied to college.

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

near it. Table A.3 and Figure A.7 show how my estimates vary with other definitions of near.

I also examine whether students who are just above and below each of the typical acceptee lines have different responses to the availability of admissions information. Here, I focus on students who are within .1 GPA points or 50 SAT points of the typical acceptee, so that students just above and below the lines are very similar. Within these bandwidths, I estimate α_1 in equation 1 separately for students who are above and below the lines. I do this separately for the SAT and GPA lines.[41] Figure 4 shows that students who could not see the scattergrams have nearly identical probabilities of applying to the college, regardless of whether they are above or below the line, but students who can see the scattergram differ in their probabilities of applying based on whether they are above or below the GPA line.

3.2 Role of Admissibility Signals

Next, I dig into how students' application choices respond to signals about their probability of admission. Naviance's scattergrams provide two clear signals about a student's probability of admission. First, they show students how similar their GPAs and SATs are to previously admitted students'. Second, they show whether a student's scores are above or below the average admitted student's. The first signal is arguably the most important because it tells the student something about her probability of admission, as well as whether the student is qualified to aim for a more selective college. The 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 probabilities. Students may, however, use the lines as heuristics or reference points because of their saliency and the complex nature of the choice they are making.

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 is above or below a line (Figure A.2). For each college, the typical acceptee lines vary across the high schools and years in which the scattergram is available.[42] This generates quasi-random variation in a student's distance from the perceived admissions criteria. I use this variation to

[41] I examine the lines separately because the joint effects of being above/below both lines are driven by the GPA line.
[42] They are also fairly noisy signals because they are typically only based on a few admitted students.

identify the causal effect of one's perceived admissions probability (as captured by one's distance to the lines) on the decision to apply to a college. I also use a regression discontinuity design to estimate whether application probabilities change discontinuously when a student crosses a line.

I estimate this using the specification in equation 3, which includes college by year and high school fixed effects (δ_{tk} and ψ_s), as well as controls for student characteristics and academic achievement ($StuCharacteristics_i$).[43] I allow for application probabilities to change discontinuously when a student moves below the typical acceptee line to account for the potential effect of this signal on student outcomes. This amounts to a regression discontinuity design around a typical acceptee line, where the coefficient β_1 indicates the extent to which being below a line has a causal impact on students' applications.[44] 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.

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

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. $LineDist_{i,k}$ represents the distance of student i's GPA or SAT from the typical acceptee's for college k. $Y_{i,k}$ is an indicator for whether student i applies to or attends college k.[45] Standard errors are clustered by student.[46] I separately estimate the impact of the GPA and SAT lines. Appendix Table A.8 describes the results when I jointly estimate the impacts of these lines (following Papay, Murnane

[43]These controls include indicators for gender, special education receipt in grade 12, free or reduced-price lunch receipt, race/ethnicity category (white, Asian, Black, or Hispanic), grade 11 GPA and maximum SAT score.

[44]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.

[45]Since there are multiple observations per student, and the treatments to which they are exposed are correlated, I focus on outcomes that are specific to each treatment. This happens because all students who are above a line at 4.1 will also be above a different college's line at 4.02. For outcomes which are constant across all of a student's observations (e.g. elite college attendance), I look separately at the impact of each college's treatment. For example, I examine the impact of each individual college's typical acceptee lines on elite college attendance (Figure 7).

[46]I cluster at the student level because there are multiple observations per student. Standard errors are similar when I cluster them by the level of treatment (school, by year, by college).

& Willet (2015) and Robins & Reardon (2012)). I focus on the separate estimates because students' responses are virtually always driven by the GPA line. I use weighted GPAs and 2400 scale SAT scores, but the results are very similar with the unweighted GPAs and 1600 scale SAT scores.

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. The results are similar but muted if I include scattergrams that would be available if five were the universal minimum. The results are robust to a triangular kernel specification as well as to a donut specification, which excludes students whose Naviance dots are on top of the typical acceptee line (Tables A.8, A.9 and A.10). The optimal bandwidths are .5 GPA points and 150 SAT points, which are consistent with the definition of near described in the previous section.[47] Columns (6) and (7) in Table 1 describe the set of observations that fall in these bandwidths. The average student is within the GPA bandwidth for 18 scattergrams and the SAT bandwidth for 11 scattergrams.

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, 2018), but I find no evidence that students manipulate their SATs to be just above the typical acceptee's (Figure A.11).[48] 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.11).[49]

3.3 Cumulative Effects of Scattergrams

Finally, I examine how the full set of scattergrams and admissibility signals a student can access influences her application portfolio, college attendance, and college selectivity. The average student in my sample had access to 47 scattergrams, but was only near the typical acceptee for 21

[47] The optimal bandwidths are calculated using the methods described in Calonico, Cattaneo, & Titiunik (2014). Results for alternative bandwidths are in Tables A.9 and A.10.

[48] The density of SAT scores and the distance of SAT scores from the lines are smooth, as per McCrary (2008).

[49] One reason I do not use unweighted GPAs is the significant bunching at the upper bound of 4.0. The density of weighted GPAs, and their distance from the typical acceptee's weighted GPA, are also smooth as per McCrary (2008). My results are based on the distance from a student's 11th grade GPA to the line apparent in the fall of her 12th grade. 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.7 shows no significant differences in observable characteristics for students just above and below the typical acceptee's GPA or SAT.

of these. My results indicate that applications were most sensitive to a college's information when the student was near the typical acceptee, so I define these as the set of relevant scattergrams. For each student, I calculate the number of relevant scattergrams to which she had access, the number that were *reach*, *match*, and *safety* colleges, as well as how many were in-state public colleges. *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 score is in the inter-quartile range.[50]

Variation in how many relevant *reach* , *match* or *safety* college scattergrams a student has access to comes from quasi-random variation in which colleges met the visibility thresholds and in the typical acceptee lines (which influence the relevance of a scattergram). To assess the cumulative effect of the scattergrams, I regress Y_i, a characteristic of the student's application portfolio or college attended, on a characteristic of the scattergrams available, SGs_i. I flexibly control for academic achievement, student characteristics and high school and year fixed effects.[51]

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

Some of the variation in the types of relevant scattergrams a student can access is due to the inaccurate nature of the typical acceptee lines. If these lines were truly accurate, then most relevant scattergrams would be match scattergrams. This may be problematic for students trying to find match schools, but it helps my analysis by generating quasi-random variation in the match quality of colleges that a student perceives to be relevant.[52]

[50] This is based on the inter-quartile range of accepted students' SAT scores reported to IPEDS in 2015. Varying definitions of *reach* , *match* and *safety* have been used in the literature. I use this measure because it is straightforward to calculate for all students. The measures used by Hoxby & Avery (2013) and Dillon & Smith (2018) are more complicated to calculate, especially since I have a non-representative sample of applicants to each college.

[51] I flexibly control for academic achievement using dummy variables for 50-point intervals of maximum SAT scores and .1 intervals of students' weighted 11th grade GPAs. I also control for race/ethnicity, free or reduced-price lunch status, gender and special education status in grade 12.

[52] This is part of why I focus on college match rather than selectivity. The selectivity of relevant scattergrams is less random because students with higher academic achievement will have access to more relevant highly selective colleges. Higher performing high schools also have more scattergrams for highly selective colleges than the other high schools.

4 Results

4.1 Access to Admissions Information

Students are significantly more likely to apply to colleges for which admissions information is visible than colleges which just miss the visibility cutoffs. Panel (A) of Figure 3 shows the 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 the visibility threshold. Panel (A), Column (1) of Table 2 reports that a student's probability of applying to a college jumps by .27 percentage points, from 1.37 percentage points to 1.64 percentage points, when a college crosses a visibility threshold.[53] This result indicates that 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.

The dashed lines in Figure 3 compare the magnitude of 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. Table 2 shows that the visibility effect is larger the more similar a student's scores are to the typical acceptee's (and Figure A.7 shows that it shrinks as students become further from the SAT or GPA lines). 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 scattergram 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).

Scattergram visibility also has larger impacts for in-state public colleges, increasing application

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

rates by 62% (1pp).[54] 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 many of these options before they saw scattergrams.[55]

The impact of admissions information on applications translates into an effect on attendance for some subgroups. The dashed lines in Panel (B) of Figure 3 show a discontinuity of .1 percentage points, in attendance rates, for students who are near the typical acceptee lines. There does not appear to be a discontinuity for students who are far from the lines or the pooled sample. Table 2 shows that once I add college by year (and student) fixed effects, this drops to an insignificant .01 percentage points. This is may be due to limited power. The college fixed effects absorb a lot of the variation in student outcomes.[56] I may also find weaker effects for attendance than applications because students have to be admitted to a college before they can choose to attend it, and students can apply to many colleges but they can only attend one.[57]

Table 3 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.[58]

[54]The null result for students near the lines at in-state public colleges may be due to the small number of students near these lines. At the in-state public colleges and less selective institutions, students' application choices do not appear very sensitive to admissions probabilities. (See Figure 6).

[55]The large effects for the FRPL and Black/Hispanic students may also be related to the fact that about 30% of the scattergrams which are relevant to them are for in-state public colleges.

[56]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.2.

[57]Attendance occurs less than 20 % as much as applications, which may inhibit my power to detect effects on it.

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

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

My main estimates are similar with alternate bandwidths, fixed effects and types of standard errors (Table A.2). They are also similar if I look at proximity to alternate versions of the typical acceptee lines. One limitation of these results is that they are local average treatment effects around the thresholds of five and ten prior applications. Thus, I also examine the impact of scattergram visibility for the full set of colleges, controlling for student and college by year fixed effects.

Table 4 shows that scattergram visibility increases applications at the average college by .9 percentage points. This is more than three times the effect I estimate for colleges near the visibility thresholds. For students near the typical acceptee lines, visibility increases applications by 1 percentage point (which is approximately double the LATE estimate). I also detect significant effects on attendance for the full set of scattergrams. Visibility increases attendance by .07 percentage points with an effect of .42 percentage points for in-state public colleges. These estimates indicate that admissions information is having large impacts on where students are applying and attending, with larger impacts for the more popular colleges (which are excluded from the LATE).

4.2 Role of Admissibility Signals

Next, I look at how application choices respond to what scattergrams signal about students' probabilities of admission. The previous section showed that access to admissions information has the largest impact on the choices of students whose scores are most similar to the typical acceptee's. Figure 4 shows that this effect is driven by both the GPA and SAT lines. Students who are 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 typical acceptee's increase applications more in response to visibility than students who are just below the GPA line.[59]

This suggests that students are most likely to respond to information when it signals something positive about their admissibility. It also indicates that students may use the typical acceptee lines as heuristics to help them determine where to apply. To more directly understand how students use admissibility signals to choose where to apply, I examine how applications vary above and below the typical acceptee lines and based on a student's perceived probability of admission.[60]

Panel (A) of Figure 5 shows how application probabilities vary by a student's distance from the typical acceptee's GPA. The navy lines indicate that students are most likely to apply to colleges at which their GPA matches or slightly exceeds the average GPA of previously admitted students. Application probabilities are decreasing in a student's distance from the GPA line. For students below the line, moving away from the typical acceptee's GPA reduces their perceived probability of admission, which then reduces the probability of applying to the college. Students also reduce applications as they move further above the typical acceptee's GPA, possibly because it signals that they are capable of gaining admission at a more selective college.

Panel (A) of Figure 5 also shows a significant gap in application probabilities at the point where a student's GPA crosses above the typical acceptee line. This is consistent with students interpreting the line as a signal about their admissibility, or using it as a heuristic 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.[61]

Row one of Table 5 shows that students just below the typical acceptee's GPA are 1.1 percentage points, or 8%, less likely to apply to the college than students just above it. Rows (2) and (3) indicate that, relative to a student whose GPA matches the typical acceptee's, being .1 GPA points

[59] These point estimates are in Column (4) of Table A.3.

[60] I separately examine the impacts of these lines because I find that students' responses are almost entirely driven by the GPA lines. In the appendix (Table A.8) I show results based on both lines as well as for other versions of the lines.

[61] 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 college is due to mean reversion and to students spreading out their applications over a larger set of colleges. This pattern is described in the next section. Figure 4 also shows a placebo test of the line effect.

above the typical acceptee reduces a student's probability of applying to the college by 1 percentage point and being .1 GPA point below it reduces the probability of applying by .85 percentage points. Thus, students prefer to apply where they have a high admissions probability, unless they are likely to be admitted to a more selective college.

Panel (B) of Figure 5 shows application rates by a student's distance from the typical acceptee's SAT. Students are most likely to apply to colleges at which their SAT score is similar to previously admitted students'. Relative to students whose scores are equal to the typical acceptee's, students whose SAT scores are 100 points above the typical acceptee's are 2 percentage points less likely to apply to the college, and those 100 points below are 1 percentage points less likely to apply. There is no apparent discontinuity at the point where a student's SAT crosses above the SAT line. This may be because there are many sources of information about colleges' SAT admissions criteria. If students find information online, or even within Naviance, which is inconsistent with what they see on scattergrams, they may not place much weight on Naviance's SAT signals.[62]

Figure 6 shows that admissibility signals are most important for decisions about applying to highly selective colleges. In terms of percent change, moving closer to the typical acceptee has the largest impact on application probabilities at highly selective colleges, and there is a significant discontinuity at the GPA line. Table 5 reports that students just below the typical acceptee's GPA at a highly selective college are 1.9 percentage points (15%) less likely to apply to the college than students just above it. No discontinuity is apparent for less selective or in-state public colleges.

Admissibility signals may be most relevant to decisions about applying to highly selective colleges because admissions probabilities are much lower at these colleges. This heterogeneity could also be driven by the types of scattergrams available to students. Students have access to more scattergrams for highly selective colleges than less selective ones.[63] 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 seem a daunting task, which could lead students to

[62]In Naviance, on a separate page, students can see the inter-quartile range of accepted SAT scores that colleges report to IPEDS for all of their applicants. There may also be a link to the college's admissions webpage which sometimes includes the SAT admissions criteria.

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

rely on a heuristic, such as the lines, to narrow down their choice set.

This story is also consistent with the larger impact of admissibility signals on the application choices of students who never received FRPL (and white and Asian students) compared to students who received FRPL, Black and Hispanic students. The latter group had access to about half as many relevant scattergrams as the first group, making the choice of which of these colleges to apply to simpler. The higher income and white and Asian students had many relevant colleges to choose among, and appear to use the admissibility signals to narrow down their choice set.

This may also explain why, on average, I find no significant effect of being below a college's GPA line on attendance at that college. The students who are using the lines to decide where to apply seem to use them to determine the highly selective colleges to which they should apply. The probability of admission at these colleges is low, and since students can only attend if they are admitted, the probability of attendance is also low. These students also tend to apply to a lot of colleges, so the probability of attending any individual college is low. These facts both contribute to limited power to detect effects on attendance. I do, however, find that a student's distance from the typical acceptee lines has a significant impact on attendance. Panel (B) of Tables 5 and 6 show that students are most likely to attend colleges where they are similar to previous admits. This is consistent with the effects on applications.

Finally, the admissibility signals in Naviance have the largest effects on students who logged onto Naviance the most and for the class of 2017. The discontinuity at the typical acceptee's GPA is more than twice as large for students who logged onto Naviance the most, compared to those who logged on the least (Table A.11). Frequent users were also most responsive to their distances from the lines. 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. In addition, the class of 2017 may have been more responsive to the typical acceptee lines than the earlier cohorts (Table A.11) because they had access to many more scattergrams (86 vs. 47) and thus faced a more complicated application choice problem (among colleges with scattergrams). This is consistent with the students using the typical acceptee lines to simplify complex choice problems.

Overall, these estimates indicate that application choices are sensitive to what the typical acceptee profiles signal about a student's probability of admission. Students application and attendance choices respond to the probability of admission conveyed in their distance to these lines and in some cases whether they are above or below the line.

4.3 Cumulative Effects of Scattergrams

So far, I have shown how information about or admissibility signals for a particular college influence applications and enrollment at that college. Now I show how the full set of information and admissibility signals that a student can access influences her choices. Students' application portfolios and attendance choices reflect the set of relevant scattergrams they could view. This set is determined by the colleges to which previous students from their school applied and the scores of those who were admitted. Whether this pushes students to attend better or worse colleges depends on the quality of the colleges to which previous students applied and how accurately the admits' scores (and self-reported admissions) reflect the college's true admissions criteria.

Table 7 indicates that students who could see more relevant scattergrams for *reach* colleges were more likely to apply to and attend *reach* colleges, while students who could see more *safety* colleges were more likely to apply to and attend *safety* colleges. A similar pattern is apparent for *match* colleges, and students who could see more relevant scattergrams were more likely to apply to and attend a *match* college, possibly because it was easier to find a match.

Most notably, among Black, Hispanic and FRPL students, every additional and relevant scattergram they saw for an in-state public college increased their probability of attending a four-year college by 2.3 percentage points. This effect is driven by the smaller and less-selective in-state public colleges.[64] It is possible that students were unaware of these options before Naviance, and that learning about nearby and in-expensive options, other than the local community college and state flagship, shifted their attendance from the local community college to one of these schools.[65]

[64] The district is located in a state with several small in-state public colleges.

[65] 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.

Scattergrams characteristics do not appear to impact the number of applications that students submit or the probability of being accepted to a college.[66] This is consistent with the fact that, on average, students apply to the same number of colleges in the years with and without scattergrams. Thus, the scattergrams are leading to substitutions in applications across colleges, not changes in the number of colleges to which a student applies. Some of this substitution is driven by the switches among *reach*, *match* and *safety* schools, as shown in Table 7. I also find that students shift applications from medium popular colleges, such as neighboring states' flagship universities, to less popular colleges in the first year with scattergrams.[67] 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 must, on average, 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.[68] Instead, students appear to shift applications to less selective colleges (Figure A.15). Figure 7 shows that students who are below the typical acceptee lines at highly selective colleges are less likely to attend an elite college than their peers who are above the lines.[69]

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 highest quality college for which they are qualified. The net effects 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.[70]

[66] There is a marginally negative impact of the number of *safety* scattergrams on the number of applications and a marginally positive impact of the number of in-state public scattergrams on acceptances.

[67] Medium popular is defined as colleges 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.

[68] 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.

[69] Elite colleges are the public and private institutions defined as elite by Barron's *Profiles of American Colleges*. I also find that students substitute their enrollment away from private colleges to in-state public colleges when they are below the typical acceptee's GPA at a private college.

[70] Among colleges near the visibility threshold, Figure 4 shows that the visibility effect is larger than the line effect.

5 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 their college choices. Providing students access to data on schoolmates' admissions experiences at a college 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).

Students increase applications most when the information conveys a positive signal about their admissibility and fit at the college. Students are most likely to apply where their scores are just above the average admit from their high school. The probability of applying to a college decreases as one's perceived probability of admissions decreases, or when the information signals that a student is likely to be accepted at more selective colleges. This suggests that students care about their admissibility when choosing where to apply to college, but biases in the admissions criteria they see can lead students to attend less selective colleges than they are qualified to attend.[71]

Information on the typical acceptee lines also deters some students from attending highly selective colleges. Students just below the GPA line are less likely to apply to a college than students just above it, especially among highly selective colleges. This reduces the selectivity of students' application portfolios and college attended, which is concerning because less selective colleges tend to have lower completion rates and are associated with lower earnings (Chetty et al., 2017; Dillon & Smith, 2018; Cohodes & Goodman, 2014). Students also appear to use the lines as heuristics to help simplify their complex application choice problem. Students who can see the most relevant scattergrams are the most likely to rely on these lines to guide their application choices.

The overall effects of Naviance's admissions information depends on the set of colleges with

[71] These biases can come from inacurracies in students' self-reported admissions outcomes, selection bias in who applies to some colleges, and noisy averages based on small samples of admitted students.

visible information and what it signals about students' probabilities of admission. Students' application portfolios reflect the set of available scattergrams that were relevant to them. Whether or not 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 college's admissions criteria. Some students are pushed to apply to and attend more *reach* colleges but others are pushed to apply to and attend more *safety* colleges. Before and after comparisons indicate that the set of scattergrams available to students in my sample may have pushed more students towards *safety* colleges than *reach* ones.[72] This may harm students' degree attainment and earnings because many students experience significant gains when they attend more selective colleges (Cohodes & Goodman, 2014; Chetty et al 2017; Dillon & Smith, 2018).

In the future, it may be valuable to more carefully curate the set of colleges for which students receive information since they are nudged towards these colleges. One way to do this would be to make information on some colleges easier to find than others. Counselors or Naviance could also identify a set of target schools and focus on disseminating information on these schools to students. In cases where insufficient data on prior applicants from a high school exist to make a scattergram visible, districts could pool data across schools to meet these minimums. Doing so may also be a good way to improve the accuracy of the admissions criteria shown to students.[73]

While the typical acceptee lines have some negative consequences, their capacity to make admissions information very salient may drive some of the positive effects associated with providing information. On net, the positive effect of providing information has a larger impact on application and enrollment choices than the negative effect of the GPA line.[74] Future work could explore how to positively harness the power of salient information while minimizing the sub-optimal responses documented here. A few ways Naviance may be able to do this include adding an inter-quartile range to the graphs (as is typically done with SAT scores) or adding a gradient of

[72]This problem may be even more severe in a lower-income district where prior students are less likely to have attended highly selective colleges.

[73]This would, however, eliminate the value that comes from the school-specific nature of the data. It is unclear if the school-specific value is more important than the accuracy that would come from a larger sample. It is also not clear that pooling across a district would increase accuracy, because this would obscure any differences in the admissions criteria applied to different schools.

[74]In particular, Figure 3 shows that the impact of showing a student a scattergram is positive for students below the typical acceptee line.

shading around the lines to depict how application probabilities change throughout the scatterplot. They could also stop making a user's score red when it is just below the typical acceptee's, perhaps turning these scores yellow, and only turning them red at a lower threshold.

Counselors or Naviance staff could also do more to assist students in accurately interpreting the scattergram data. Naviance usage, as well as the impacts of the scattergrams and the lines, vary based on the counselor to which a student is assigned.[75] Some counselors are particularly effective in eliminating student responses to the exact value of the typical acceptee line, which suggests that they may be explaining its irrelevance to students. Other counselors have students who are extremely likely to respond to the lines, so these counselors may also be using it as a tool to help with application choices. Finally, variation in the impact of information access across counselors may be driven by how much the counselor knows about or encourages certain types of colleges. Access to scattergrams may have a larger impact on students whose counselors are providing little information. Thus, a technology of this sort may be a natural complement to counselors' roles, enabling them to more efficiently serve their students.

This paper's most encouraging result is that providing low-income and minority students access to information on relevant in-state public colleges increases four-year college enrollment. These are students who were likely to be on the margin between attending a two-year or four-year college. They may have considered their only nearby and inexpensive options to be the state flagship or local community college. The scattergrams introduce them to other nearby and inexpensive options, where they are more likely to be admitted than the state flagship. This pushes them to attend a four-year college, which will likely increase their probability of earning a degree and future earnings (Goodman, Hurwitz & Smith, 2017; Zimmerman, 2014).

While I find that the presence and format of the scattergrams affect students, I cannot determine the overall effect of Naviance and how effects may vary with different implementations. 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' data are added, making more

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

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.6), however, the importance of the lines grows as students have more scattergrams to sort through.[76] Furthermore, responses to the typical acceptee's GPA lead the average GPA line to increase over time, which can reduce the accuracy of the information students see.[77]

Finally, the results presented here may understate the true effect of providing admissions information on the college choices of the average student. My estimates for the scattergram access portion of the paper are underestimates because 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 and are high performing 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. This impacts where low-income and minority students enroll in college. 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.[78]

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 (Bursztyn & Jensen, 2015). This paper, how-

[76]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.

[77]On average, they increase by 0.008 GPA points per year, with a p-value of 0.002. If they continue to increase for several years, the positive effect of scattergram visibility may become overtaken by the negative effect of information on the typical acceptee's GPA, so that the net effects of visibility are no longer positive for all application choices. Surprisingly, the SAT lines appear to get lower over time (4 points per year), but this is consistent with the lack of discontinuity at the SAT line. The reduction could be because students in this district have high SAT scores and they can see the college reported inter-quartile range of SAT scores on another page within Naviance.

[78]Cost data is unavailable for the district studied here, but a few other school districts pay less than ten dollars per student for access to Naviance.

ever, shows that nudging individuals towards social norms may not be optimal if the norms, in this case college choices, are suboptimal. In addition, data framing may lead to adverse reactions. Designers of information interventions should carefully consider how students may respond to information, and recognize that information may not have the desired, or even positive, effects.

Finally, this paper indicates that information about admissibility at a college is an important piece of the application choice problem. Students may, however, place too much weight on their admissibility. I find that small changes in perceived admissions probabilities reduce applications to highly selective colleges. Given the high returns to many of these colleges, and the low cost of applying to them, it is probably not optimal for students to respond as strongly as they do. These reactions can reduce the quality of college that a student attends, which may impact future labor market outcomes (Cohodes & Goodman, 2014; Chetty et al 2017; Dillon & Smith, 2018).

6 References

Akerlof, George A., and Rachel E. Kranton. "Identity and schooling: Some lessons for the economics of education." *Journal of Economic Literature* 40, no. 4 (2002): 1167-1201.

Avery, Christopher, and Thomas J. Kane. "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, 2004.

Barr, Andrew, and Sarah Turner. "A Letter and Encouragement: Does Information Increase Postsecondary Enrollment of UI Recipients?." *American Economic Journal: Economic Policy* 10, no. 3 (2018): 42-68.

Bergman, Peter, and Eric Chan. Leveraging technology to engage parents at scale: Evidence from a randomized controlled trial. *Working paper* (2017).

Bettinger, Eric P., Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu. "The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment." *The Quarterly Journal of Economics* 127, no. 3 (2012): 1205-1242.

Bond, Timothy N., George Bulman, Xiaoxiao Li, and Jonathan Smith. "Updating human capital decisions: Evidence from SAT score shocks and college applications." *Journal of Labor Economics* 36, no. 3 (2018): 807-839.

Bursztyn, Leonardo, and Robert Jensen. "How does peer pressure affect educational investments?." *The Quarterly Journal of Economics* 130, no. 3 (2015): 1329-1367.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. "Robust nonparametric confidence intervals for regressiondiscontinuity designs." *Econometrica* 82, no. 6 (2014): 2295-2326.

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

Castleman, Benjamin L., and Lindsay C. Page. ”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 (2015): 144-160.

Castleman, Benjamin L., and Lindsay C. Page. ”Freshman year financial aid nudges: An experiment to increase FAFSA renewal and college persistence.” *Journal of Human Resources* 51, no. 2 (2016): 389-415.

Chade, Hector, Gregory Lewis, and Lones Smith. ”Student portfolios and the college admissions problem.” *Review of Economic Studies* 81, no. 3 (2014): 971-1002.

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

Clinedinst, Melissa, and Anna-Maria Koranteng. “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. ”Merit aid, college quality, and college completion: Massachusetts’ Adams scholarship as an in-kind subsidy.” *American Economic Journal: Applied Economics* 6, no. 4 (2014): 251-85.

Dettling, Lisa J., Sarena Goodman, and Jonathan Smith. ”Every little bit counts: The impact of high-speed internet on the transition to college.” *Review of Economics and Statistics* 100, no. 2 (2018): 260-273.

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

Drezner, D.W. (2017, September 20). “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* 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

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

Fu, Chao. ”Equilibrium tuition, applications, admissions, and enrollment in the college market.” *Journal of Political Economy* 122, no. 2 (2014): 225-281.

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

Goodman, Joshua, Oded Gurantz, and Jonathan Smith. ”Take Two! SAT Retaking and College Enrollment Gaps.” *National Bureau of Economics Research* Working Paper No. 24945 (2018).

Goodman, Joshua, Michael Hurwitz, and Jonathan Smith. ”Access to 4-year public colleges and degree completion.” *Journal of Labor Economics* 35, no. 3 (2017): 829-867.

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

Hastings, Justine S., and Jeffrey M. Weinstein. ”Information, school choice, and academic achievement: Evidence from two experiments.” *The Quarterly Journal of Economics* 123, no. 4 (2008): 1373-1414.

Hoxby, Caroline, and Christopher Avery. ”The missing” one-offs”: The hidden supply of high-achieving, low-income students.” *Brookings Papers on Economic Activity* 2013, no. 1 (2013): 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. ”Student responsiveness to earnings data in the College Scorecard.” *Economic Inquiry* 56, no. 2 (2018): 1220-1243.

Kahneman, Daniel. ”Reference points, anchors, norms, and mixed feelings.” *Organizational behavior and human decision processes* 51, no. 2 (1992): 296-312.

Kahneman, Daniel. ”Maps of bounded rationality: Psychology for behavioral economics.” *American Economic Review* 93, no. 5 (2003): 1449-1475.

Kolesár, Michal, and Christoph Rothe. ”Inference in regression discontinuity designs with a discrete running variable.” *American Economic Review* 108, no. 8 (2018): 2277-2304.

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

Luca, Michael, and Jonathan Smith. ”Salience in quality disclosure: evidence from the US News college rankings.” *Journal of Economics & Management Strategy* 22, no. 1 (2013): 58-77.

Pallais, A. (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. ”High-school exit examinations and the schooling decisions of teenagers: Evidence from regression-discontinuity approaches.” *Journal of Research on Educational Effectiveness* 7, no. 1 (2014): 1-27.

Reardon, Sean F., and Joseph P. Robinson. ”Regression discontinuity designs with multiple rating-score variables.” *Journal of Research on Educational Effectiveness* 5, no. 1 (2012): 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.

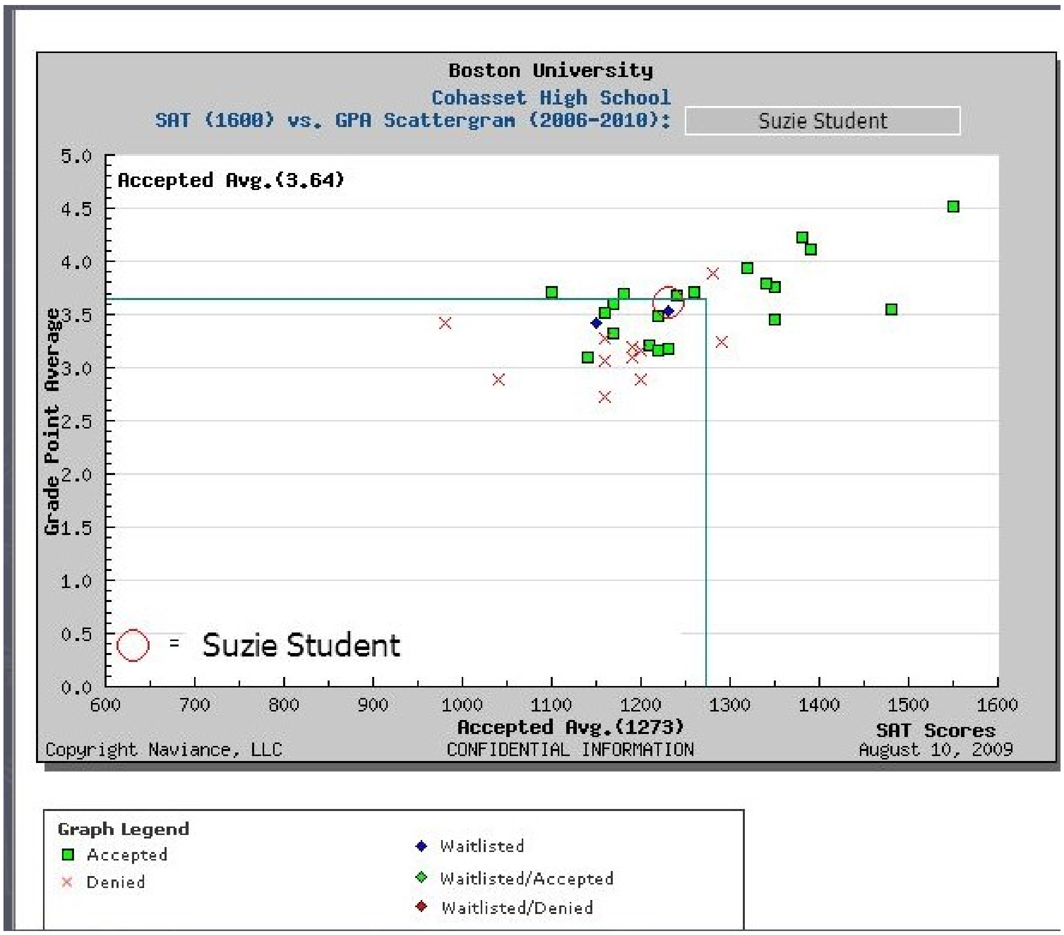
Shellenbarger, Sue. “College-Search Quandry? There’s an App for That: Dozens of digital tools aim to help students find their dream school.” *The Wall Street Journal*. September 19, 2017. Retrieved from: https://www.wsj.com/articles/college-search-quandary-theres-an-app-for-that-1505832484

Shulman, Robyn D. “EdTech Investments Rise To A Historical $9.5 Billion: What Your Startup Needs to Know.” *Forbes.* January 26, 2018. 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. ”The returns to college admission for academically marginal students.” *Journal of Labor Economics* 32, no. 4 (2014): 711-754

7 Tables and Figures

Figure 1: Example Scattergram

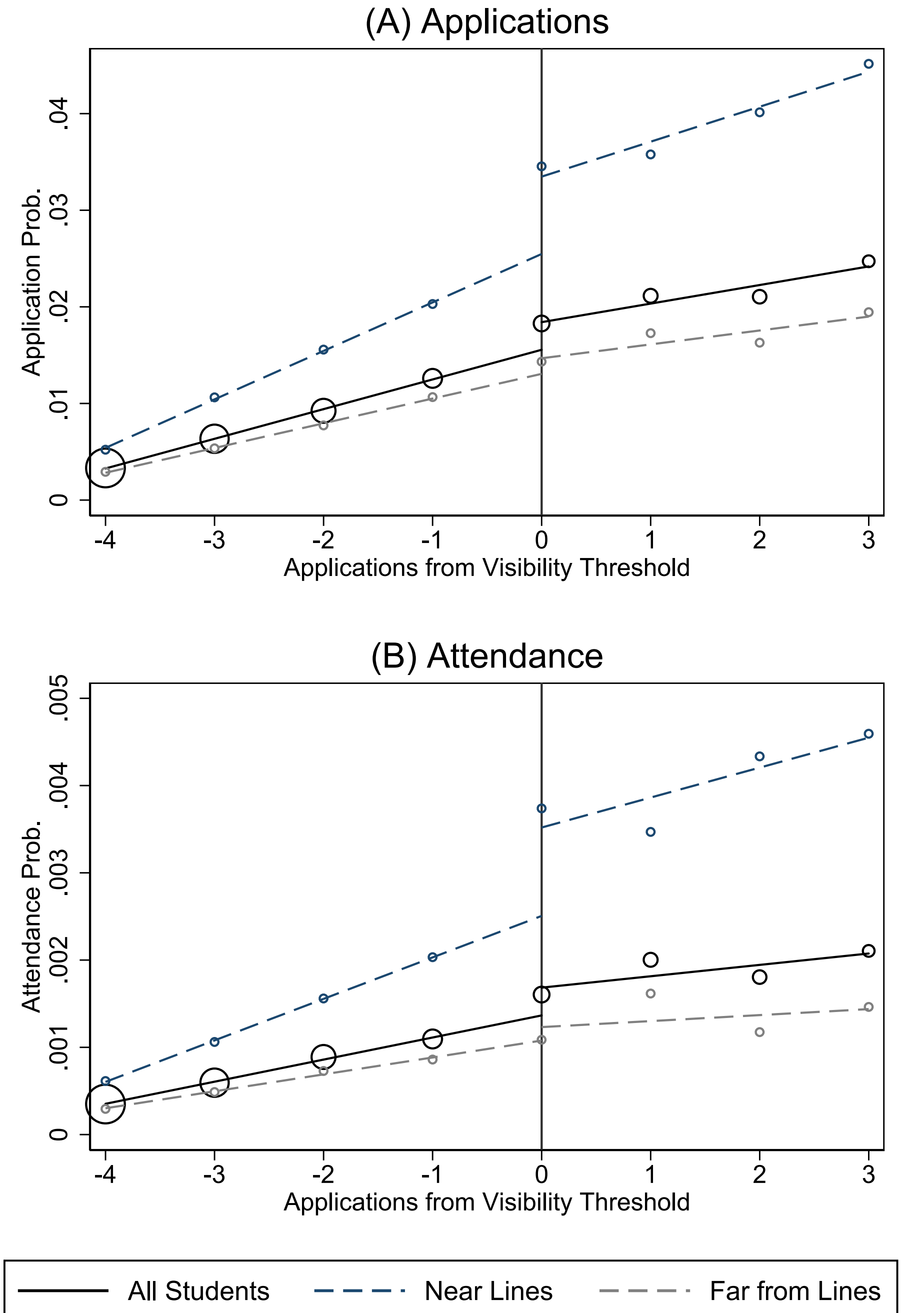


Notes: Photo credit: Naviance LLC. This is a fictional example of a scattergram. The red circle represents the GPA and SAT score of the student currently viewing the scattergram. The blue lines represent the average GPA and SAT score for students from the same high school. Naviance updated the scattergram 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

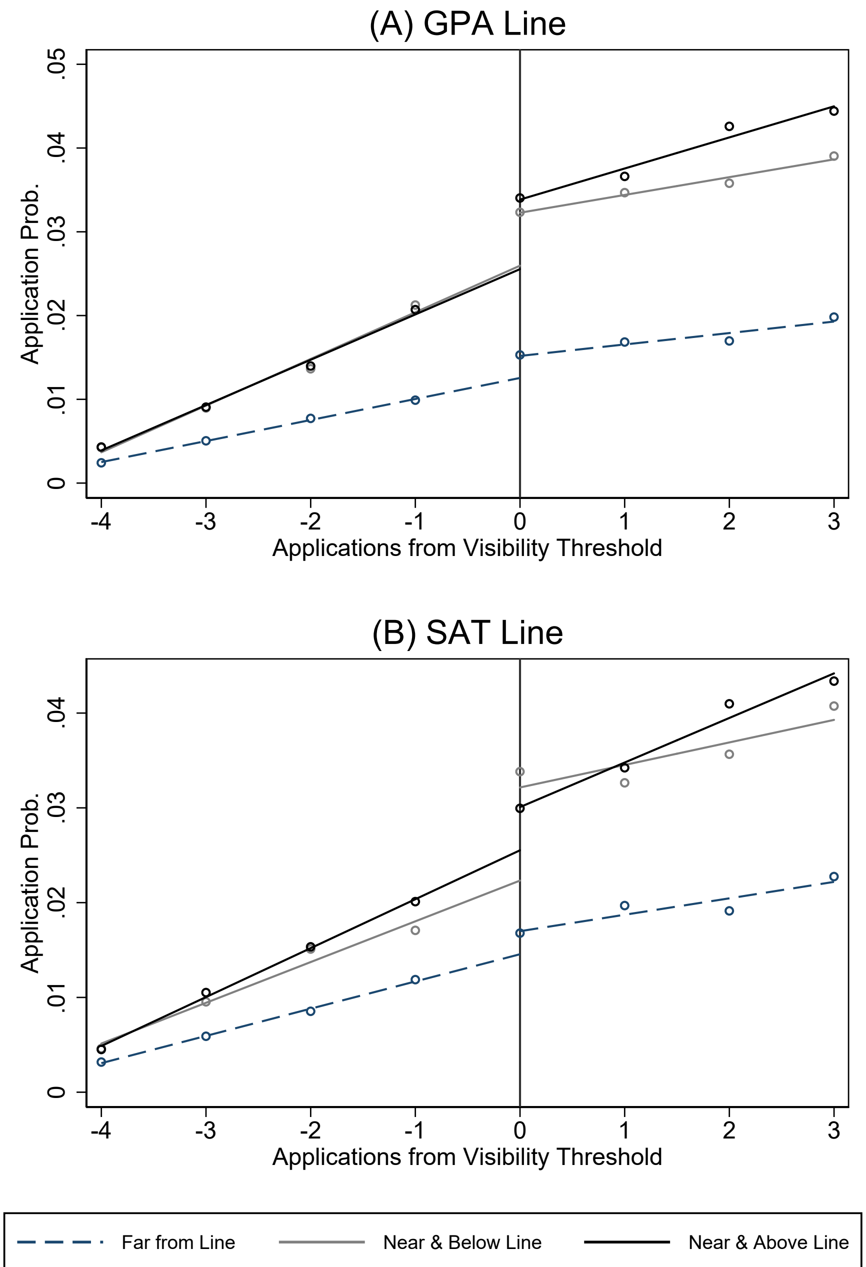
	2014: Naviance purchased (no data for scattergrams)	**2015:** Scattergrams Available (based on 2014 data)	**2016:** Additional Scattergrams and Data Available	**2017:** Additional Scattergrams and Data Available	*Data Available*
Class of **2014**	**Grade 12:** Applying to College	**Entering College/ 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



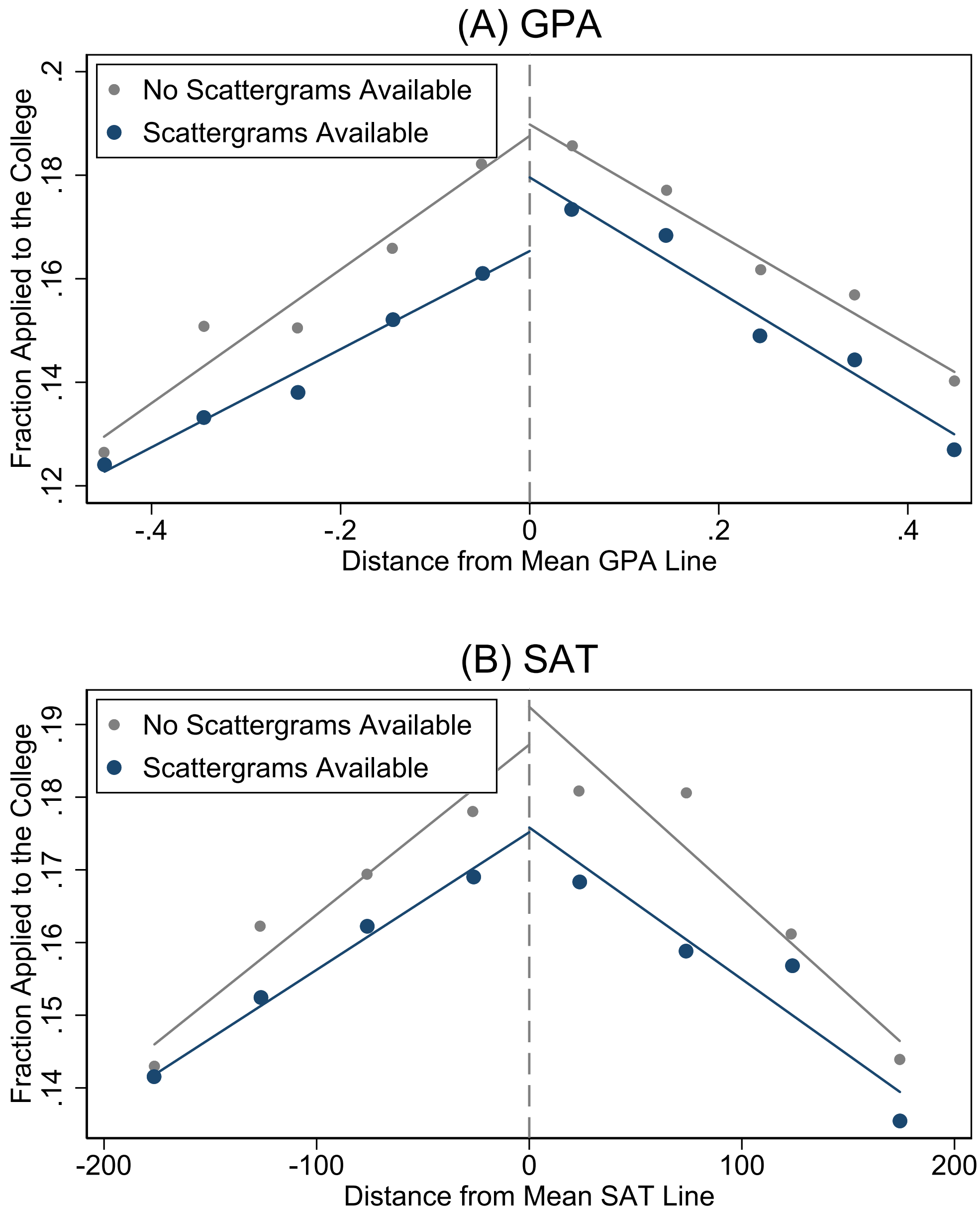
Notes: The figures above show how the probability of applying to (A) or attending (B) a college changes when a college crosses a scattergram visibility threshold, and how this varies based on the student's similarity to previously accepted students (as measured by proximity to the typical acceptee lines). 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 Presence on Applications by Proximity to Typical Acceptee Lines



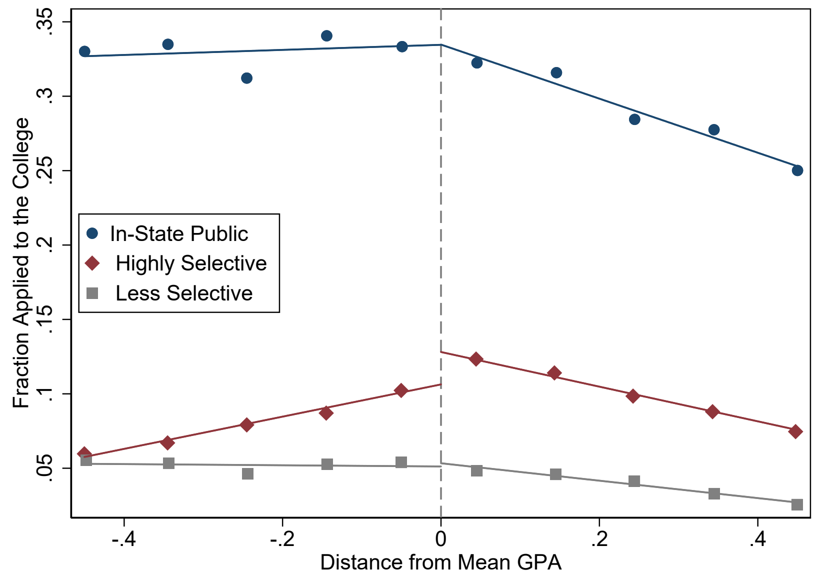
Notes: The figures above show how the probability of a student applying to a college changes when a college crosses a scattergram visibility threshold, and how this varies based on the proximity of the student to the typical acceptee lines. Panel (A) is based on the weighted GPA lines and near is defined 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



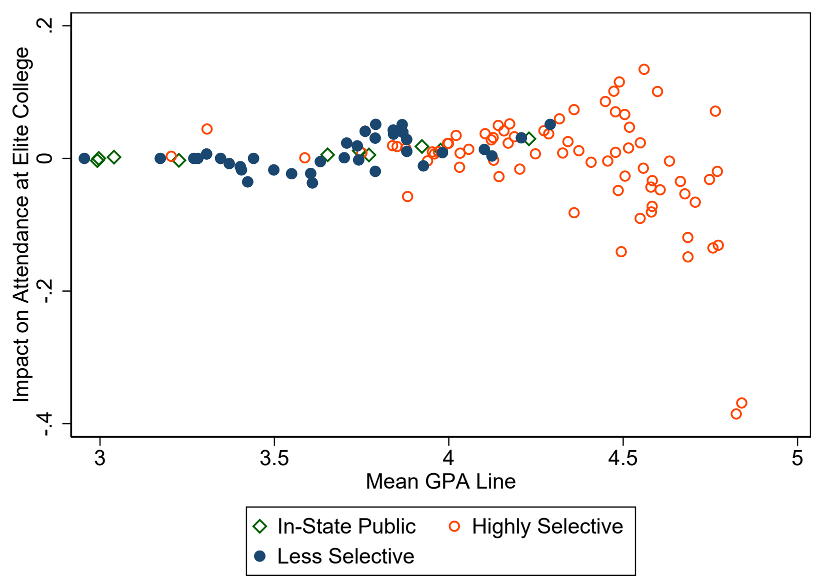
Notes: The figures above show how application rates varied based on a student's position on a scattergram relative to the typical acceptee's GPA (A) and SAT (B). The navy lines are based on students from the years in which scattergrams were available (2015-2016) and the gray lines are based on students 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

	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	Year 2014	Year 2015	Year 2015
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(A) Demographics										
White	0.49	0.58	0.16	0.75	0.00	0.57	0.57	0.52	0.49**	0.49**
Asian	0.17	0.17	0.16	0.25	0.00	0.22	0.22	0.16	0.17	0.17
Black	0.20	0.14	0.45	0.00	0.72	0.11	0.11	0.19	0.20	0.21**
Hispanic	0.08	0.05	0.18	0.00	0.28	0.05	0.05	0.07	0.08	0.08
Free/Reduced Lunch	0.21	0.00	1.00	0.10	0.47	0.10	0.09	0.19	0.21**	0.22**
(B) Academics										
GPA (11th gr. weighted)	3.41	3.58	2.79	3.64	2.88	3.97	3.89	3.41	3.42	3.40
SAT (M+V+W)	1689	1740	1444	1765	1477	1821	1836	1698	1695	1683
Attend 4-yr Coll	0.60	0.67	0.32	0.68	0.42	0.81	0.82	0.59	0.59	0.60
Attend 2-yr Coll	0.24	0.20	0.38	0.21	0.32	0.11	0.11	0.23	0.24	0.25
(C) Applications										
Applications	5.15	5.50	3.89	5.41	4.60	6.51	6.70	5.21	5.17	5.13
Num Reach Apps	1.53	1.40	2.10	1.23	2.34	1.40	1.26	1.59	1.48*	1.59
Num Match Apps	2.31	2.53	1.38	2.55	1.73	3.01	3.25	2.40	2.34	2.29*
Num Safety Apps	1.29	1.48	0.50	1.55	0.63	1.84	1.91	1.30	1.31	1.27
Highly Selective	3.98	4.30	2.63	4.33	3.07	5.23	5.25	4.09	3.97	3.99
Less Selective	1.73	1.62	2.22	1.48	2.41	1.42	1.44	1.72	1.72	1.74
Acceptances	2.51	2.78	1.53	2.77	1.92	3.36	3.43	2.35	2.55***	2.47***
(D) Attendance										
Reach	0.19	0.17	0.31	0.15	0.33	0.26	0.09	0.18	0.18	0.19
Match	0.54	0.54	0.54	0.55	0.51	0.43	0.63	0.56	0.55	0.54
Safety	0.27	0.29	0.16	0.30	0.16	0.31	0.29	0.26	0.28*	0.27
Highly Selective	0.56	0.60	0.32	0.62	0.37	0.71	0.69	0.55	0.56	0.55
Less Selective	0.44	0.40	0.68	0.38	0.63	0.28	0.31	0.45	0.44	0.44
(E) Scattergrams										
Total	47	50	38	51	39	59	59	33	33	62
In GPA Bandwidth	18	20	8	21	10	32	30	12	12	22
In SAT Bandwidth	11	13	5	13	6	19	22	0	8	14
Relevant	21	23	10	25	12	36	35	12	15	26
Relevant In-State Public	4	4	3	4	3	5	5	3	3	4
Relevant High Selectivity	14	17	6	18	7	27	26	9	11	18
N	7,647	6,004	1,643	5,005	2,156	134,023	85,451	3,758	3,733	3,914

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. Column (3) contains students who received it at least once 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). 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. Column (8) contains all students who graduated from the district in 2014. These students could not see any scattergrams. Columns (9) and (10) 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. Selectivity ratings are based on Barron's 2009 selectivity index. Where this is missing, selectivity rankings from IPEDS in 2002 are used.

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. 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 3: Impact of Scattergrams by Student and College Characteristics

		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) Apply								
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)
Visitble & Near Lines	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	0.0137	0.0146	0.0121	0.0136	0.0144	0.0185	0.0159	0.0106
CCM Near Line	0.0278	0.0279	0.0287	0.0270	0.0340	0.0782	0.0375	0.0178
(B) Attend								
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)
Visible & Near Lines	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	0.0011	0.0013	0.0008	0.0012	0.0011	0.0022	0.0009	0.0013
CCM Near Line	0.0034	0.0037	0.0023	0.0037	0.0024	0.0192	0.0033	0.0030
N	2,565,375	2,031,177	534,198	1,696,273	708,692	63,947	1,001,540	1,499,888
N Near Lines	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. 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. 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 4: Impact of All Scattergrams 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) Apply								
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)
Visitble & Near Lines	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	0.0079	0.0086	0.0056	0.0081	0.0075	0.0096	0.0094	0.0062
CCM Near Lines	0.0180	0.0181	0.0174	0.0172	0.0207	0.0228	0.0251	0.0119
(B) Attend								
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)
Visible & Near Lines	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	0.0007	0.0008	0.0004	0.0007	0.0006	0.0016	0.0007	0.0007
CCM Near Lines	0.0018	0.0018	0.0011	0.0017	0.0020	0.0039	0.0019	0.0016
N	8,914,720	7,018,780	1,895,940	5,844,978	2,503,108	300,304	2,939,362	5,062,656
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 or attendance 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 5: Impact of Mean GPA Lines

		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) Applied								
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
(B) Attendance								
Below GPA	0.000	0.000	0.000	-0.001	0.001	0.007	-0.000	0.002
	(0.002)	(0.002)	(0.006)	(0.002)	(0.004)	(0.009)	(0.002)	(0.002)
Dist Above GPA	-0.011*	-0.012**	-0.016	-0.016**	-0.019**	-0.010	-0.001	-0.001
	(0.006)	(0.006)	(0.019)	(0.006)	(0.009)	(0.026)	(0.009)	(0.003)
Dist Below GPA	-0.025***	-0.020***	-0.045***	-0.022***	-0.024***	-0.111***	-0.009	-0.006
	(0.005)	(0.005)	(0.016)	(0.006)	(0.007)	(0.029)	(0.007)	(0.004)
CCM	0.026	0.025	0.034	0.026	0.028	0.113	0.010	0.005
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, which are on a five point scale. The outcome in panel (A) is applying to the college associated with the scattergram treating the student. The outcome in panel (B) is attending that college. 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. 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. CCM refers to the mean attendance or application probability for students with GPAs just above the typical acceptee's.

Table 6: Impact of Mean SAT Lines

	All	Free/Reduced Lunch Never	Free/Reduced Lunch Ever	White or Asian	Black or Hispanic	In-St. Public Colleges	Other Colleges High Sel.	Other Colleges Less Sel.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) Applied								
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
(B) Attendance								
Below SAT	0.0013	0.0020	-0.0011	0.0031*	0.0083*	0.0088	-0.0013	0.0037*
	(0.0019)	(0.0019)	(0.0065)	(0.0018)	(0.0048)	(0.0095)	(0.0016)	(0.0019)
Dist Above SAT	-0.0001***	-0.0001***	-0.0001	-0.0001***	-0.0000	-0.0002**	-0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0000)
Dist Below SAT	-0.0000*	-0.0000***	0.0000	-0.0001***	-0.0000	-0.0001	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0000)
CCM	0.025	0.024	0.037	0.027	0.019	0.111	0.007	0.005
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 in panel (A) is applying to the college associated with the scattergram treating the student. The outcome in panel (B) is attending that college. 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. 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. CCM refers to the mean attendance or application probability for students with SATs just above the typical acceptee's.

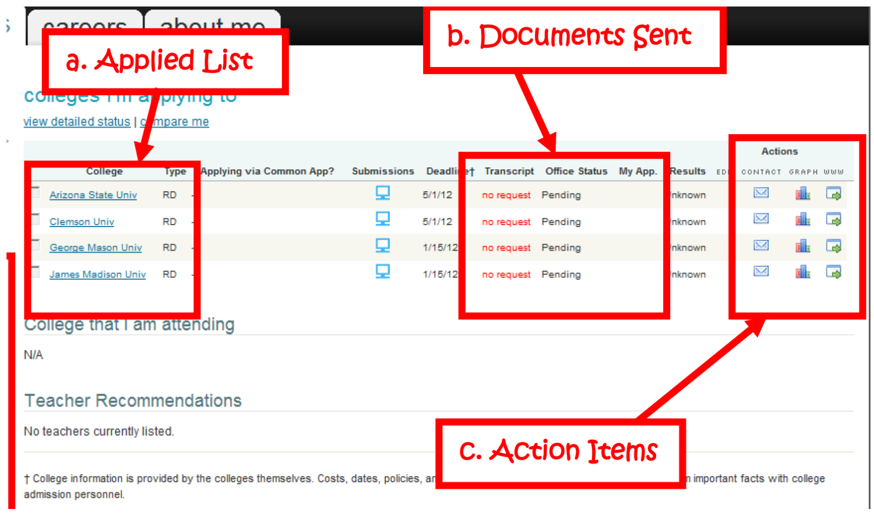
Table 7: Cumulative Impact of Scattergrams

	Applications				Acceptances	Attend College			
	Num.	Reach	Match	Safety		Reach	Match	Safety	4-yr
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(A) All Students									
Total SGs	-0.010	-0.028***	0.029***	-0.011***	0.001	-0.003***	0.005***	-0.003***	-0.001
	(0.008)	(0.004)	(0.005)	(0.004)	(0.005)	(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
	(0.037)	(0.020)	(0.023)	(0.017)	(0.022)	(0.003)	(0.005)	(0.003)	(0.004)
Match SGs	-0.006	-0.029***	0.072***	-0.051***	0.002	-0.003***	0.010***	-0.009***	-0.001
	(0.012)	(0.006)	(0.007)	(0.005)	(0.007)	(0.001)	(0.002)	(0.001)	(0.001)
Safety SGs	-0.028*	-0.078***	-0.035***	0.083***	-0.003	-0.011***	-0.001	0.011***	-0.000
	(0.016)	(0.009)	(0.010)	(0.007)	(0.009)	(0.001)	(0.002)	(0.002)	(0.002)
In-St Public SGs	0.036	-0.089***	0.164***	-0.042*	0.056*	-0.014***	0.034***	-0.011**	0.008
	(0.053)	(0.029)	(0.033)	(0.024)	(0.031)	(0.005)	(0.007)	(0.005)	(0.005)
N	5,176	5,176	5,176	5,176	5,176	5,176	5,176	5,176	5,176
(B) Minority & FRPL									
Total SGs	-0.007	-0.026**	0.039***	-0.015**	0.010	-0.004**	0.008***	-0.004***	-0.000
	(0.019)	(0.013)	(0.010)	(0.006)	(0.010)	(0.002)	(0.002)	(0.001)	(0.002)
Reach SGs	0.002	0.143***	-0.003	-0.130***	0.023	0.023***	-0.009	-0.020***	-0.006
	(0.080)	(0.054)	(0.043)	(0.026)	(0.044)	(0.008)	(0.010)	(0.006)	(0.009)
Match SGs	-0.015	-0.036**	0.056***	-0.029***	0.015	-0.004	0.010***	-0.006***	0.001
	(0.027)	(0.018)	(0.014)	(0.009)	(0.015)	(0.003)	(0.003)	(0.002)	(0.003)
Safety SGs	-0.019	-0.113***	0.056**	0.042***	-0.001	-0.022***	0.018***	-0.001	-0.004
	(0.048)	(0.032)	(0.026)	(0.016)	(0.026)	(0.004)	(0.006)	(0.004)	(0.005)
In-St Public SGs	0.053	-0.049	0.163***	-0.045	0.030	-0.014	0.043***	-0.010	0.023**
	(0.103)	(0.070)	(0.055)	(0.034)	(0.056)	(0.010)	(0.012)	(0.008)	(0.012)
N	1,409	1,409	1,409	1,409	1,409	1,409	1,409	1,409	1,409

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. 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. 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 be accepted to a college in order to attend it.

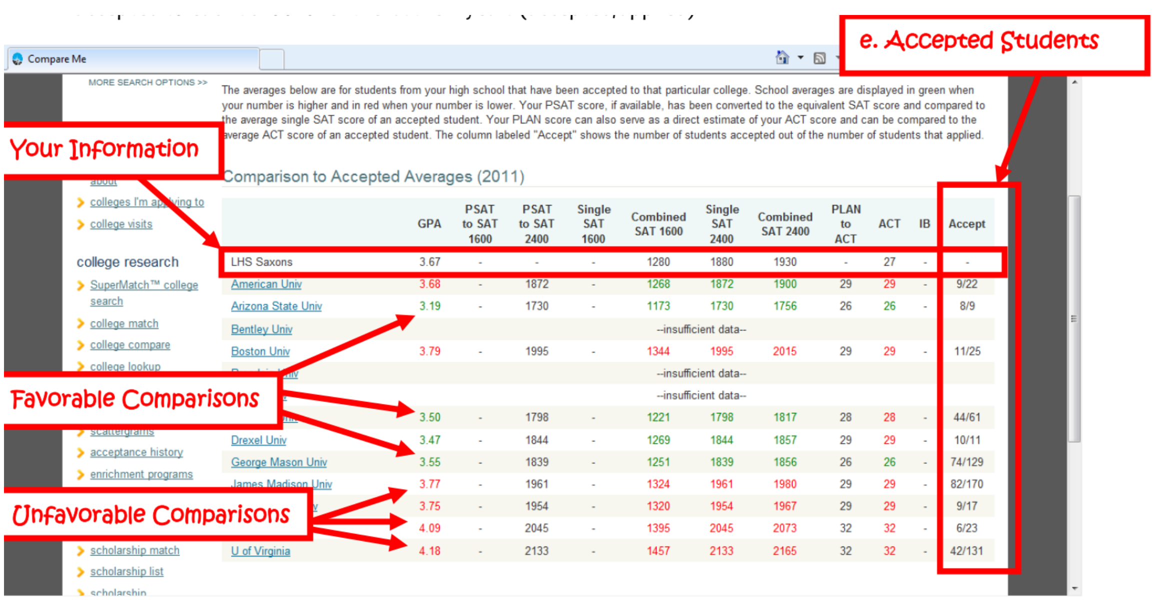
A Appendix

Figure A.1: Example of College Dashboard on Naviance



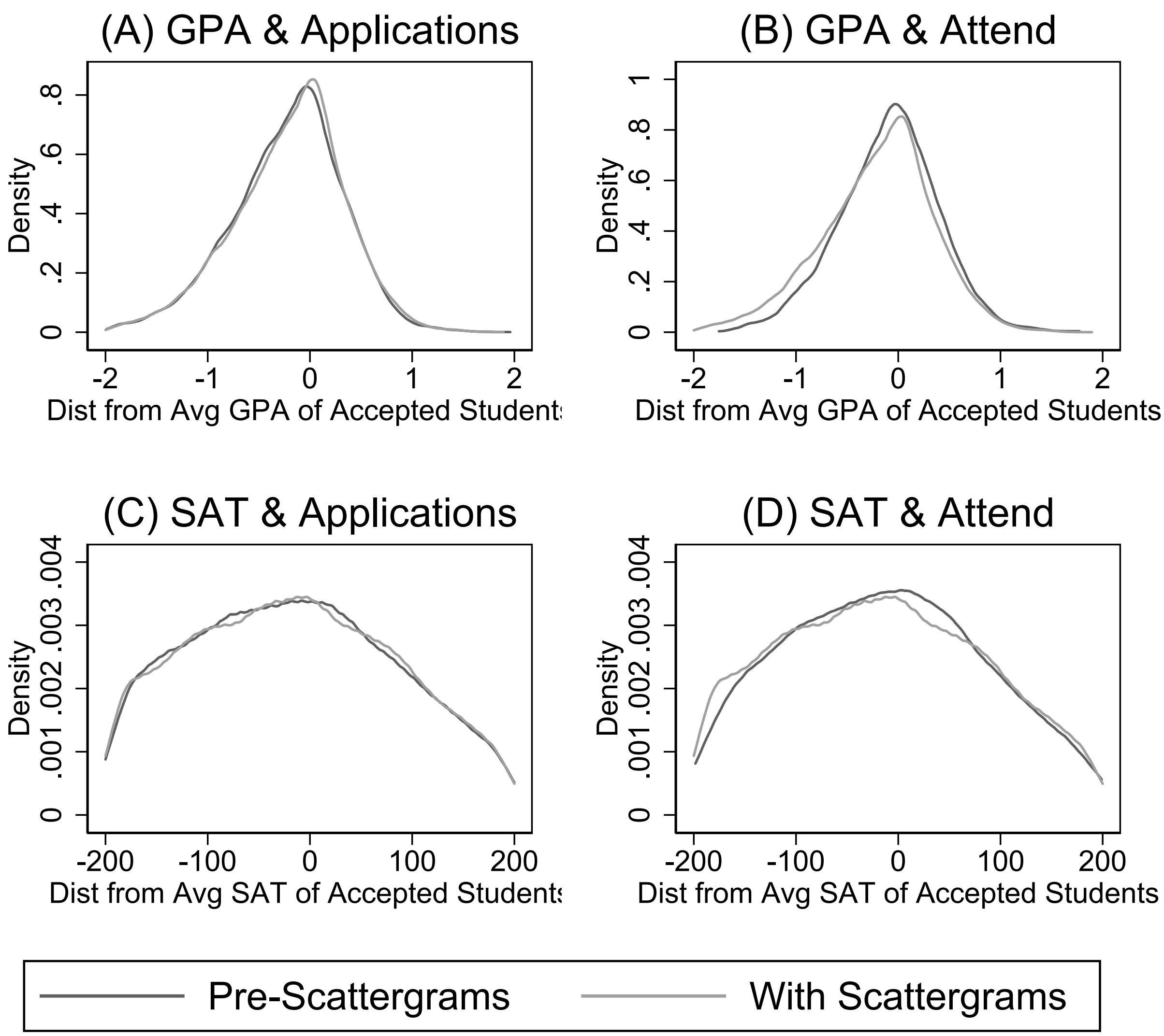
Notes: This is an example of the college dashboard on Naviance. (This is not from the district studied). The red boxes were added as notes by the high school posting instructions on how to use Naviance. Source: Langley High School and Naviance.

Figure A.2: Example of College Comparisons on Naviance



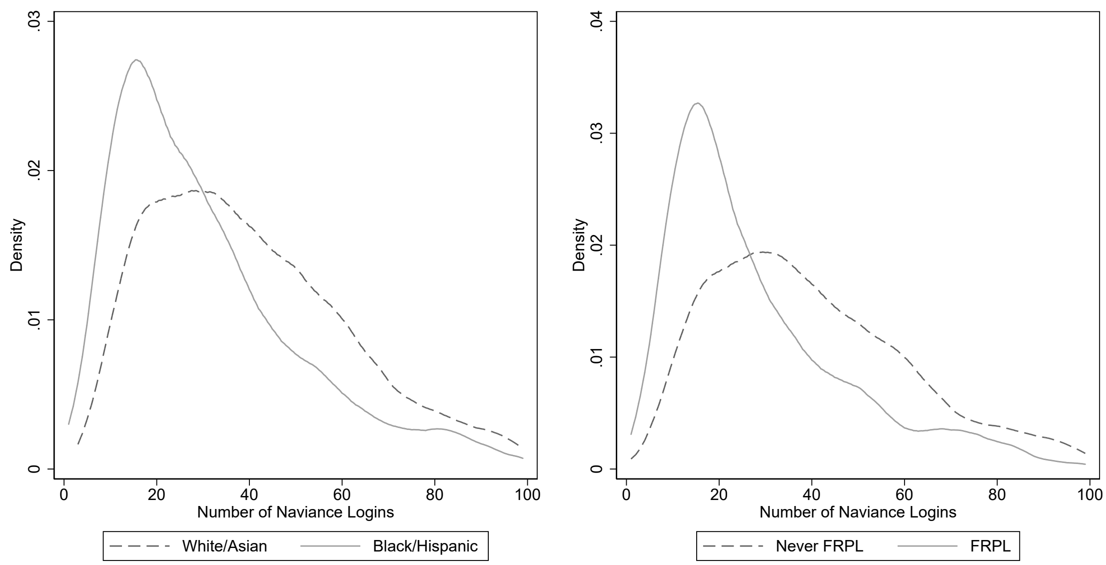
Notes: This is an example of the how students can compare colleges on Naviance. (This is not from the district studied). The red boxes were added as notes by the high school posting instructions on how to use Naviance. Source: Langley High School and Naviance.

Figure A.3: Application and Attendance Density by Distance from Mean GPA or SAT



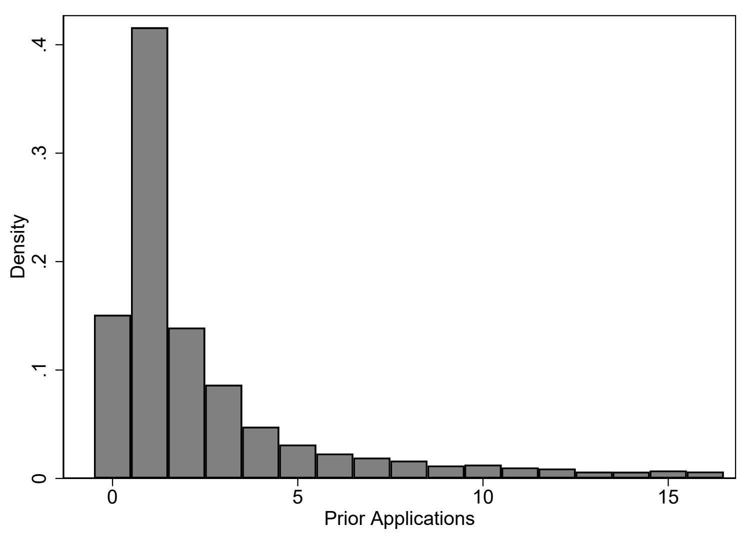
Notes: The figures above show how the types of colleges to which students applied or attended shifted when scattergrams became available. In particular, they show the densities of applications (A and C) and attendance (B and D) as a function of the student's GPA or SAT distance from the average GPA or 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. These 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.4: Density of Naviance Login Rates for the Class of 2017



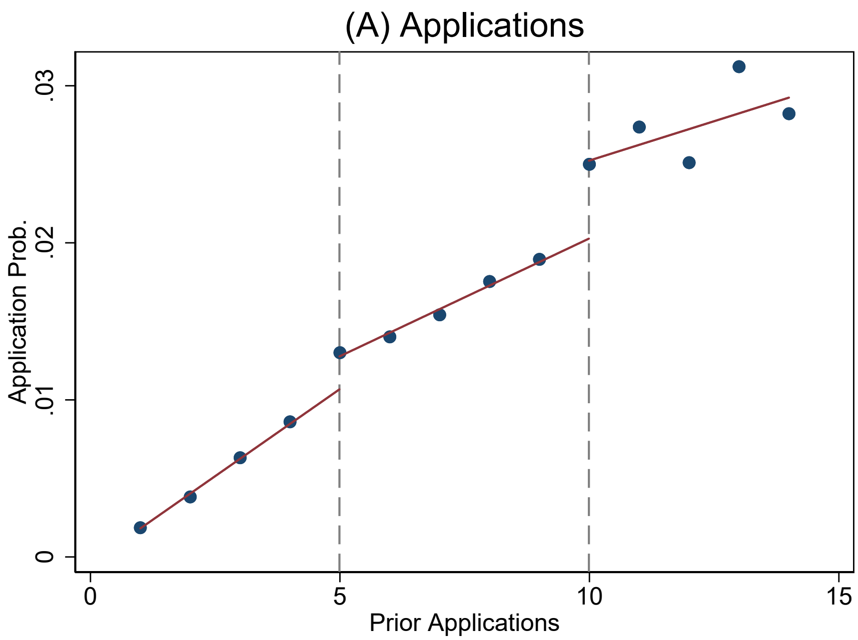
Notes: The figures above indicate the densities for the number of times students logged onto Naviance. These login rates are only available for the class of 2017 but they include all logins since the fall of 2014. (These include all logins through the student's account and thus may capture parents or 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.

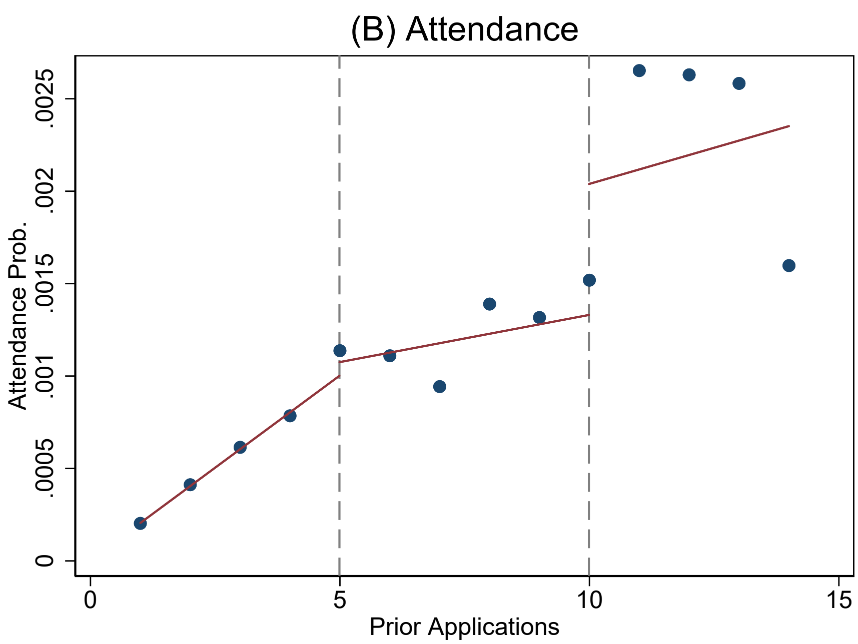
Figure A.5: Density of Prior Applications



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.

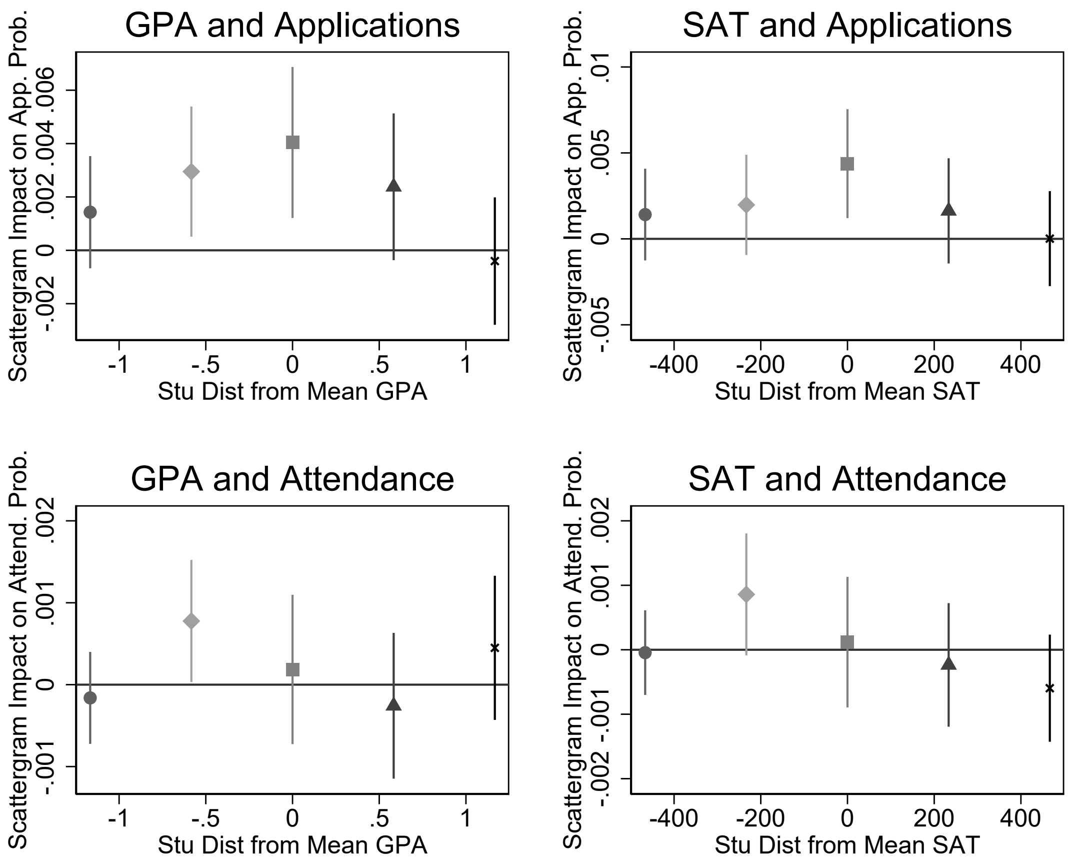
Figure A.6: Discontinuities at 5 and 10 Prior Applications





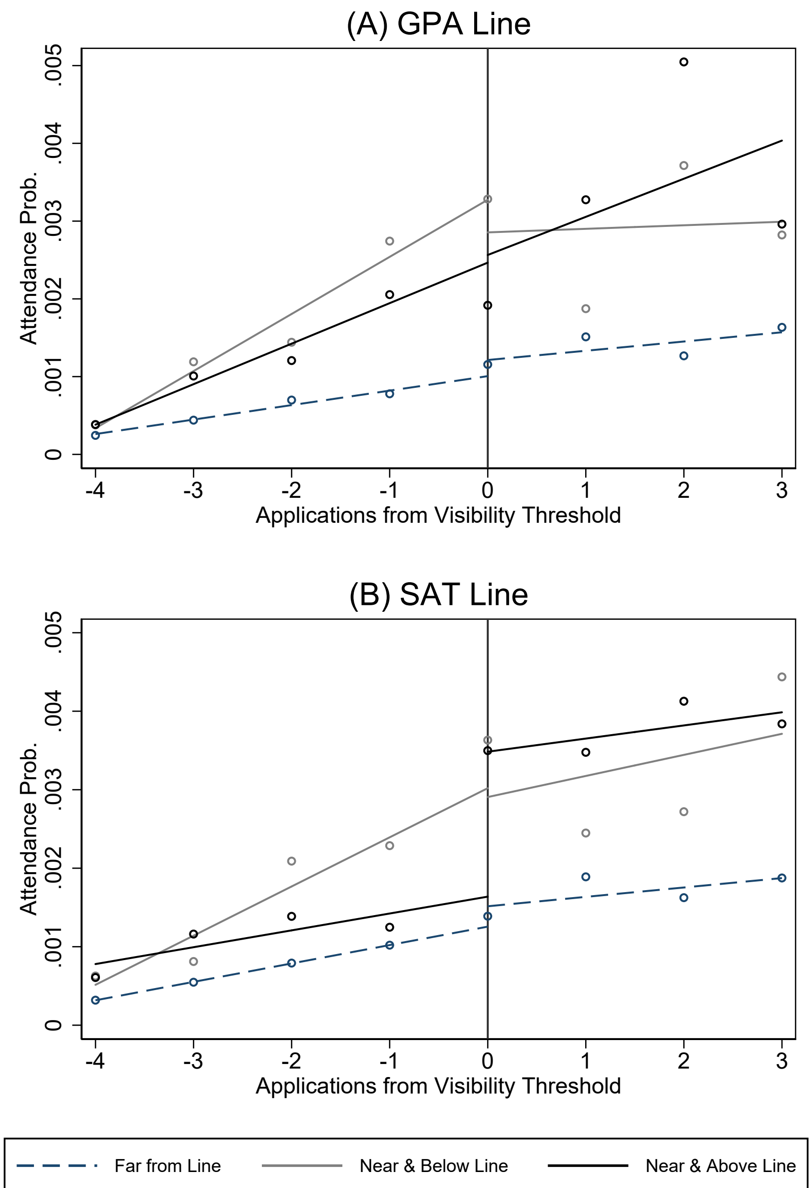
Notes: The figures above show how the probability of a student applying to (A) or attending a college (B) changes based on the number of applications a college has received from the student's high school prior to the student's year of graduation (2015 or 2016). Each dot on the x-axis represents the exact 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.7: Magnitude of Scattergram Impact by Proximity to Typical Acceptee



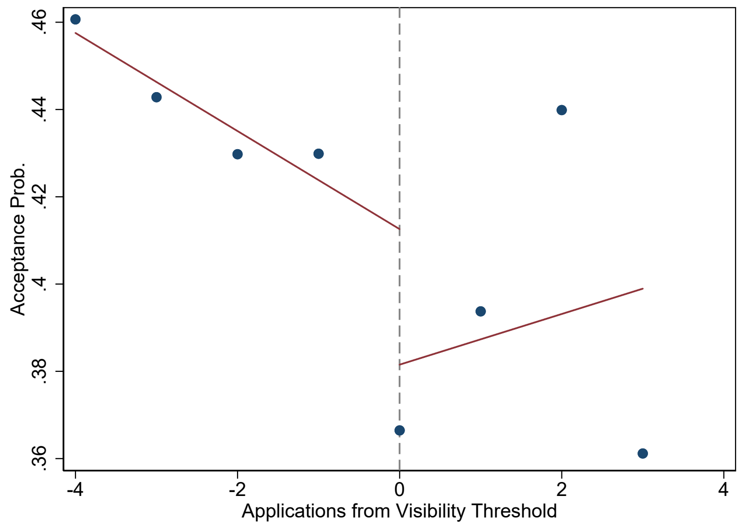
Notes: The above figures show how the magnitudes of the discontinuities in application or attendance probabilities at the visibility thresholds vary based on how similar the students were to the typical acceptee's from their high school. I construct these by separately estimating discontinuities for 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.8: Impact of Scattergram Presence on Attendance by Proximity to Typical Acceptee



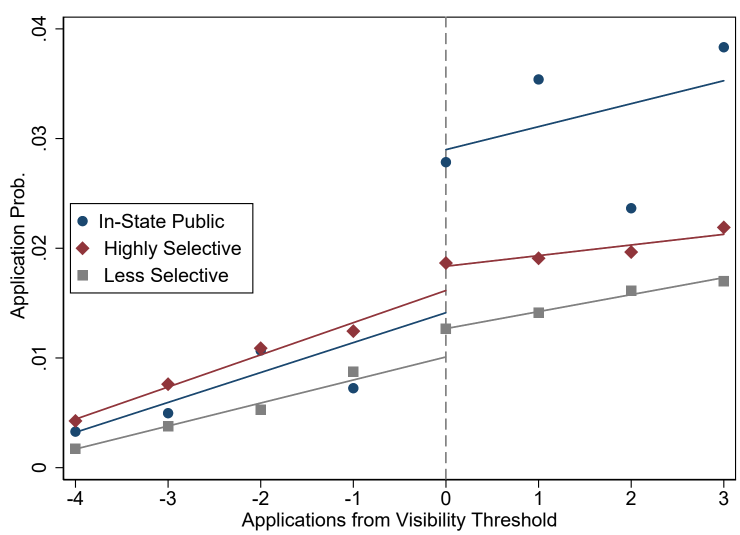
Notes: The figures above show how the probability of a student attending a college changes when a college crosses a scattergram visibility threshold, and how this varies based on the proximity of the student to the typical acceptee lines. Panel (A) is based on the weighted GPA lines and near is defined 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.9: Impact of Scattergram Visibility on Acceptance Probabilities (Conditional on Applying)



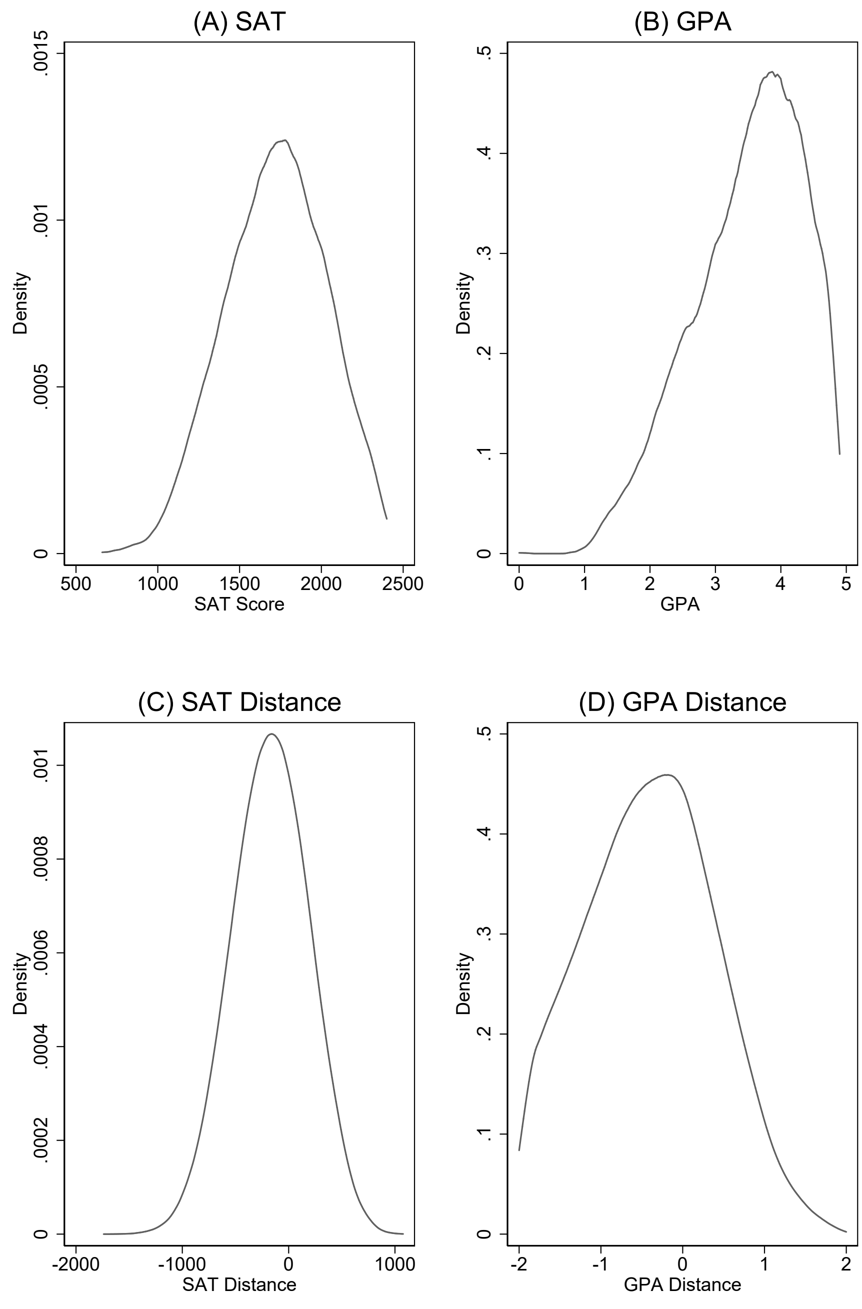
Notes: This figure shows how the probability of a student being accepted to a college, conditional on applying, changes when a college crosses a scattergram visibility threshold. A college's scattergram becomes visible to students after it receives five or ten applications from the student's high school. (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.10: Impact of Scattergram Visibility on Application Probability by College Type



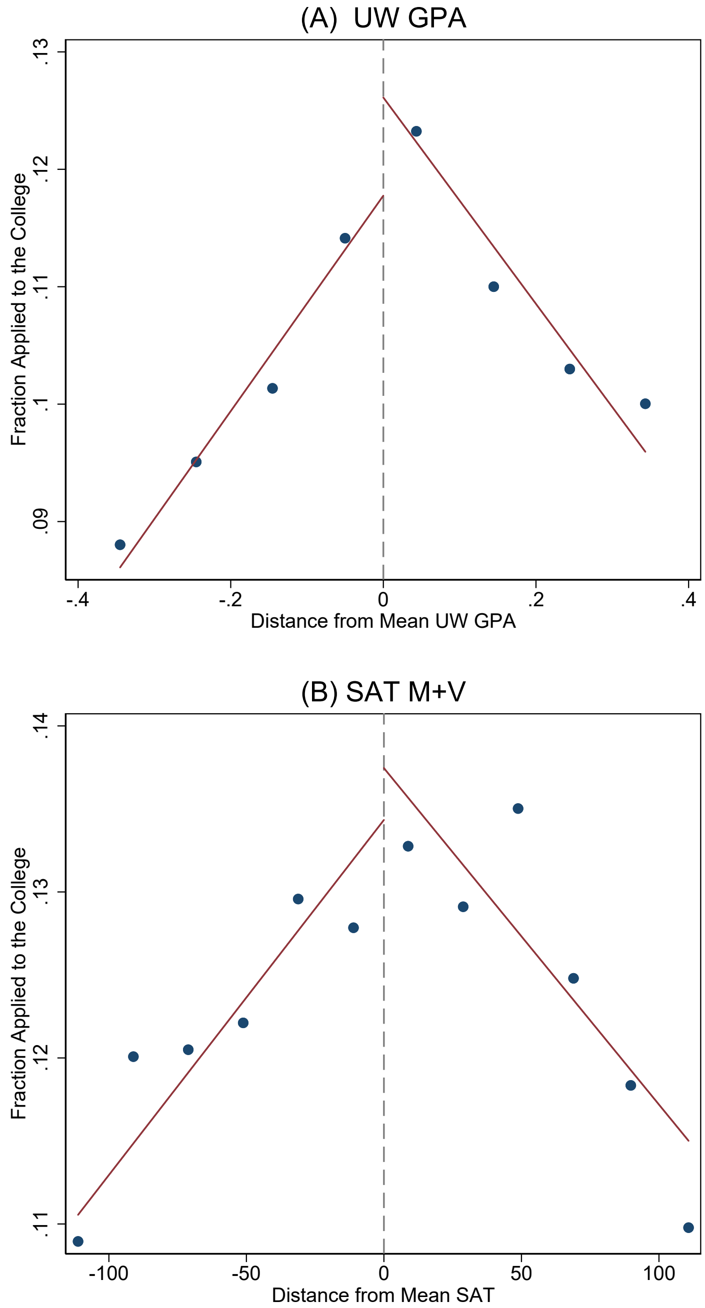
Notes: This figure compares the fraction of students who attended a college with the distance of the student's weighted 11th grade GPA from the typical acceptee line she could see and the type of college. Observations are student-college combinations, and the college in this pair must have received at 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.11: Densities of SAT scores and GPAs



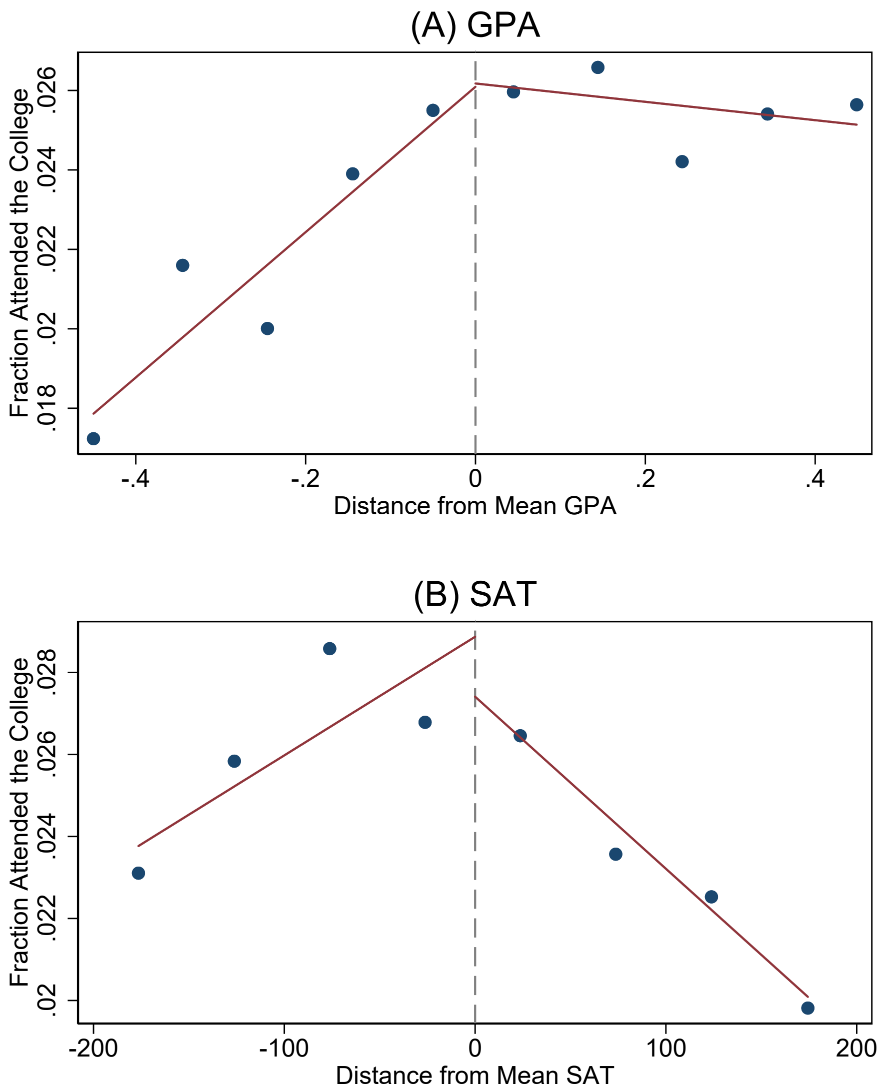
Notes: The top row shows the densities for SAT scores (the old version on the 2400 scale) in panel (A) and weighted 11th grade GPAs in panel (B). The bottom row shows the densities for the distance of the student's SAT score (old 2400 version) and weighted GPA from the typical acceptee's SAT or GPA on 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.12: Application Probabilities by Distance from Other Lines



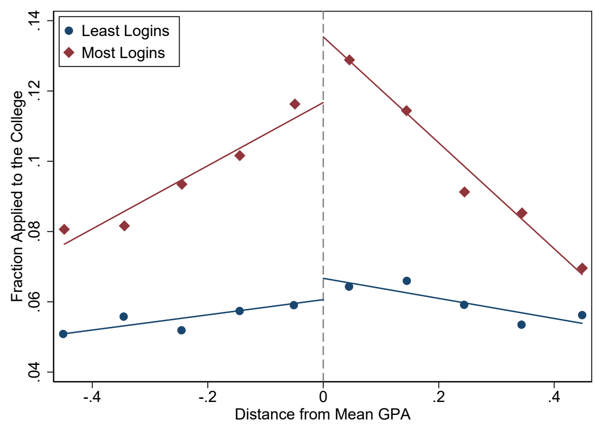
Notes: The figures above show how attendance rates varied based on a student's position on a scattergram relative to the typical acceptee's GPA (in panel A) and SAT (in panel B). For panel (A), unweighted GPAs from 11th grade are used to determine the distance from the mean unweighted GPA line depicted 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).

Figure A.13: Attendance Probability by Distance from Typical Acceptee Lines



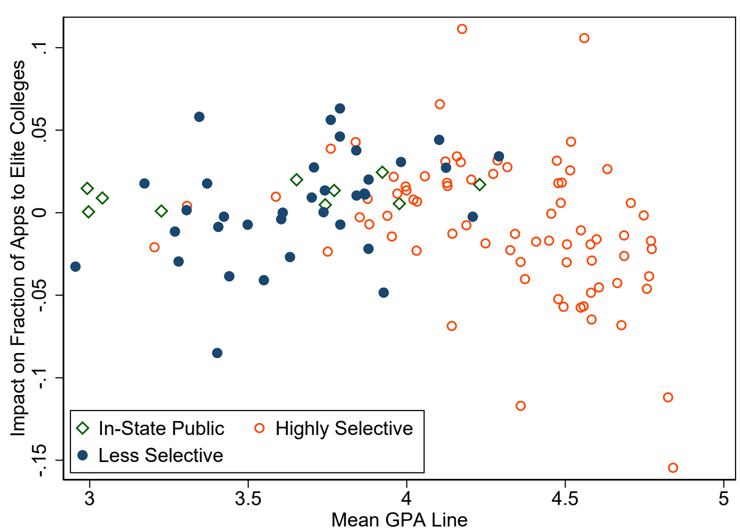
Notes: The figures above show how attendance rates varied based on a student's position on a scattergram relative to the typical acceptee's GPA (in panel A) and SAT (in panel B). For panel (A), Weighted GPAs from 11th grade are used to determine the distance from the mean weighted GPA line depicted on 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.14: 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.15: 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.

Table A.1: Balance Table for Colleges with and without Scattergrams

	Private (1)	Out-of-State Public (2)	In-State Public (3)	Selectivity Tier (4)	Net Price (5)	Enrollment (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.2: 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	0.0078***	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) and (6)-(8) include student and college by year fixed effects. Column (2) contains the standard errors described by Kolesar & Rothe (2018) for discrete running variables and the coefficients generated from a regression with no fixed effects, a bandwidth of 4 and a smoothness constant of .1. 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.3: Scattergram Impact on Applications and Attendance by Distance to GPA and SAT Lines

		W GPA			SAT M+V+W		
	All	BW 1	BW .5	BW .1	BW 300	BW 150	BW 50
	(1)	(2)	(3)	(4)	(5)	(6)	(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.4: Bandwidth Comparisons for Subgroup Results

		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) BW: +/- 4								
Visible	0.003***	0.003***	0.002*	0.002***	0.003***	0.010***	0.003***	0.002***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)
N	2,565,375	2,031,177	534,198	1,696,273	708,692	63,947	1,001,540	1,499,888
(B) BW: +/- 3								
Visible	0.003***	0.003***	0.002	0.003***	0.002**	0.008***	0.003***	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)
N	1,236,352	981,868	254,484	822,763	336,925	38,327	591,279	606,746
(C) BW: 1-14								
Visible	0.003***	0.003***	0.001	0.002***	0.003***	0.010***	0.003***	0.002***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)
N	2,619,088	2,074,990	544,098	1,733,040	722,306	64,786	1,037,882	1,516,420
(D) BW: 0 - 20								
Visible	0.002***	0.002***	0.001	0.002***	0.002***	0.007***	0.001***	0.003***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)	(0.002)	(0.000)	(0.000)
N	4,513,998	3,549,164	964,834	2,953,417	1,273,067	108,947	1,520,290	2,884,761

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.

Table A.5: 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.6: Scattergram Impacts by Year

	2015	2016	2017	2015-2016	2015-2017
	(1)	(2)	(3)	(4)	(5)
(A) Applied					
Visible	0.0030***	0.0023***	0.0009**	0.0026***	0.0019***
	(0.0008)	(0.0005)	(0.0003)	(0.0004)	(0.0003)
Visible & Near Lines	0.0046	0.0060***	0.0019*	0.0055***	0.0035***
	(0.0031)	(0.0019)	(0.0010)	(0.0017)	(0.0010)
(B) Attended					
Visible	0.0002	0.0002	0.0002		
	(0.0002)	(0.0001)	(0.0001)		
Visible & Near Lines	0.0001	0.0004	0.0003		
	(0.0010)	(0.0006)	(0.0005)		
N	116,292	165,105	273,977	281,397	555,374
N Near Lines	1,091,913	1,527,175	1,721,299	2,619,088	4,340,387
Avg Number Scattergrams	33	62	86	47	53

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.7: 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.8: 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.9: 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.10: 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.11: 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).