Every Little Bit Counts: The Impact of High-Speed Internet on the Transition to College

Lisa J. Dettling[1] - Federal Reserve Board of Governors
Sarena Goodman - Federal Reserve Board of Governors
Jonathan Smith - College Board

October 2015

Abstract

This paper investigates the effects of high-speed Internet on students' college application decisions. We link the diffusion of zip code-level broadband Internet to millions of PSAT and SAT takers' college testing and application outcomes, and find that students with access to high-speed Internet in their junior year of high school perform better on the SAT and apply to a higher number and more expansive set of colleges. Effects appear to be concentrated among higher-SES students, indicating that while, on average, high-speed Internet improved students' postsecondary outcomes, it may have increased pre-existing inequities by primarily benefiting those with more resources.

[1] Email: lisa.j.dettling@frb.gov, sarena.f.goodman@frb.gov, and jsmith@collegeboard.org

The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the Board of Governors, the College Board, or the research staff at either institution. We thank Michael Lovenheim, Amanda Pallais, Melissa Kearney, Bekah Richards, and seminar participants at the 2015 Association for Education Finance and Policy and Society for Research and Educational Effectiveness conferences for helpful comments and suggestions.

I. Introduction

College planning is complicated: there are thousands of colleges in the United States, each with countless attributes, and students face uncertainty in both admission and completion. Moreover, students encounter hurdles and barriers throughout the lengthy college planning and application process. These barriers could be procedural, such as lack of instructions on how and when to register for an admissions exam, as well as informational, such as lack of detailed information about colleges that may help produce a better fit. These complications, along with the many steps involved in the application process (Klasik, 2012), have contributed to widespread inequality in college access (Bowen et al., 2009; Hoxby and Avery, 2013; Smith et al, 2013).

While college access efforts, programs, and organizations are designed to help students overcome the optimization problem they face, there is an emerging literature that demonstrates that students often do not have full or correct information when making this consequential decision (Dillon and Smith 2013; Hoxby and Turner, 2015). A natural question follows - how can we encourage students to apply to and attend colleges that best fit their needs? Prior successful interventions have aimed to provide students with additional information (Avery and Kane, 2004; Carrell and Sacerdote, 2013) and change the way information is presented (Bettinger et al., 2012; Hoxby and Turner, 2013 and 2015). Other studies have found that students' decisions are quite sensitive to small changes in information or costs.[2] Thus, in theory, a technology that increased the availability or improved the presentation of information could also generate large changes in college-going.

This paper investigates whether the dramatic and conditionally-random diffusion of high-speed broadband Internet over the last decade affected students' college application behavior. Between 2000 and 2013, high-speed Internet usage increased from 3 percent to 70 percent.[3] Recent research concludes that the rollout of broadband to households had an impact on a host of outcomes, such as academic achievement (Vigdor et al., 2014; Faber et al., 2015), labor force participation (Dettling, forthcoming), voter turnout (Falck, Gold and Heblich, 2014), and criminal behavior

[2] Evidence of students' responsiveness to small changes in information or costs include interventions that change 1. rules of thumb (Pallais, 2015), 2) the salience of college rankings (Luca and Smith, 2013), 3) financial aid offers (Cohodes and Goodman, 2014), 4) application fees and essays (Smith et al. 2015), and 5) admissions exam taking (Bulman, forthcoming; Goodman, 2013; Klasik 2014; Hurwitz et al. 2015).

[3] High-speed internet usage rates were obtained from PEW Research at www.pewinternet.org/data-trend/internet-use/connection-type/.

(Bhuller et al., 2011). An important mechanism emphasized in this research is ease and speed of information acquisition. In this paper, we tie the two strands of literature together and examine whether the diffusion of high-speed Internet – which could make it easier and less costly to obtain information about the college application process – affected students' application decisions.

Our main empirical strategy links administrative test-taker data from the College Board to a zip code-level measure of broadband Internet availability during students' junior year of high school. We derive Internet availability from data collected by the Federal Communications Commission (FCC) on the number of broadband Internet Service Providers (ISPs) in each zip code from 1999 to 2007, a time when Internet prevalence was skyrocketing. Using information on millions of students' Preliminary SAT/National Merit Scholarship Qualifying Test® (PSAT) scores, SAT scores, and SAT Score Sends (as a proxy for applications), we first probe the extent to which Internet availability induced SAT score changes and then examine a range of college application outcomes. Specifically, we investigate whether students with broadband Internet are more likely to apply to a four-year college, relatively many colleges, relatively selective colleges, an academically matched college, top liberal arts colleges, the flagship university within a student's state, and an out-of-state college.

Our findings indicate that increased availability of the Internet in a student's zip code unambiguously improves her admissions exam scores and her application set. On average, her test score improves by 0.7 scale points and her application portfolio increases in both size and quality, with the magnitude of these changes ranging from 0.2 to 0.4 percentage point, depending on the specific outcome we consider. Some of these results are particularly striking when we scale by mean application rates; for example, about 7.2 percent of the SAT-taking population apply to a very selective liberal arts college, but we find that Internet availability increased that rate by almost 0.2 percentage point (or 3 percent). Further, while an applicant's performance on the SAT and her application set are undoubtedly intertwined, we find that the improvements in her application portfolio persist beyond what can be explained by her improved score: less than one-fifth of changes in her application outcomes can be traced to Internet-driven increases in average test scores.

For a causal interpretation of our estimates, we must assume that within our framework, high-speed Internet is exogenous to student's testing and application outcomes. We examine the validity of this assumption in three ways. First, we confirm that early and late adopting zip codes had similar trends in outcomes before high-speed Internet was available in the late 1990s. Second, we show that the availability of high-speed Internet in the student's home zip code in subsequent years – i.e., when she is presumably in college – has no effect on her application outcomes. Last, because we observe outcomes only for students who take an exam administered by the College Board, to mitigate selection concerns, we demonstrate that our main estimates are robust to the exclusion of states in which the ACT—a college admissions exam that competes with the SAT—prevails among college-going students.

Even as high-speed Internet became more accessible across the United States, it is likely that not every type of student benefited equally from its availability.[4] In particular, it is well known that high-speed Internet take-up rates and computer accessibility vary by socioeconomic status.[5] That said, the research on college under-match suggests teens from a lower socioeconomic background may be the least informed about college, and hence, may stand to gain the most from high-speed Internet. When we estimate the effects of Internet availability separately by group, we find that application improvements appear to be driven by higher-income students, students in urban areas, white students, and those with more-educated parents. These findings likely reflect an array of known impediments between availability and effective use for lower-resource groups – i.e., differential household adoption rates, within-household access to a computer, and use of the Internet to acquire and distill information on colleges. Altogether, our results indicate that while, on average, the Internet improved students' postsecondary outcomes, it may have widened existing inequities, favoring those with more resources.

The contribution of our paper is fourfold. First, we provide the first causal estimates of the effects of high-speed Internet availability on college admissions testing and

[4] Moreover, even if there were equally improved access to the efficiencies of high-speed Internet—which there is not —the general equilibrium effects are less clear, since students may be competing with one another for spots. This is beyond the scope of this paper.

[5] For example, the October 2003 CPS indicates that among 15-18 year olds whose mother did not complete high school, 8 percent had access to broadband at home and 55 percent had access to a computer at home. Among 15-18 years olds whose mother has a post-graduate degree, 46 percent had access to broadband at home and 97 percent had access to a computer at home. By 2009, when high speed Internet access was nearly universal, only 58 percent of 15-18 year olds with a mother who did not complete high school had access to broadband at home, compared with 95 percent of 15-18 year olds whose mother has a post-graduate degree. Computer ownership was not asked in 2009.

application behavior.[6] Second, we add to a literature which finds that small cost reductions, and, specifically, improved access to information, can improve postsecondary outcomes. We demonstrate that significantly many students, if given the opportunity, appear to be able to obtain and distill information on colleges and universities for themselves. We also show that some students are potentially being left behind by the information age; because unobserved barriers limit their access, because these students require more guidance on how to benefit from their access, or some combination of the two. Third, given that universal broadband Internet is a central policy goal, our finding that Internet availability can change college-going outcomes, but that the benefits may unequally accrue to higher-resource students, suggests a more intensive intervention may be necessary to realize the full benefits of increased availability in underserved areas. Last, we make a methodological contribution by developing a way to measure Internet availability at the local level that matches aggregate usage patterns.

II. Conceptual Framework and Related Literature

There is a large and growing literature in economics on college choice from which we derive three key findings that serve as the bedrock of our analysis. One, there are substantial returns to college quality (e.g., Card, 1995; Black and Smith, 2006; and more recently, Zimmerman, 2014), particularly among disadvantaged students (Dale and Krueger, 2002, 2011). Two, despite the potential for large returns, many students, especially disadvantaged students, do not apply to or attend a college commensurate with their abilities (Bowen et al., 2005; Pallais and Turner, 2006; Spies, 2001). In the literature, this is typically referred to as “under-matching.” Three, information constraints appear to play a sizable role in under-matching (Hoxby and Avery, 2012; Hoxby and Turner, 2013 and 2015).

Several recent papers indicate that high school students exhibit large changes in behavior in response to small changes in costs and pieces of information, and the strongest responses usually occur among subsets of students lacking sound pipelines to college. Smith, Hurwitz, and Howell (2015), for instance, find that a small increase in

[6] There exists a related literature on the effects of computer and Internet technology on academic achievement (e.g., Faber et al. 2015; Vigdor, Ladd, and Martinez, 2014; Belo et al 2013; Fairlie and Robinson, forthcoming). We view our paper as complementary to those papers since we examine the role of Internet technology in reducing informational and transactional frictions in the college application process, which could also indirectly improve academic achievement if students attend better matched or higher quality colleges.

an application fee or an additional essay heavily influences application behavior. Further, students' application sets are seemingly guided by defaults or perceived "recommendations." Pallais (2015) finds that students apply to more schools when given an additional free Score Send, a cost savings of $6. Cohodes and Goodman (2014) find that students forego large expected earnings for small offers of financial aid from the state. Several papers also find that state mandated admissions exams and proximity to test centers induce large enrollment changes (Bulman, forthcoming; Goodman, 2013; Klasik 2014; Hurwitz et al. 2015).

Such large behavioral responses to defaults, nudges, and small changes in costs suggest students are not well informed of optimal strategies in the application process. Several recent experimental studies have explored this idea by examining whether targeted information provision can improve application behavior. These experiments have directly intervened with the college application process and made tailored recommendations to particular student groups. Examples of interventions include filling out financial aid forms, counseling on the application process generally, and helping students obtain fee waivers that they were already eligible to receive (Avery and Kane, 2004; Bettinger et al., 2012; Carrell and Sacerdote, 2013; Hoxby and Turner, 2013 and 2015).[7] Each intervention has generated large improvements in students' attendance outcomes.

The diffusion of broadband Internet serves as a hands-off experiment that could reduce some of the transactional frictions in the college search. High-speed Internet allows prospective applicants to quickly and easily conduct tailored searches for the information they desire –information that once required conversations with high school counselors, parents and peers, who may be differentially knowledgeable of the college-going landscape and/or the students' needs and interests. Moreover, the time and monetary cost of registering for exams, obtaining study materials, soliciting and submitting applications, and applying for financial aid and scholarships could be reduced. For example, college rankings, prospective college characteristics, admissions requirements, study guides, practice exams, deadlines, costs of attendance, application

[7] Hoxby and Turner's experiment, in particular, produced large enrollment effects from targeted mailings even though their student sample was drawn from the universe of college admissions test-takers, so that both their treatment and control groups were equally college-bound and eligible to receive less-streamlined marketing materials from colleges. Consistent with prior observational studies, a survey of both groups found that untreated students were dramatically under-informed about college quality – particularly at top-ranked liberal arts schools and out-of-state schools but also at state flagships – relative to students who received their intervention.

forms, and information on financial aid and scholarship programs are all available online in some form. In each case, high-speed Internet can make information more accessible and/or reduce the time cost of search and submission. Because we cannot separately identify the impact of each, we remain largely agnostic about how the Internet operates to improve postsecondary outcomes.

Compared with the more targeted interventions described above, the nature of the relationship between Internet access and our outcomes is more complex. Students with Internet access must search for information and distill it largely by themselves. If students do not use the Internet to study for exams and search for information on college application, we would see no effect on student outcomes. If students primarily use the Internet for leisure activities, and the Internet serves as a distraction leading students to substitute time towards leisure, we could see worse outcomes. Ultimately, whether Internet availability, on average, generates positive, negative, or no effects on student outcomes is an empirical question we strive to answer through the paper.[8]

Driven in part by this ambiguity, we also investigate heterogeneity by demographics and geography. We consider two broad hypotheses. First, effects could be concentrated among groups found to under-match or who have been otherwise shown to be sensitive to information nudges. Per the literature, these tend to be students with the fewest inroads to elite colleges. As such, we examine whether minority students or students in low-income neighborhoods, students with less-educated parents, or students from remote places are relatively more affected by Internet access.

Still, the literature on under-match is fairly new, and many of the most relevant studies were conducted after the Internet had become commonplace. This alone suggests that the mere prevalence of the Internet may not have been particularly beneficial to these traditionally- underserved populations, either because of a lack of take-up or differential usage of Internet services. (Certainly, these populations on a

[8] Indeed, surveys offer ambiguous and inconclusive evidence on whether students use the Internet to effectively search for information on college application. A 2005 survey of Internet users found that 45 percent of Internet users, or 30 percent of all adults, had used the Internet to search for information on prospective college or universities for themselves or a family member (PEW 2005). At that time, this was similar to reported usage rates for banking online (44 percent) and looking for information about a job (44 percent), and higher than reported usage rates for reading a blog (27 percent), playing online games (32 percent), looking at online classifieds (36 percent), or using social networking sites (7 percent) (PEW, 2014). A 2012 survey found that three-fourths of teachers think the Internet and digital search tools have a mostly positive impact on students' research habits, suggesting students may be able to conduct effective online searches (Purcell et al., 2012). However, 64 percent of teachers think digital technologies do more to distract students than help them academically (Purcell et al., 2012), suggesting that students may use the Internet for counterproductive activities that detract from study and search activities.

whole are demonstrably less likely to have and use Internet services.) Indeed, within North Carolina, Vigdor, Ladd, and Martinez (2014) recently found that for younger students, Internet access increased the wedge in testing outcomes between high- and low-income students, which they attribute to a digital divide in how productively the Internet is used by the two groups at home. Thus, another hypothesis is that the greatest beneficiaries of the Internet as it rolled out were the students with the most initiative and/or the greatest resources, especially if such students tend to use the Internet for education-related activities more than their peers. This might imply a concentration of effects among groups with *greater* inroads to colleges.

III. Data

The main empirical approach used in this paper is to relate zip code-level broadband Internet availability to individual-level testing and application data for students who took SAT exams. This section describes our main data sources and how we construct our relevant variables.

a. Testing and Application Data

Our primary data source is the College Board (CB), an organization that administers the Preliminary SAT/National Merit Scholarship Qualifying Test® (PSAT) and the SAT to high school students. The PSAT is an assessment taken prior to the SAT that serves as a qualifying exam for a nationally competitive scholarship program. Approximately 3.5 million students take the PSAT each year, typically either in the fall of their sophomore or junior year or both. The SAT is primarily a college admissions exam, and it yields a key metric on admissibility as well as application and demographic information for a majority of college-bound students. Over 1.5 million students in the graduating high school class of 2014 took the SAT, typically as juniors or seniors and often as both. Using the population of PSAT takers for our analysis of SAT outcomes provides us with a pre-“treatment” ability metric, which can be related to test-taking and application outcomes in the SAT sample.

Our sample is composed of test-takers in the graduating high school cohorts of 2001 to 2008, with roughly one million students per cohort.[9] CB data contain self-reported information on high school graduation cohort, student race, gender, cumulative high school GPA, home zip code, parental income and education, and high school attended. The SAT contains math and critical reading sections, each of which is scored on a scale

[9] We use the 2001 to 2008 cohort in order to be able to examine Internet access in both junior and senior year of high school. Internet access is available from December 1999 to December 2007.

between 200 and 800 for a maximum composite score of 1600.[10] The PSAT has a similar format, but the section scoring is between 20 and 80.

Along with exam scores and basic demographics, the CB data also identify colleges to which students send official copies of their SAT scores (Score Sends), which serve as good proxies for actual college applications (Card and Krueger, 2005; Pallais, 2015). When registering for the SAT, the student has the option to send her scores to four colleges for no additional cost. Scores may also be sent at a later date for a fee of approximately $11 per Score Send.[11] Low-income students are eligible for additional free Score Sends.

For every Score Send, we merge characteristics of each college, including data on quality (average SAT score of incoming freshmen), control (public or private), level (two- or four-year), type (e.g., liberal arts, state public flagship), and location (state), which come from Integrated Postsecondary Education Data System (IPEDS). Our final analytic sample consists of the 7,452,302 students who take both the PSAT and SAT and are in the high school graduating cohorts of 2001 to 2008. The sample only includes students with a valid zip code.[12] Table 1 displays the summary statistics on the sample.

The top panel of Table 1 displays scoring and application means of SAT takers. The average combined PSAT score on both sections is just over 99 and, similarly, the average SAT score, which combines the results of these same testing areas, is just over 1000. According to the students' Score Sends, about 79 percent submitted an application to at least one four-year college. In addition, 40 percent applied to at least five colleges, which, importantly, is one more than the default of four free Score Sends.[13] Approximately 50 percent applied to a selective college (average SAT score greater than 1200) and 30 percent applied to a very selective college (average SAT score greater than 1300), though almost 70 percent of students applied to an academically matched college (average SAT score at least as good as their own or 1300, whichever is lower). Zooming in on school type, about 30 percent of test-takers applied

[10] The writing section was introduced in 2005, making the maximum composite score 2400. For consistency across classes, and because colleges typically do not rely on the writing section, we focus only on the math and critical reading sections.
[11] The cost changed slightly over the sample period.
[12] A few students were excluded because all their demographic data were missing or because they live in a zip code with no information on broadband access.
[13] The modal number of Score Sends is four and Pallais (forthcoming) shows that students tend to apply to the number of free Score Sends. Therefore, sending at least five Score Sends is a deliberate act.

to the state flagship, 50 percent to at least one out-of-state college, and less than 10 percent to a top private liberal arts college.

The middle panel of Table 1 displays demographic characteristics of our sample. The sample consists of 67.2 percent who identify as white, 10.7 percent as black, and 9.5 percent as Latino/Hispanic. About 45 percent of the sample is male. High school GPA is a categorical variable where 0 is a non-response, 1 is the lowest, and 12 is the highest. We use the categorical variable in the analyses but present the average of the continuous version (equal to 3.363) in the table.

Finally, we add zip code-level economic characteristics to our data in order to control for changes in local economic conditions (Table 1, bottom panel). Mean adjusted gross income is $77,444, which is measured at the zip code-level and was obtained from the IRS Statistics of Income (SOI) data.[14] Population and housing data at the zip code-level were obtained from the 2000 Census and made time-variant by merging with zip code-level trends in SOI counts of filers and households. We capture local labor market trends by including information on county-level unemployment rates, collected from the Bureau of Labor Statistics. We also include trends in county-level house prices, obtained from the FHFA house price index and the 2000 Census.[15]

b. High-Speed Internet Access Data

Our goal is to estimate how broadband Internet affects student testing and application outcomes. Since there is no measure of a student's ability to use or access broadband in the CB data, we construct a measure of broadband availability in a student's zip code.[16] To do so, we combine information on zip code-level ISP coverage with national trends in aggregate demand for broadband services to produce a binary measure of broadband availability such that any household within a zip code where broadband is available can opt to have it. Because individual households have very little control over whether and when providers enter their zip code, and very little

[14] We interpolate missing years in this data.

[15] We construct a county-level measure of house prices by combining information on county-level median home prices from the 2000 Census with the Federal Housing Finance Agency house price index, as in Dettling and Kearney (2014). Urban counties use the Metropolitan Statistical Area (MSA) version of the index and rural counties use the rural index.

[16] We might alternatively attempt to derive variation from household or student usage rates. Unfortunately, usage rates cannot be constructed for all years at the subnational level. The PEW data are available frequently but are only available at the national level. The CPS data include state identifiers but are only available in 2000, 2001, 2003, 2007, and 2009. Moreover, a measure of access is preferable to usage rates, which capture a student's ability to access online content but are also endogenous to our question of interest if, for example, parents take up high-speed Internet in order to improve their children's educational outcomes.

impact on aggregate demand, the measure we derive is interpretable as an exogenous shock to a student's ability to use high-speed Internet.[17] In this section, we describe the data and construction of our measure in more detail.

We derive our measure of broadband ISP coverage from FCC Form 477 Filing data.[18] The FCC requires every facilities-based provider with at least 250 high-speed lines to report basic information about its service offerings and end users twice a year.[19] The FCC releases summary statistics to the public aggregated to the zip code-level, namely a list of zip codes with the number of ISPs who have at least one subscriber within the zip code receiving speeds of 200 kbps or more.[20] The data are available bi-annually from December 1999 to June 2008, and to protect confidentiality, do not distinguish between one, two, or three providers in a zip code. Over that time, there is considerable variation both across and within zip codes.

Ideally, we would like to operationalize the data on the number of ISPs in a zip code to be able to compare, across zip codes and time, the average resident's ability to access high-speed Internet in her home. While we do not have a direct measure of zip code-level access from which to derive a correspondence between the number of ISPs in a zip code and accessibility, we can compare nationally aggregated coverage rates implied by the FCC data to survey-reported national usage rates of high-speed Internet. This is a reasonable litmus test for how well the raw FCC data capture market penetration – i.e., how provider entry translates into usage – at least at the national level. Figure 1 compares trends in the fraction of the population residing in a zip code with at least one provider to national trends in survey-reported usage, which were obtained from PEW Research.[21] There are large discrepancies between the two series, which do not match one another well in either levels or trends.[22] Figure 1 also indicates

[17] We discuss this issue at length in the analysis section.
[18] The FCC data can be downloaded from https://www.fcc.gov/encyclopedia/form-477-data-zip-codes-number-high-speed-service-providers.
[19] Small providers, many of whom serve sparsely populated areas, are not required to report to the FCC and sometimes provide information on a voluntary basis. In our analysis we will provide separate treatment for rural and urban areas, in part to address concerns of measurement error arising by differential reporting standards across the two densities.
[20] A customer, per the filing requirements, can include residential or small business customers.
[21] The PEW data can be found at: www.pewinternet.org/data-trend/internet-use/connection-type/. Note that the these rates of usage are extremely similar to those found in the Current Population Survey, which asked respondents about broadband Internet usage in 2000, 2001, 2003, 2007 and 2009.
[22] We are not the first to note that the raw data on provider presence may not be able to accurately capture local access. A 2005 study by the GAO concludes that defining access according to provider presence alone "…may overstate deployment in the sense that it can be taken to imply there is deployment throughout the zip code even when deployment is very localized." The paper also provides the results of a

that using the next available cutoff – four or more providers – does little to improve the divergence in levels or trends.

The inability of the raw FCC data to capture national trends in usage is not surprising once one considers the vast heterogeneity in geographic and population sizes across the roughly 32,000 zip codes in the United States. Consider, for example, two zip codes which each reported one to three providers in 2000: 82332 is a rural zip code in Savery, WY with 134 residents in 1,422 square miles, and 10030 is an urban zip code in New York, NY with 25,847 residents in 0.30 square mile. By 2008, 82332 had 4 providers and 10030 had 11. Thus, it seems unlikely that all residents of each of these zip codes has equal access to broadband, suggesting a "one size fits all" measure of zip code-level coverage is inappropriate.[23] Instead, we propose a measure that scales the number of providers in a zip code by its size. Further, because the literature on broadband roll-out suggests that supply-side constraints limiting roll-out were structurally different in urban and rural areas, we allow for how we define "size" to vary across these concepts.[24] In particular, we scale the number of providers by square mile in rural areas and by population in urban areas.[25] In the appendix, we present alternative specifications using only the population-scaled measure and only the mileage-scaled measure across all zip code types (as well as separately for urban and rural zip codes).

---

case study in Kentucky, where it was found that 95 percent of residents had a provider in their zip code, but only 70 percent had access in their area.

[23] Prior research using the FCC data attempts to circumvent the comparability issues across zip codes by drawing variation in ISP coverage from fairly homogeneous zip codes, either by investigating outcomes within a single state--where zip codes are quite similarly structurally – or by removing high- and low-density zip codes (Vigdor, Ladd, and Martinez, 2014; Xiao and Orazem, 2010). Unfortunately, such cuts are less desirable in our setting, because much of the foundational work in higher education suggests that these are precisely the sorts of comparisons we wish to make, as it is at extreme population densities where there is the most substantial divergence in both information and application behavior.

[24] In rural areas, where zip codes are much larger and less densely populated, coverage was constrained by the cost of extending additional lines long distances to reach relatively few customers (GAO, 2005). In urban areas, population density can be a problem because too many customers using a single line at once can exhaust the system (Faulhaber, 2002; Greenstein and Price, 2007; Grubesic and Murray, 2002). Thus, to be most flexible in our definition of coverage, we allowed for the possibility that geographic size may be most relevant for rural zip codes and population size may be relevant for urban zip codes. Note that in our empirical match (described in the appendix) we allowed for a single definition but found that the best match was to include different definitions for rural and urban consumers. In the appendix we provide results using a single definition for both urban and rural areas.

[25] Zip code population data are based on a combination of Census 2000 population data and SOI population data. SOI data provides a count of the number of income tax filers, which was used to create population trends to move the 2000 Census zip-code population total forward. Zip-code land area estimates are from the 2000 Census.

The results become similar to our main estimates the better the measure fits usage trends.[26]

To operationalize the concept of Internet availability and facilitate interpretation, we convert the scaled measure into an indicator variable that takes a value of one when it crosses a specific threshold.[27] Since there is no theoretical guidance for what an appropriate threshold might be, we identify the threshold empirically by targeting the national trends in survey-reported high-speed Internet usage displayed in figure 1. We construct an algorithm to test varying thresholds and ultimately find that the best-fit measure defines penetration in a rural zip code as "at least one provider per 12 square miles," and in an urban zip code as "at least one provider for every 2,700 people."[28] More details on the construction of this measure can be found in the data appendix.

The red line in figure 1 displays bi-annual trends in our measure of broadband Internet access rates based on our measure of Internet penetration. We see that, unlike the providers-only-based measures, our measure closely follows levels and trends in the PEW survey reported usage rates. In 2000, both measures are close to zero, and by 2008, both are around 60 percent. Figure 2 displays maps of our measure of zip code-level high-speed Internet penetration over time, where dark gray zip codes are those that our measure indicates have Internet access available and light gray zip codes are those

[26] The root mean squared error between our main measure of access and national trends in usage is 0.92. Using just the urban population measure for the whole sample, the root mean squared error is 1.34 and the results are quite similar to our main measure. Using just the rural mileage measure for the whole sample, the root mean square error is much larger at 2.79. In this case, the results begin to differ from our main results. We attribute this to the relatively poor fit of this measure, as a different mileage measure which better fits overall usage patterns leads to more similar results to those obtained from our main measure. We also examined the sensitivity of our results to models using discrete bins characterizing the number of providers (one to three and four or more), with the caveat that we believe this substantially overstates national broadband availability and fits the data poorly (as evidenced by figure 1). The measure, though it is difficult to interpret, offers inconsistent testing results (additional test-taking drawing in equally able students but then lower SAT scores overall) and somewhat inconsistent application results (more applications to selective schools; fewer applications). The measure performs better in both respects when we investigate North Carolina alone, as in Vigdor, Ladd, and Martinez (2014), likely because zip codes within a single state are at least conceptually similar. Results are available upon request.

[27] We do so because we wish to interpret our results as the effect of Internet access on student outcomes. Interpretation of the linear measure would be difficult, as the literature provides no guidance on what it means for a zip code to have one additional provider per person or square mile. Additionally, the provider measure is not continuous and is listed in bins; thus, a continuous measure of coverage cannot be used and interpreted as such. While we prefer a dichotomous measure, we note the caveat that we can only interpret our measure as the average resident's ability to access high-speed Internet. It is possible that, for some parts of a zip code, Internet will be available prior to a zip code being "turned on" by our measure, and for others, Internet may never be available.

[28] We define "best fit" as minimizing the root mean squared error between each measure and the survey-reported trends. More details can be found in the appendix.

that do not.[29] It is clear there is both across-time and across-zip code variation in high-speed Internet coverage. In December 1999, few zip codes have access, and those that do are clustered in the major population centers on the East Coast. By December 2007, coverage has expanded across the United States, although there are still many areas that coverage has not reached. Over the full sample period, around 30 percent of our SAT takers are considered to have high-speed Internet available in their area in their junior year (table 1).

Finally, we note the caveat that our measure is designed to capture Internet "availability" or "access" to the average resident in a zip code, but an individual student's ability to use the Internet is much more nuanced, whereby household take-up can play either a mediating or amplifying role. High-speed Internet subscriptions are not free, and there are large observable differences in high-speed Internet take-up by education and income. These gaps remain to the present day when local Internet access is nearly universal, suggesting these differences are due to differential take-up rates beyond the question of availability.[30] Moreover, even if a home has a high-speed Internet subscription, differences in purchased speeds and the number of devices available to a student can lead to differences in a student's ability to use online services.[31]

IV. Analysis

In this section, we investigate whether Internet access affects a student's college admissibility and application set, and, to some extent, how the two interact. To separate the two, we need to understand the evolution of a student's application and distinguish behavioral changes in how she targets her application from structural changes in the quality of her application. Thus, we begin by examining whether our measure of Internet access coincides with shifts in the quantity and quality of SAT test-takers drawn from that population. Then, in our core analysis, drawing from existing research on college quality and under-match, we consider application outcomes designed to capture behavioral patterns deviating from defaults or broad recommendations. We then

[29] The white areas on the map are zip codes for which we do not have information. Most represent unpopulated areas like national parks and bodies of water.

[30] The 2009 October CPS indicates that 58 percent of 15-18 year olds with a mother who did not complete high school had broadband at home compared with 95 percent of 15-18 years with a mother with a post-graduate degree.

[31] The 2003 October CPS indicates that 55 percent of 15-18 year olds with a mother who did not complete high school had a computer at home compared with 97 percent of 15-18 years with a mother with a post-graduate degree.

examine the validity of the assumptions underlying our interpretation of our estimates as causal. Finally, we extend our analyses in two ways. First, we wed our testing and application results to examine the extent to which observable differences in applicant quality induced by Internet access can explain the application effects we detect. In addition, we investigate whether our results appear to be concentrated within particular demographic and socioeconomic groups.

a. Empirical Specification

Our estimating equation is a generalized difference-in-differences:

$$y_{izc}=\alpha+\beta*broadband_{izc}+\mathbf{A_i}\boldsymbol{\theta}+\mathbf{X_{zc}}\boldsymbol{\lambda}+\gamma_z+\gamma_c+\varepsilon_{izc}$$

where y_{izc} is a binary variable[32] capturing a testing or application outcome for student i from zip code z in cohort c; $broadband_{izc}$ is an indicator for broadband availability in her junior year; $\mathbf{A_i}$ is a set of student-specific demographic and ability controls; $\mathbf{X_{zc}}$ is a set of cohort-varying zip code economic conditions coincident with the timing of the broadband availability measure (i.e., also in her junior year);[33] and γ_z and γ_c are zip code and cohort effects. Note that $\mathbf{A_i}$ includes the student's PSAT verbal and math scores as well as high school GPA, which capture "latent ability" as well as an early signal of admissibility to selective colleges.

Our primary coefficient of interest is β such that when y is "applied to a four-year college," our estimate represents the increase in the likelihood a student applies to a four-year college if Internet is available in her home zip code. This characterization of β holds under the assumption that, all else equal, trends in testing and application outcomes in zip codes with and without Internet would have evolved similarly over time, save for the availability of high-speed Internet. Of course, zip codes which received Internet access earlier may have been different than zip codes which received Internet access later. That is why it is imperative that our specification include zip code fixed effects (γ_z) so that our estimates are net of any time-invariant differences across zip codes. $\mathbf{X_{zc}}$ also contains a set of zip-code-level economic indicators, including local income, unemployment rates, house prices, and population density in an effort to

[32] An exception is when we estimate the effect of Internet availability on PSAT and SAT scores, which is an integer-based variable, ranging in increments of 10 from 400 to 1600.

[33] Student-specific controls include dummies for gender, race, and high school GPA, as well as PSAT math and verbal scores; zip code controls include adjusted gross income, population, number of houses, unemployment rate, and median home price. Additionally, in Appendix Table 4 we add a control for student survey responses to parental education and income, questions that do not appear on the PSAT survey and that are associated with considerable non-response on the SAT survey. Results are qualitatively unchanged to their inclusion.

capture any observable changes in zip-code characteristics that may have been correlated with the availability of high-speed Internet.

Our identifying assumption is that, all else equal, our measure of high-speed Internet availability is exogenous to a student's testing and application outcomes. Recall that we derive our measure from zip-code-level access based on counts of ISPs benchmarked to trends in usage. We might be concerned that usage is endogenous to our outcomes, and, at the individual level, it almost certainly is. However, recall that our specification includes cohort effects (γ_c), so that our estimated effects are net of any national trends in Internet adoption and the availability of online content. Thus, for identification purposes, we need only assume that provider entry is exogenous to student's outcomes. This means that threats to identification come in the form of either student-demand-related pressures on Internet service providers to enter their zip code, or any omitted variables that co-vary with student outcomes and provider entry. We view such threats to be small, if not negligible.

Per the former, there is abundant evidence that supply-side constraints restricted high-speed Internet access, and that the supply of high-speed Internet lagged consumer demand. To provide high-speed Internet, Internet service providers (ISPs) – typically the existing phone or cable company – had to make substantial infrastructure investments, retrofitting existing phone and cable lines and installing new switches and servers (Faulhaber, 2002; Greenstein and Prince, 2007; Grubesic and Murray, 2002). There is a general consensus that the costs slowed rollout and access did not keep up with consumer demand (Greenstein and Prince, 2007; Faulhaber, 2002). Dettling (forthcoming) discusses how variation in the underlying housing infrastructure and the availability and quality of existing telephone and cable wiring made these infrastructure investments differentially costly across locations, which created differences in the timing of the availability of high-speed Internet services across locations that was unrelated to consumer demand-related pressures.[34] Based on these known barriers to

[34] There are two main transmission modes for high-speed Internet in the United States: cable-based and telephone-based digital subscriber line (DSL) service. Each service requires the installation of fiber-optic wiring, which provides high-speed Internet service up to a certain point, from which the signal travels over traditional coaxial cable or copper telephone wiring the rest of the way. These fiber-optic lines may reach the ISPs' central office, some remote terminal in the neighborhood, or the home. The main issue that prevented timely rollout for the cable companies was capacity. Cable companies had installed some fiber lines in the 1980s to provide digital cable service, but each additional customer on a single fiber line reduces the "downstream" capacity, meaning that multiple simultaneous users reduces speeds and could exhaust the system. Thus, to provide reliable, high-speed Internet service, cable companies needed to add more fiber lines that came closer to residences. For

entry, it seems unlikely, but not impossible, that student-demand-related pressures induced Internet service provision. Thus, net of state and year fixed effects, and the economic controls included in our model, the roll-out of high-speed Internet is arguably random.

For the latter to be plausible, it must be that our extensive set of student- and zip-code-level controls are inadequate. We investigate the exclusion of key variables by re-estimating our main outcomes over other timings of Internet availability that are less likely to affect our outcomes. For example, as a preview of our results, while we find that junior year Internet availability leads to SAT score improvements, when we instead use a measure of Internet availability in December of a student's senior year of high school – after most students would take the SAT – there is no discernible effect on SAT scores. We view this as suggestive evidence that our estimation framework does not omit key variables, though we acknowledge it is a possibility.

Finally, before we turn to results, we make two notes regarding their interpretation. First, the coefficient on $broadband_{izc}$ captures the effect of *potential* access, as opposed to Internet use, because $broadband_{izc}$ is a measure of Internet penetration over a geography rather than student-specific Internet adoption. Arguably, this coefficient is the most relevant for policy. As noted earlier, the effect of Internet use will depend on how many students use it, how often they use it, and in what ways – all of which the government is unable to fully control or measure. Moreover, local peer effects, wherein local access changes the college-going culture of students' neighborhood or high school even for those students who do not personally have access at home, could amplify or reduce the effectiveness of use in ways that are difficult to separate from own access but that a measure of local access will generally subsume.

Second, because our analysis focuses on how the prevalence of Internet in junior year coincides with late high school outcomes and includes both zip code and year fixed effects, our estimates will reflect only the testing and application effects of coverage

---

DSL Internet from the phone companies, rollout was prevented by the need to upgrade the existing telephone wiring, much of which was old and had been split too often to be capable of carrying high-speed two-way traffic. In either case, the key insight is that existing wiring leading up to the home or apartment building was insufficient to carry highs-peed traffic, while the wiring already in a home or building was typically sufficient. Dettling (forthcoming) demonstrates that this incentivized entry into markets where the existing housing infrastructure offered lowered costs of provision. That is, areas with more multiple-family dwellings received access earlier because ISPs could take advantage of economies of scale in the provision by bringing one line to multiple consumers.

that are immediately detectable. In practice, a nudge based on new information could affect outcomes at any pre-collegiate stage. Early high school students, for instance, could alter their coursework and career paths based on information gleaned from the Internet, which, in turn, could also affect postsecondary outcomes. We leave this possibility to future work.

b. Main Results: Test-Taking, Scoring, and Application Outcomes

We begin our analysis by examining whether SAT test-taking rates systematically vary with Internet availability. Relative to the full population of U.S. students, SAT takers tend to be college-aspiring students, so our sample of testing and application outcomes are likely positively selected from the distribution of student ability. Moreover, there exists a competing exam for college-bound students, the ACT, that students can elect to take instead. Thus, depending on how the Internet is used and by whom, we might see shifts in SAT test-taking and SAT test-taker quality that result from increased broadband availability.

Note that while test-taking is a separately interesting outcome that may be affected by Internet availability, any systematic differences that we detect in the amount or quality of SAT takers would also affect our interpretation of β when the dependent variable is an SAT score or a college application decision. This is because we can only observe these outcomes for students who take the SAT, and such results would imply that we do not observe a comparable set of SAT test-takers across zip codes with and without Internet. Consider, for example, if, when Internet is available, more students find it worthwhile to take the SAT, but those who elect to take the exam only in this state tend to earn below-average scores. Were we to then observe less-favorable average application outcomes in zip codes where broadband=1, we might incorrectly infer that the Internet negatively affects application outcomes, when instead the composition of students for whom we can observe outcomes has also changed. Thus, if there is observable selection into (or out of) SAT-taking when Internet is available, estimated βs for outcomes that can only be observed conditional on taking the SAT would reflect a combination of true application effects on students who take the exam in any state and broadband-driven sample selection.

To estimate shifts in test-taking and test-taker quality, we leverage information available from the PSAT. The PSAT is the qualifying test for the National Merit Scholarship Program and is often mandatory within a state or district, so that the population of PSAT takers approximates the at-risk (i.e. non-selected) population of SAT takers. Using the full set of PSAT takers, we first examine whether more students elect to take the SAT once broadband is available, so that y_{izc} is an indicator value that takes the value of 1 if a student takes the SAT exam and 0 otherwise. Table 2, column 1 indicates that the estimated effect of high-speed Internet on SAT-taking is extremely small (0.00049) and statistically indistinguishable from zero. Thus, on average, the likelihood a student takes the SAT is unchanged by increased Internet prevalence.

Still, there remains the possibility that there are changes in test-taker quality underlying this non-effect if, for instance, more-able students are induced to take the exam and directly offset less-able students who are discouraged from taking the exam. Thus Column 2 considers whether the PSAT score for our analysis sample – i.e., students who take both the PSAT and the SAT – varies systematically with Internet availability. Again, we obtain results indistinguishable from zero. From this combination of results, we can make two related inferences. First, the availability of high-speed Internet in junior year apparently does not induce students to alter their SAT test-taking decisions. And second, the set of test-takers for whom we observe applications and SAT scores does not meaningfully differ in ways that would raise sample selection issues in examining these outcomes.

Next, we examine whether broadband availability affects outcomes later in the application cycle. We begin by looking to see whether we can detect systematic differences in SAT scores. Since, as we have demonstrated, the underlying test-taker populations are equally able, any observable changes would indicate that test-takers become differentially admissible to selective colleges when broadband is available. Indeed, column 3 shows that students in zip codes with broadband outperform their academic peers in zip codes without broadband by an average of 0.7 SAT point (or 0.3 percent of a standard deviation).

Finally, we examine the effect of broadband on application rates. Drawing from existing research on college quality and under-match, we consider outcomes designed to capture behavioral patterns deviating from defaults or broad recommendations, such as applying to more schools than the number allotted with a test registration (i.e. applying to five or more), applying to a four-year school, applying to a selective school, applying to a match school, applying to a state flagship, applying out of state, and applying to a highly ranked liberal arts school. We construct each application measure as an indicator variable and we test, in line with the literature, whether the likelihood a student will deviate from perceived defaults increases with more information.

The remainder of Table 2 displays estimates of the effect of broadband on the application measures. The signs on all of the coefficients are consistent with improved outcomes: we estimate statistically significant gains in the probability a student applies to more than the default number of colleges, a four-year college, a selective college, a college commensurate with her own score (i.e., a match college), a top liberal arts college, and an out-of-state college ranging from 0.2 to 0.4 percentage point. In some instances, these changes reflect quite meaningful deviations from mean behavior; for example, only about 7.2 percent of SAT takers apply to top liberal arts colleges, and Internet availability induces a 0.17 percentage point change over that baseline.

Interestingly, while both are positive in sign, the effect of broadband Internet on whether students apply to a very selective college or an in-state flagship are statistically indistinguishable from zero. The first null result could reflect disparities in student groups to whom the gains of Internet availability tend to accrue – perhaps the most elite students are not the largest beneficiaries. The second null result is consistent with this line of thinking as well, but also with prior work that has found that groups of students who under-match generally tend to favor public, in-state colleges in their applications (Hoxby and Avery, 2013), so we might not anticipate large effects within this class of institutions. Moreover, flagship colleges tend to be large and well-known; thus, if simple awareness was preventing applications, we might not expect this outcome to be affected.[35]

c. Validity of Research Design

Our identifying assumption is that barring the emergence of broadband, trends in our outcomes would have evolved similarly. One way to check this assumption is to examine trends in our main outcomes before broadband became available to residential customers in 1999. Using SAT data for the 1996 to 1998 graduation cohorts, we classify zip codes according to whether Internet became available during the early (1999–2001), middle (2002–2005), or late (2005–2007) years of our sample period, and chart the evolution of our main outcomes over this prior period. To conserve space, we limit this

---

[35] While Hoxby and Avery (2013) find that, within the state, these groups do not seem to favor flagships and that they often instead apply to less selective public universities in their state, in a related survey, Hoxby and Turner (2015) find that the reasons such students give for not applying to the state flagship seem more related to the social environment at the school than to unawareness of its academic quality.

discussion to only outcomes that were statistically significant in Table 2 and present the results for the remaining outcomes in the appendix.

Figure 3 displays the results of this analysis, which indicate that there are substantial and persistent level differences by timing in several of our outcomes before broadband technology became commonplace: test-takers from zip codes that received broadband relatively early tend to have better application outcomes. This result is not surprising, since ISPs tended to enter wealthier zip codes earlier than less well-off zip codes. Recall, however, that our specification includes zip code effects. This means we need only assume that trends – not levels – are statistically similar across the timings we consider in the absence of broadband. Figure 3 indicates that this assumption appears to be valid for the 1996–1998 period: there do not appear to be any differential trends in our outcomes by broadband availability category. Thus, prior trends do not appear to be driving our estimates.

Another way to examine the validity of our design is to estimate outcomes for students who were older in the application cycle when broadband became available. Specifically, we would not expect to see systematic differences in applications among students who were already freshmen in college by the time broadband became available to their home zip code. Thus, we re-estimate our model, adding new indicators for broadband availability in a student's home zip code during her senior year of high school and freshman year of college. Although obtaining broadband access during one's freshman year of college should not theoretically affect application outcomes, in practice we have an imperfect measure of individual-level access and it is possible that some students have access to high-speed Internet before our measure officially turns on in their zip code. Thus, we do not necessarily expect to see precise zeroes for freshman year of college, but rather broadly more muted effects for cohorts who generally would have been too old to benefit.

Figure 4 summarizes the results of our analyses of zip code broadband access in junior year of high school, senior year of high school, and freshman year of college on our main outcomes. Again, for brevity we present the outcomes that were statistically significant in Table 2. In these figures, the x-axis plots the year of schooling and the y-axis displays the estimated coefficients and 99 percent confidence intervals on the indicator for zip code-level broadband access in the year listed. In each case, broadband access in junior year has a positive and significant effect on the outcome, and an imprecisely estimated null effect in the freshman year of college.

Finally, we have noted that in our data we can only observe testing and application outcomes for SAT takers, even though there exists a competing admissions exam students can elect to take instead. Thus, we might be concerned that some of the behavior we detect reflects selection into (or out of) our analysis sample, driven by switching between the ACT and SAT. While colleges throughout the United States generally will accept either exam, the tests historically tend to prevail in particular geographic regions; students on the coasts and in the South more often take the SAT and students in the Midwest favor the ACT. In addition, students planning to apply to highly selective universities often take the SAT. Over the time period we consider, as accountability pressures grew, some states, and often those where the ACT had already prevailed, began requiring their students to take an admissions exam. Since the resulting score could be used in the application process, state mandates potentially interfered with some of the competitive dynamics between the two exams.

To examine whether the existence of the ACT is introducing selection concerns into our analysis, we restrict our analysis sample to those states in which the SAT historically has prevailed, and re-estimate all of the outcomes from Table 2 in Table 3. The results broadly mirror those in Table 2. There is again no evidence of sample selection, either by differences in test-taking rates or by PSAT score. The SAT score increase owing to availability is again 0.7 SAT points, and application outcomes increase across the board similar to our main estimates.

d. Exploring Mechanisms

We showed earlier that application outcomes improved for students with broadband Internet. However, we also detect broadband-driven increases in their average admissibility, as measured by their scores. We have not yet shown whether application improvements occur independent of scoring improvements. In other words, to understand how broadband availability is operating to improve outcomes, we would also like to know whether students who are observably equally admissible apply differently when broadband is available.

To probe this, we re-estimate our application outcomes, adding a separate control for the SAT score in addition to the PSAT score. Table 4 displays the results of this exercise and indicates that broadband has an effect on student applications, above and beyond improved admissibility. Dividing the estimates in Table 4 by those in Table 2 indicates that score improvement explains at most 20 percent of changes in the application set (as indicated by the italicized percent changes). While there may be some complementarity between changes in SAT scores and other features of a student's application that are unobservable to us – for example, a student with a higher score might feel encouraged to enroll in more rigorous courses or participate in additional extra-curricular activities – these results indicate that broadband availability is likely also acting through channels independent of student admissibility, such as reducing the work involved in submissions or making information on schools easier to obtain.

We can further examine this separability by considering whether a more short-term treatment – broadband availability in December of senior year of high school – affects our outcomes. In this case, students would have much less time to improve the strength of their application. Table 5 displays the effects of senior year access on application outcomes. Similar to the results for junior year, we find broad improvements in outcomes, although the results are muted.[36] Moreover, while still positive, we no longer detect significant effects on the probability a student applies to a selective school or out-of-state school. Importantly, we do not estimate an SAT score change consistent with the later timing. Since students do not typically take the SAT after fall of senior year, we find this (non-)result to be a compelling affirmation of our strategy and indication that broadband availability is acting beyond score increases to improve application outcomes. Overall, Table 5 suggests that admissibility alone is not driving our results.

e. Heterogeneous Effects

The literature on under-match identifies the existence of substantially many high-achieving students left behind because of a lack of peers, role models, and necessary information that could help them with the application process. Broadly, findings in this area tend to connect measures of relative disadvantage—e.g., lower-income geographies, less-educated parents, students in rural areas, and students who identify as a racial minority—to information constraints limiting students' postsecondary opportunities. In this regard, it seems plausible that high-speed Internet access might have a larger impact on postsecondary outcomes for students from these groups.

[36] Because we focus on the same cohorts of students and allow the years in which Internet access was obtained in the zip code to vary, the magnitude of the coefficients are not directly comparable to the junior year results. For instance, the first cohort we use graduates in 2001. That cohort could have obtained Internet access in December of their junior year, or 1999. In the senior year analysis, that cohort could have obtained access in December of 2000.

However, our setting offers more than just pure information, such that there may also be unobserved countervailing forces, such as whether the student's family can afford to or wants to purchase a subscription, whether the student has devices at their disposal, and how effectively the Internet is being used. If those differences are correlated with relative disadvantage – which most evidence suggests that they are – Internet access may not translate into improved academic outcomes according to the groups we consider.[37] This would be consistent with recent work which suggests students from lower socioeconomic backgrounds (i.e., lower-SES students) with access to Internet suffer academically relative to their peers, while students from better socioeconomic backgrounds (i.e., higher-SES students) gain (Vigdor, Ladd and Martinez, 2014).

Thus, our next exercise is designed to probe whether we observe systematic differences in outcomes by particular student groups, with the caveat that the analyses that follow are somewhat difficult to interpret because the dimensions we consider are also correlated with lower rates of household Internet take-up and device access. As before, our βs are interpretable as the effects of potential access to Internet; if we observe no effect on the students who are most constrained, we are unable to test whether this is due to lack of take-up or lack of a treatment effect.

Table 6 displays estimates of the effect of broadband Internet by group. We find that improvements in applications are concentrated among higher-income, more-educated, white, and urban students. Interestingly, our estimates for lower-SES students appear to indicate that high-speed Internet induces considerable changes in the test-taking population, as evidenced by the reduction in average PSAT scores, and likely the characteristics of our applicant sample. While interesting from a policy perspective, this complicates interpretation of the SAT scoring and application results, and we cannot be sure what is driving the large changes in SAT scores or the null effect on applications for these groups. Still, in the other lower-resource populations, we observe no such evidence of selection but pervasive null effects in applications. Thus, these results may

[37] There is evidence that disadvantaged students are less likely to have broadband at home, have a computer at home and use the Internet for education-related activities: for example, the October 2003 CPS indicates that among 15-18 year olds whose mother did not complete high school, 8 percent had access to broadband at home and 55 percent had access to a computer at home. Among 15-18 years olds whose mother has a post-graduate degree, 46 percent had access to broadband at home and 97 percent had access to a computer at home. Among students with access to Internet at home, 74 percent of 15-18 year old Internet users whose mothers did not finish high-school use the Internet for education-related purposes, compared to 93 percent of Internet users whose mothers had a post-graduate degree.

reflect how well such populations are able to use the Internet to improve outcomes on their own.

Altogether our findings suggest that the benefits of high speed Internet may be lopsided and favor those who already had more resources. Still, our set of results may be somewhat specific to our setting, when high-speed Internet was first becoming available to households. It may be that in its early stages, the benefits of high-speed Internet primarily accrued to resource-rich students, but as it became more diffuse, lower-resource students were able to experience large benefits themselves. While outside the scope of this paper, future research could explore how policy could mitigate the gaps we detect and whether lower-resource students experienced gains over a longer horizon.

V. Conclusion

This paper examines the effect of high-speed Internet on testing and college application decisions. We find that the availability of high-speed Internet in a student's zip code unambiguously improves her scores and her application set. Students appear to diverge from defaults and broad recommendations by sending more applications, applying to more-selective schools, applying to schools outside their state or smaller in presence, and applying to schools more commensurate with their abilities. Some of this improvement can be traced to increases in average test scores, also owing to broadband. Moreover, consistent with prior literature, we find that interventions fairly late in adolescence can have considerable effects on students' postsecondary outcomes.

We find that the primary beneficiaries of Internet availability appear to be higher-resource student groups, so that the digital divide may be substantially neglecting students who already tend to have fewer inroads to elite academic institutions. If this gap is due to differences in student's ability to access and use broadband – perhaps because their parents do not take up broadband or because students do not have access to devices – these results suggest that policies aimed at increasing broadband availability and affordability could reduce inequality in postsecondary outcomes. If these gaps are due to differences in the ability to find and digest the relevant online content, policies aimed at providing guidance on how the Internet can enhance opportunities for these students might be effective. Nonetheless, our results imply that students can benefit from content available online to improve their outcomes. And, even though our results cannot speak directly to Internet usage at school, it is possible that in-school programs that encourage and monitor Internet searches could be an effective tool for improving college outcomes. We leave it to future work to uncover exactly which margins are relevant for policy to mitigate these gaps.

Of course, an important lingering question is how the application improvements we estimate translate into differences in attendance. We cannot observe enrollment outcomes for our analysis sample, but we can use comparable estimates from the literature to derive anticipated effects. Hoxby and Turner (2013) launched a national information intervention designed to improve application and attendance outcomes. Students who were sent material were 12 percentage points more likely to apply to "peer" institutions and then 5.3 percentage points more likely to enroll in them. In other words, in their setting, about 43 percent of the detected change in application behavior effectively translated into improved attendance outcomes. If we apply those estimates to our "match" application coefficient, we can expect a 0.12 percentage point increase in the likelihood that college-bound students enroll at schools commensurate with their ability. Assuming no supply-side constraints or general equilibrium effects, this decreases the extent of under-match by about 3,500 students per year.[38]

[38] Statistic derived from the Digest of Education Statistics 2013 count of first-time freshmen.

References

Avery, C. and T. Kane. (2004). Student Perceptions of College Opportunities: The Boston COACH Program. In C. Hoxby, ed. College Choices: The Economics of Where to Go, When to Go, and How to Pay for It. University of Chicago Press.

Bailey M. and S. Dynarski (2011, December). Gains and Gaps: Changing Inequality in U.S. College Entry and Completion. Working Paper 17633, NBER.

Belo, R., P. Ferreira, and R. Telang (2013). "Broadband in School: Impact on Student Performance," *Management Science*. 60(2): 265-282.

Bettinger, E. P., Long, B. T., Oreopoulos, P., & Sanbonmatsu, L. (2012). "The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment," *Quarterly Journal of Economics*, 127, 1205–1242.

Bhuller, M., T. Havnes, E. Leuven and M. Mogstad (2011) "Broadband Internet: An Information Superhighway to Sex Crime?" *IZA Discussion Papers Series* No. 5675

Bulman, G. (forthcoming). "The Effect of Access to College Assessments on Enrollment and Attainment." *American Economic Journal: Applied Economics*.

Card, D (1995). "Using Geographic Variation in College Proximy to Estimate the Return to Schooling," in *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp*, ed. by Louis N. Christofides, E. Kenneth Grant, and Robert Swidinsky. Toronto: University of Toronto Press, 201-222.

Carrell, S. E. and B. Sacerdote (2013). "Late Interventions Matter Too: The Case of College Coaching in New Hampshire," NBER Working Paper, 19031.

Cohodes, S. and J. Goodman (2014). "Merit Aid, College Quality and College Completion: Massachusetts' Adams Scholarship as an In-Kind Subsidy," *American Economic Journal: Applied Economics*.

Dale, S. and A. Krueger (2002). Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables. *Quarterly Journal of Economics* 117(4), 1491–1527.

Dale, S. and A. Krueger (2011, June). Estimating the Return to College Selectivity over the Career Using Administrative Earnings Data. Working Paper 17159, NBER.

Dettling, L. (forthcoming) "Broadband in the labor market: The Impact of High-Speed Internet on Married Women's Labor Force Participation" *Industrial and Labor Relations Review*

Dillon, E. and J. Smith (2013). "The Determinants of Mismatch Between Students and Colleges," NBER Working Paper.

Fairlie, R. and J. Robinson (forthcoming). "Experimental Evidence on the Effects of Home Computers on Academic Achievement among Schoolchildren," *American Economic Journal: Applied Economics.*

Falck, O., R. Gold, and S. Heblich (2014) "E-lections: Voting Behavior and the Internet" *American Economic Review* 104(1) pp.329-389

Faulhaber, G. R. (2002). "Broadband deployment: Is policy in the way?" In R. H. Crandall and J. H. Alleman (Eds.), Broadband: Should We Regulate High-Speed Internet Access, AEI-Brookings Joint Center for Regulatory Studies. pp. 223–244.

Goodman, S. (2013) "Learning from the Test: Raising Selective College Enrollment by Providing Information" *Finance and Economics Discussion Series*, 2013-69.

Greenstein, S. and J. Prince (2007) "The diffusion of the Internet and the geography of the digital divide in the United States" In D.Q. Mansell, Robin and R. Silverstone (Eds.), *Oxford Handbook on ICTs*, Oxford University Press.

Grubesic, T. H. and A. T. Murray (2002). "Constructing the divide: Spatial disparities in
Broadband access." *Papers in Regional Science 81*, 197–221.

Hoxby, C. and C. Avery (2012, December). The Missing "One-Offs": The Hidden Supply of High-Achieving, Low Income Students. Working Paper 18586, NBER.

Hoxby, C. and S. Turner (2013). Expanding College Opportunities for High-Achieving, Low Income Students. Discussion Paper 12-014. Stanford Institute for Economic Policy Research.

Hoxby, C. and S. Turner (2015). What High-Achieving Low-Income Students Know About College. Working Paper 20861, NBER.

Hurwitz, M., J. Smith, S. Niu, and J. Howell (2015). "The Maine Question: How is Four-Year College Enrollment Affected by Mandatory College Entrance Exams?" *Educational Evaluation and Policy Analysis.* 37(1): 138-159.

Kane, T. (1995, July). Rising Public College Tuition and College Entry: How Well Do Public Subsidies Promote Access to College? Working Paper 5164, NBER.

Klasik, D. (2012). The college application gauntlet: A systematic analysis of the steps to four-year college enrollment. *Research in Higher Education*, 53(5), 506-549.

Klasik, D. (2013). "The ACT of enrollment: The college enrollment effects of required college entrance exam taking," *Educational Researcher*, 42(3), 151-160.

Luca, M. and J. Smith (2013). "Salience in Quality Disclosure: Evidence From U.S. News College Rankings." *Journal of Economics & Management Strategy*. Vol. 22(1): 58-77.

Pallais A (2015). “Small Differences that Matter: Mistakes in Applying to College.” *Journal of Labor Economics* 33(2):493-520.

Pallais A. and S. Turner (2006). Opportunities for Low Income Students at Top Colleges and Universities: Policy Initiatives and the Distribution of Students. National Tax Journal 59(2), 357-386.

PEW Research (2014) “Internet use over time” http://www.pewinternet.org/data-trend/internet-use/internet-use-over-time/ Accessed September 30, 2014

Smith, J. M. Hurwitz, and J. Howell (2015). “Screening Mechanisms and Student Responses in the College Market,” *Economics of Education Review*, 44:17-28.

United States Government Accountability Office (2006) “ Broadband deployment is extensive throughout the United States, but it is difficult to assess the extent of deployment gaps in rural areas” *GAO Report 06-426*

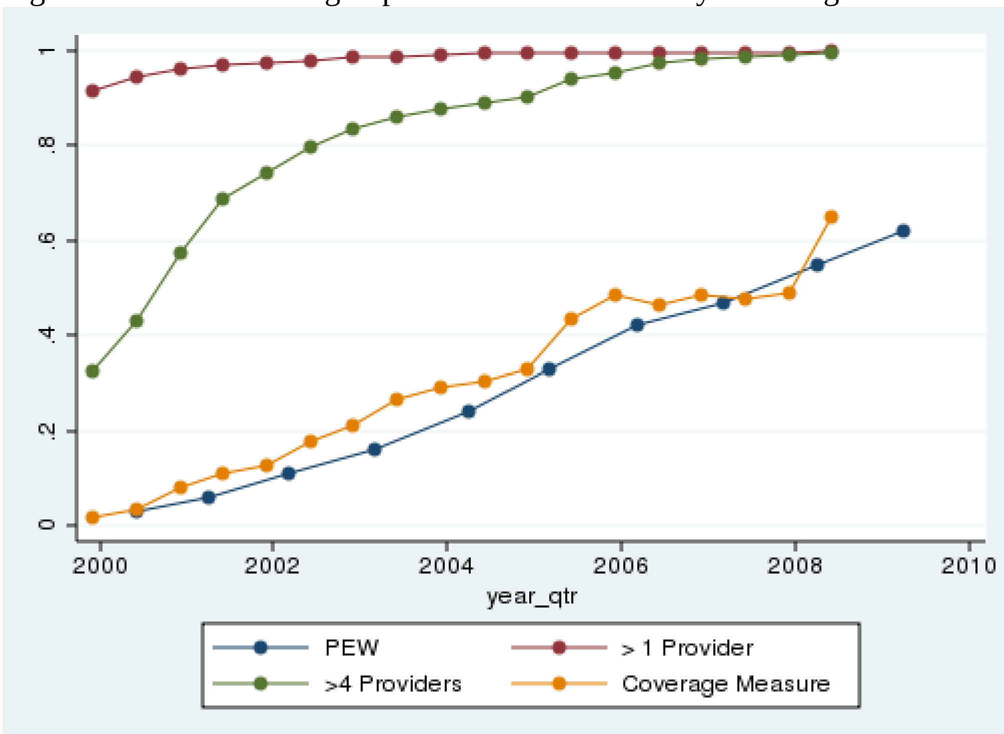
Vigdor, J., H. Ladd, and E. Martinez (2014). "Scaling The Digital Divide: Home Computer Technology And Student Achievement." *Economic Inquiry* 52(3): 1103-1119.

Zimmerman, S. (2014). The Returns to College Admission for Academically Marginal Students *Journal of Labor Economics* 32(4), 711-754.

Zucker, K. and A. Smith (2013). “Home Broadband 2013” August 26, 2013 http://www.pewinternet.org/2013/08/26/home-broadband-2013/

Tables and Figures

Figure 1: Measures of High Speed Internet Availability and Usage

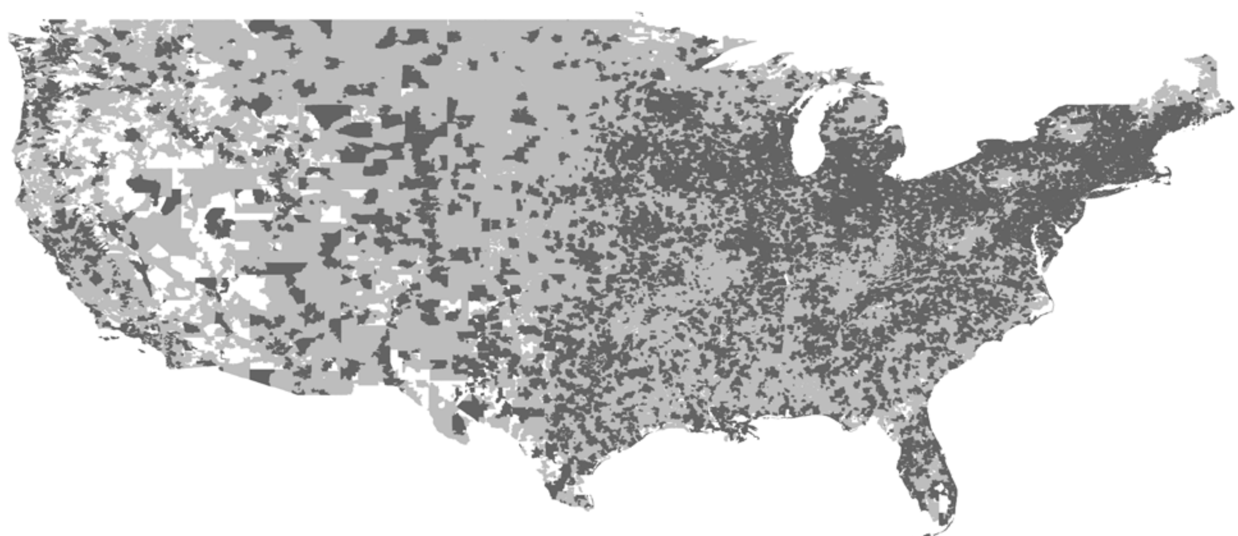


Notes: Displayed are trends in high-speed Internet availability and usage according to different measures. PEW refers to survey-reported national usage rates. The two provider measures (>1 and >4) refer to implied aggregate availability rates, if availability is defined as having at least one or at least four providers in the zip code. These rates are weighted by zip code population. The coverage measure refers to our preferred measure of high-speed Internet availability, as described in the text.

Figure 2: Zip Code Internet Penetration in 1999 and 2007

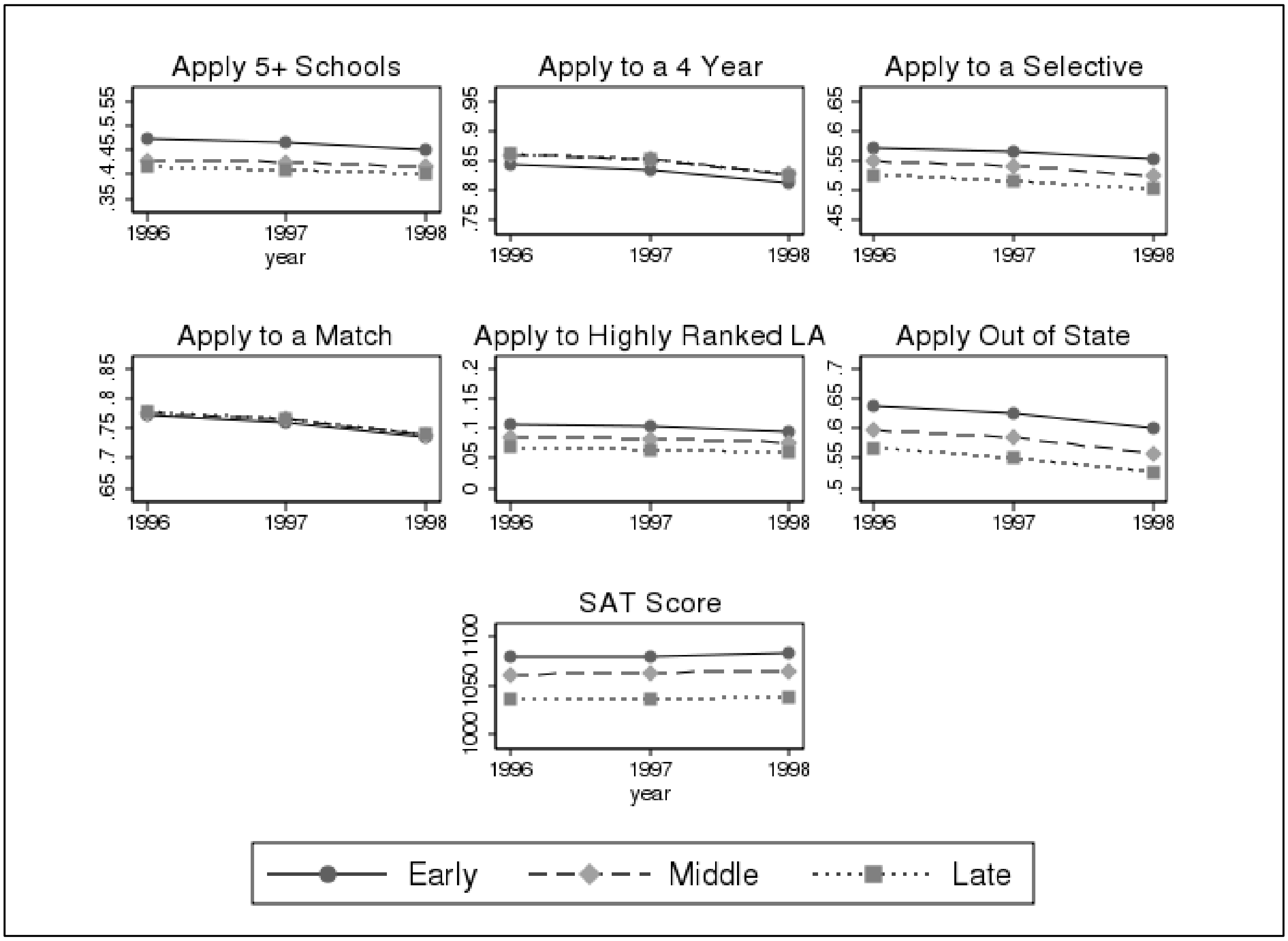
(A) December 1999

(B) December 2007



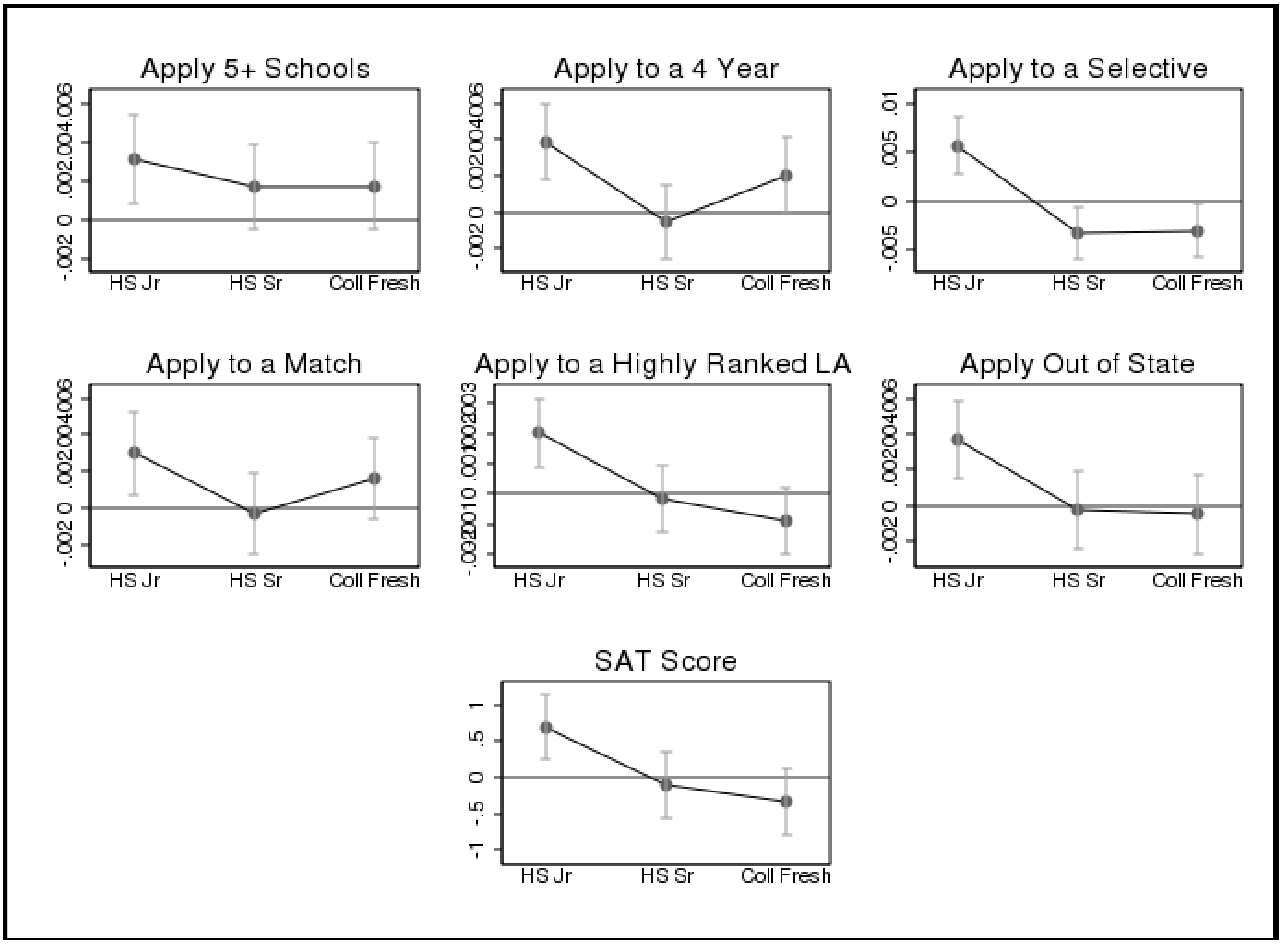
Notes: Displayed are zip code-level maps of Internet penetration. The light gray shading represents zip codes without coverage and the dark gray shading represents zip codes with coverage, where coverage is defined as our indicator for Internet availability, as described in the text. White areas are not included in the sample.
Sources: FCC and Census.

Figure 3: Means of Selected Application Outcomes in 1996-1998 by Year High-Speed Internet Entered Zip Code



Notes: Displayed are population-weighted means of the main application outcomes in the years just prior to the introduction of residential broadband Internet (1996 to 1998). Means were constructed by classifying zip codes according to the year in which high-speed Internet became available in the zip code.

Figure 4: Estimated Effects of Broadband Availability in Junior Year of High School, Senior Year of High School, and Freshman Year of College on Selected Application Outcomes.



Notes: Displayed are coefficients and 99 percent confidence intervals for the effect of broadband availability on application outcomes by year of schooling, estimated as described in the main text.

Table 1: Summary Statistics

	All Students (obs = 7,452,302)			
	Mean	*Std. Dev.*	*Min*	*Max*
Student Demographics and Exam Scores				
White	0.672	0.470	0	1
Black	0.107	0.309	0	1
Latino/Hispanic	0.095	0.293	0	1
Other Race	0.126	0.332	0	1
Male	0.447	0.497	0	1
Parental Education (0: HS; 1: Some college)	0.841	0.365	0	1
Cumulative GPA	3.363	0.610	0	4.33
PSAT Score	99.40	19.83	40	160
SAT score	1049.63	204.85	400	1600
Zip Code/County Level Variables				
Broadband Availability - Dec. of Junior Year	0.3177	0.4656	0	1
Mean Adjusted Gross Income	76,946	53,646	13,370	2,791,601
Population	4,111	9,799	-	132,944
Unemployment Rate	5	2	-	31
Median Home Price	192,769	118,034	-	943,877
Application Outcomes				
Applied to a 4-Year College	0.7897	0.4075	0	1
Applied to at least 5 Colleges	0.4023	0.4904	0	1
Applied to a College with an Avg. SAT of 1200	0.5032	0.5000	0	1
Applied to a College with an Avg. SAT of 1300	0.2972	0.4570	0	1
Applied to a Selective Liberal Arts College	0.1226	0.3280	0	1
Applied to a Top Liberal Arts College	0.0719	0.2583	0	1
Applied to a Match College	0.7071	0.4551	0	1
Applied to an In-State Flagship	0.2956	0.4563	0	1
Applied to an Out-of-State College	0.4920	0.4999	0	1

Notes: Sample is all SAT takers, high school graduating cohorts 2001 to 2008.

Table 2: Effect of Broadband Internet Availability in Junior Year on Testing and Application Outcomes

	Took the SAT	PSAT Score	SAT Score	Applied to at least 5 Colleges	Applied to a 4-Year College	Applied to a Match College
Broadband Access	0.00049	-0.01789	0.69798***	0.00327***	0.00357***	0.00292***
	(0.00081)	(0.03232)	(0.16392)	(0.00084)	(0.00081)	(0.00086)
Observations	12,556,552	7,452,302	7,452,302	7,452,302	7,452,302	7,452,302

	Applied to a College with an Avg. SAT of 1200	Applied to a College with an Avg. SAT of 1300	Applied to a Top Liberal Arts College	Applied to the Instate Flagship	Applied to an Out-of-State College
Broadband Access	0.00361***	0.00055	0.00172***	0.00109	0.00321***
	(0.00103)	(0.00074)	(0.00040)	(0.00080)	(0.00083)
Observations	7,452,302	7,452,302	7,452,302	7,452,302	7,452,302

Notes: Standard errors adjusted for clustering at the zip code-level are in parentheses (* p<.10 ** p<.05 *** p<.01). Sample includes PSAT and SAT takers in the high school graduating cohorts of 2001 to 2008. Estimates are from separate regressions that include zip code and year fixed effects, as well as controls for gender, race, high school GPA, PSAT verbal and math sections, and time-varying zip code characteristics.

Table 3: Effect of Broadband Internet Availability in Junior Year on Testing and Application Outcomes, SAT States Only

	PSAT Score	Took the SAT	SAT Score	Applied to at least 5 Colleges	Applied to a 4-Year College	Applied to a College with an Avg. SAT of 1200	Applied to a Match College	Applied to a Top Liberal Arts College	Applied to an Out-of-State College
Broadband Access	-0.06520*	0.00175*	0.70417***	0.00348***	0.00430***	0.00200*	0.00091	0.00148***	0.00408***
	(0.03571)	(0.00098)	(0.18552)	(0.00094)	(0.00092)	(0.00116)	(0.00083)	(0.00044)	(0.00092)
Observations	6,080,456	8,805,628	6,080,456	6,080,456	6,080,456	6,080,456	6,080,456	6,080,456	6,080,456

Notes: Standard errors adjusted for clustering at the zip code-level are in parentheses (* p<.10 ** p<.05 *** p<.01). Sample includes PSAT and SAT takers in the high school graduating cohorts of 2001 to 2008. Estimates are from separate regressions that include zip code and year fixed effects, as well as controls for gender, race, high school GPA, PSAT verbal and math sections, and time-varying zip code characteristics. SAT states are defined as states in which the SAT is the dominant exam.

Table 4: Effect of Broadband Internet Availability in Junior Year on Testing and Application Outcomes, Controlling for SAT Score

	Applied to at least 5 Colleges	Applied to a 4-Year College	Applied to at least 5 Colleges	Applied to a College with an Avg. SAT of 1200	Applied to Match	Applied to a Top Liberal Arts College	Applied to an Out-of-State College
Broadband Access	0.00286***	0.00325***	0.00286***	0.00313***	0.00292***	0.00154***	0.00280***
	(0.00083)	(0.00081)	(0.00083)	(0.00102)	(0.00086)	(0.00040)	(0.00082)
Percent of initial coefficient	*80*	*99*	*80*	*87*	*100*	*90*	*87*
Observations	7,452,302	7,452,302	7,452,302	7,452,302	7,452,302	7,452,302	7,452,302

Notes: Standard errors adjusted for clustering at the zip code-level are in parentheses (* p<.10 ** p<.05 *** p<.01). Sample includes PSAT and SAT takers in the high school graduating cohorts of 2001 to 2008. Estimates are from separate regressions that include zip code and year fixed effects, as well as controls for gender, race, high school GPA, PSAT verbal and math sections, and time-varying zip code characteristics.

Table 5: Effect of Broadband Internet Availability in Senior Year on Testing and Application Outcomes

	PSAT Score	Took the SAT	SAT Score	Applied to at least 5 Colleges	Applied to a 4-Year College	Applied to a College with an Avg. SAT of 1200	Applied to a Match College	Applied to a Top Liberal Arts College	Applied to an Out-of-State College
Broadband Access	-0.04729	0.00182**	-0.00338	0.00336***	0.00182**	-0.00036	0.00298***	0.00001	0.00010
	(0.03290)	(0.00080)	(0.17111)	(0.00085)	(0.00083)	(0.00101)	(0.00076)	(0.00039)	(0.00085)
Observations	7,452,121	12,556,865	7,452,121	7,452,121	7,452,121	7,452,121	7,452,121	7,452,121	7,452,121

Notes: Standard errors adjusted for clustering at the zip code-level are in parentheses (* p<.10 ** p<.05 *** p<.01). Sample includes PSAT and SAT takers in the high school graduating cohorts of 2001 to 2008. Estimates are from separate regressions that include zip code and year fixed effects, as well as controls for gender, race, high school GPA, PSAT verbal and math sections, and time-varying zip code characteristics.

Table 6: Effects of Broadband Internet Availability in Junior Year on Testing and Application Outcomes by Group

	PSAT Score	Took SAT	SAT Score	Applied to a 4-Year College	Applied to at least 5 Colleges	Applied to a College with an Avg. SAT of 1200	Applied to a Match College	Applied to a Top Liberal Arts College	Applied to an Out-of-State College
	Below median income zip								
Broadband Access	-0.16762***	0.00177	0.55126**	-0.00048	0.00054	0.00319*	-0.00109	0.00018	-0.00025
	(0.05087)	(0.00125)	(0.26502)	(0.00135)	(0.00135)	(0.00167)	(0.00150)	(0.00051)	(0.00135)
Observations	3,065,569	6,228,615	3,065,569	3,065,569	3,065,569	3,065,569	3,065,569	3,065,569	3,065,569
	Above median income zip								
Broadband Access	0.04606	-0.00073	0.22062	0.00369***	0.00230**	0.00326**	0.00348***	0.00169***	0.00286***
	(0.04224)	(0.00107)	(0.20831)	(0.00099)	(0.00106)	(0.00133)	(0.00105)	(0.00056)	(0.00103)
Observations	4,381,979	6,309,245	4,381,979	4,381,979	4,381,979	4,381,979	4,381,979	4,381,979	4,381,979
	High school or less								
Broadband Access	-0.18499***	-0.00332***	1.18410***	-0.00009	0.00231	0.00200	-0.00160	0.00114*	0.00206
	(0.06376)	(0.00084)	(0.37447)	(0.00189)	(0.00192)	(0.00206)	(0.00204)	(0.00061)	(0.00186)
Observations	980,650	1,017,740	980,650	980,650	980,650	980,650	980,650	980,650	980,650
	Some college or more								
Broadband Access	0.00278	-0.00083***	0.48472***	0.00313***	0.00288***	0.00384***	0.00291***	0.00165***	0.00261***
	(0.03402)	(0.00026)	(0.17461)	(0.00082)	(0.00092)	(0.00109)	(0.00089)	(0.00047)	(0.00090)
Observations	5,202,386	5,293,795	5,202,386	5,202,386	5,202,386	5,202,386	5,202,386	5,202,386	5,202,386

Table 6 Cont: Effects of Broadband Internet Availability in Junior Year on Testing and Application Outcomes, by Group

	PSAT Score	Took SAT	SAT Score	Applied to a 4-Year College	Applied to at least 5 Colleges	Applied to a College with an Avg. SAT of 1200	Applied to a Match College	Applied to a Top Liberal Arts College	Applied to an Out-of-State College
					Black/Hispanic				
Broadband Access	-0.01710	-0.00077	0.50816	-0.00083	0.00203	-0.00074	-0.00139	-0.00038	0.00007
	(0.03377)	(0.00168)	(0.36226)	(0.00168)	(0.00181)	(0.00199)	(0.00190)	(0.00062)	(0.00178)
Observations	5,947,624	3,006,168	1,504,678	1,504,678	1,504,678	1,504,678	1,504,678	1,504,678	1,504,678
					White/Other				
Broadband Access	-0.05844	0.00000	0.46961***	0.00320***	0.00298***	0.00422***	0.00274***	0.00208***	0.00271***
	(0.06730)	(0.00078)	(0.16955)	(0.00085)	(0.00089)	(0.00109)	(0.00090)	(0.00045)	(0.00088)
Observations	1,504,678	9,550,384	5,947,624	5,947,624	5,947,624	5,947,624	5,947,624	5,947,624	5,947,624
					Rural				
Broadband Access	-0.15285**	-0.00021	0.60706	-0.00176	0.00286	-0.00034	-0.00333	-0.00038	0.00071
	(0.07701)	(0.00175)	(0.40139)	(0.00219)	(0.00212)	(0.00231)	(0.00228)	(0.00101)	(0.00218)
Observations	887,587	1,856,087	887,587	887,587	887,587	887,587	887,587	887,587	887,587
					Urban				
Broadband Access	0.00301	0.00120	0.65184***	0.00393***	0.00302***	0.00425***	0.00345***	0.00196***	0.00302***
	(0.03519)	(0.00090)	(0.17790)	(0.00087)	(0.00091)	(0.00113)	(0.00093)	(0.00044)	(0.00089)
Observations	6,564,715	10,700,465	6,564,715	6,564,715	6,564,715	6,564,715	6,564,715	6,564,715	6,564,715

Notes: Standard errors adjusted for clustering at the zip code-level are in parentheses (* p<.10 ** p<.05 *** p<.01). Sample includes PSAT and SAT takers in the high school graduating cohorts of 2001 to 2008. Estimates are from separate regressions that include zip code and year fixed effects, as well as controls for gender, race, high school GPA, PSAT verbal and math sections, and time-varying zip code characteristics.

For Online Publication
Appendix

Appendix 1. Broadband Internet Penetration Variable Creation and Alternative Specifications

To construct our measure of zip code-level broadband Internet access, we merged the zip code-level FCC data on the number of ISPs in a zip code, to zip code-level information on (1) land area and (2) population. The land area data are from the 2000 Census. The population data comes from a combination of 2000 census data and SOI income tax data, which we used to construct zip-code-level population counts over time. We then constructed ratios of the number of ISPs to zip-code geographic size and population size. We assigned each zip code as urban or rural based on the fraction of its population which was urban, defining an urban zip code as one where more than 50 percent of the population is urban.

We also collected national and sub-national trends broadband Internet usage from the CPS/PEW data. The CPS data come from the CPS computer and Internet supplements/school enrollment surveys (August 2000, September 2001, October 2003, 2007, 2009). The PEW Survey is collected at least annually and the month of the survey was recorded. We interpolated the between months so the timing was equivalent with the FCC data (which were collected in June/December). We constructed national trends out of both series, and a separate urban and rural trends from the CPS series based on whether an individual lived in a metropolitan statistical area.

We then conduct the following exercise:

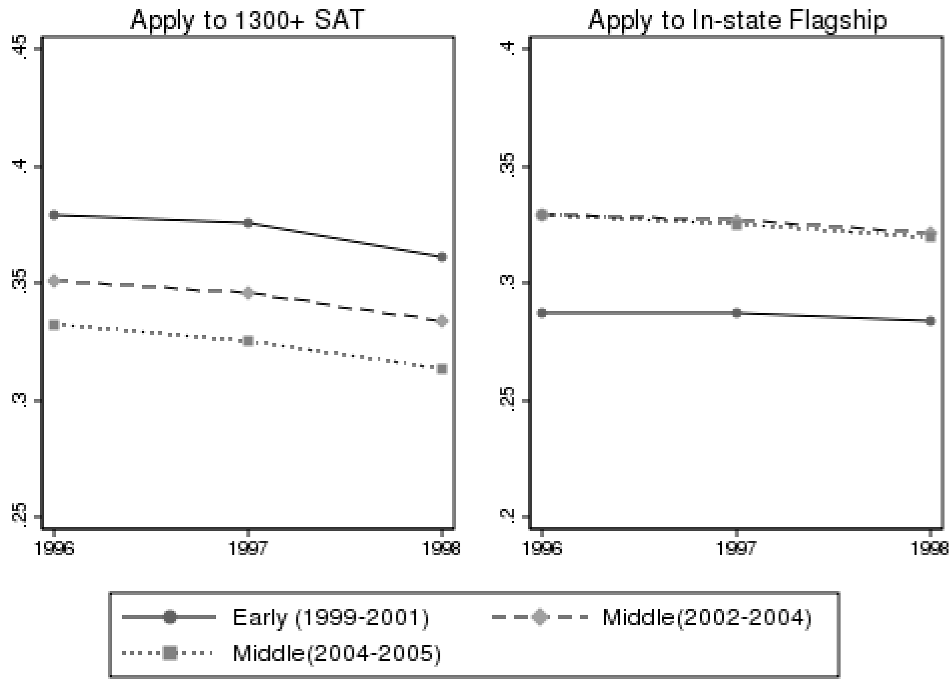
1) Identify a threshold for “coverage” (one provider per x square miles or one provider per y thousand people)
2) Construct an indicator which takes a value of one if threshold is in each zip code-year
3) Aggregate up using zip code population weights to construct national time series of high-speed Internet penetration
4) Estimate root mean squared error between CPS or PEW measure and #3

We then incrementally increase the threshold in step 1 and repeat steps 2 through 4. We tried the following combinations: one provider per 1-10,000 people (in intervals of 500 people), and one provider per 1-40 square miles (in intervals of one mile). We also allowed different thresholds for urban and rural zip codes (although we did not impose that they had to be different) by comparing to the CPS data. Finally, we minimized the root mean squared error (RMSE) and identified that threshold as the preferred definition of zip code broadband coverage. This is to define a rural zip code as “covered” when there is at least one provider per 12 square miles and an urban zip code as “covered” when there is at least one provider for every 2,700 people.

For exposition, we also present results for alternative internet penetration concepts, applying each of the population-based (i.e. “one provider per 2700 people”) and mileage-based (i.e., one provider per 12 square miles) measures to all zip codes (appendix table 1), as well as for the urban and rural subsamples separately (appendix table 2). Appendix table 1 demonstrates that the results using the population-based measure (second panel) are qualitatively similar to those obtained using our main measure (first panel). The RMSE for this single measure, however, is about 45 percent larger than the RMSE for our preferred measure, suggesting the fit to usage is worse. The third panel presents the results using the mileage-based measure (i.e. “one provider per 12 square miles”). In this case, the results are quite different from the results obtained using our main measure. However, this could be attributable to the fact that this measure fits the usage data poorly, with a RMSE of 2.79 (300 percent larger than the RMSE for our preferred measure). To further examine this, in the fourth panel of appendix table 1 we alternatively estimate the model using a mileage-based measure that better-fits the data (one provider per one square mile, which has an RMSE of 1.39). In this case, we find more similar results to both the preferred measure and the population-based measure. Taken together, these results are affirming and supportive of our strategy: alternative measures that more closely match national usage trends tend to lead to more consistent and similar results.

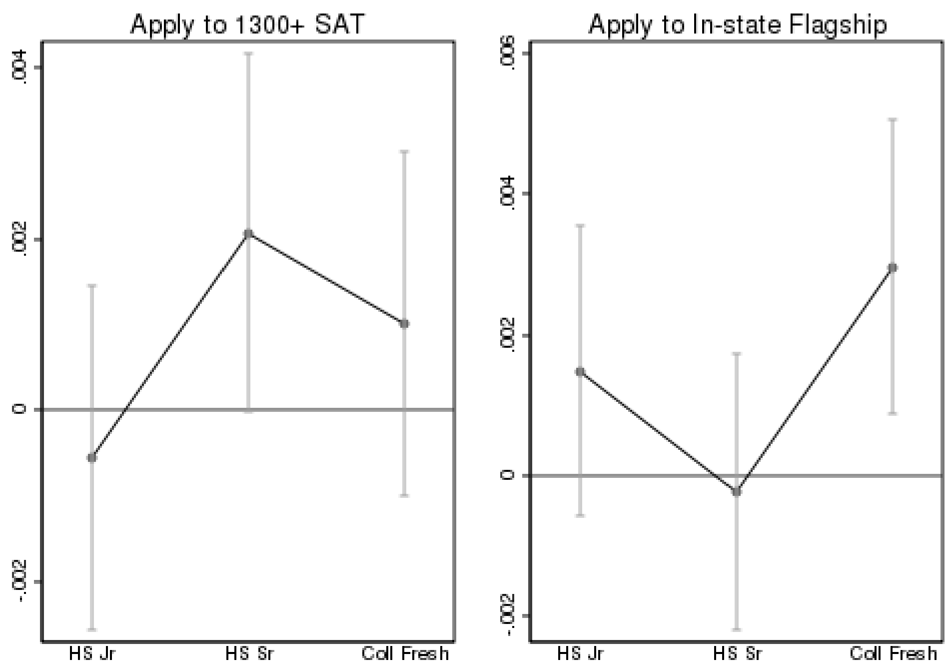
In appendix table 2 and 3 we consider the alternative three measures from the prior table separately within the urban and rural zip code subsamples. First, it is clear from the RMSEs that the separate population and mileage-based measures used in the preferred measure fit the urban and rural samples better than the alternative measures do, with RMSEs of 0.85 and 0.5, respectively. In addition, appendix table 2 indicates that the mileage-based measure (i.e., “one provider per 12 square miles”) performs poorly for the urban sample, with an RMSE of 2.89, and leads to results that differ from the population-based results. However, as in the pooled results, the mileage-based measure that fits the usage data more closely (i.e., “one provider per one square mile”) leads to results which are more consistent with the main results. Finally, appendix table 3 confirms the null results for the rural subsample in Table 6. In each case, regardless of the fit of the measure, we see little to no effect of Internet access on our outcomes for the rural subsample.

Appendix Figure 1a: Means of Additional Application Outcomes in 1996-1998 by Year High-Speed Internet Entered Zip Code



Notes: Displayed are population weighted means of the main application outcomes in the years just prior to the introduction of residential broadband Internet (1996 to 1998). Means were constructed by classifying zip codes according to the year in which high-speed Internet became available in the zip code.

Appendix Figure 1b: Estimated Effects of Broadband Availability in Junior Year, Senior Year and Freshman Year of College on Additional Application Outcomes.



Notes: Displayed are coefficients and 99 percent confidence intervals for the effect of broadband availability on application outcomes by year of schooling, estimated according to equation 1.

Appendix Table 1: Alternative Measures of Zip Code-Level Broadband Access

	Applied to at least 5 Colleges	Applied to a 4-Year College	Applied to a College with an Avg. SAT of 1200	Applied to a College with an Avg. SAT of 1300	Applied to a Match College	Applied to a Selective Liberal Arts College	Applied to a Top Liberal Arts College	Applied to the Instate Flagship	Applied to an Out-of-State College
	Main Measure of Broadband Access = 1 provider per 2700 people if urban and 1 provider per 12 square miles if rural (RMSE=0.92)								
Broadband Access	0.00327***	0.00357***	0.00361***	0.00055	0.00292***	-0.00041	0.00172***	0.00109	0.00321***
	(0.00084)	(0.00081)	(0.00103)	(0.00074)	(0.00086)	(0.00050)	(0.00040)	(0.00080)	(0.00083)
Observations	7,452,302	7,452,302	7,452,302	7,452,302	7,452,302	7,452,302	7,452,302	7,452,302	7,452,302
	Alternate Measure of Broadband Access = 1 provider per 2700 people (RMSE=1.34)								
Broadband Access	0.00134	0.00276***	0.00342***	0.00059	0.00228***	-0.00044	0.00139***	0.00073	0.00236***
	(0.00082)	(0.00079)	(0.00101)	(0.00073)	(0.00085)	(0.00049)	(0.00039)	(0.00078)	(0.00081)
Observations	7,452,302	7,452,302	7,452,302	7,452,302	7,452,302	7,452,302	7,452,302	7,452,302	7,452,302
	Alternate Measure of Broadband Access = 1 provider per 12 square miles (RMSE=2.79)								
Broadband Access	-0.00351***	-0.00526***	-0.00114	-0.00028	-0.00586***	-0.00318***	-0.00148***	0.00187*	-0.00281**
	(0.00110)	(0.00112)	(0.00133)	(0.00094)	(0.00118)	(0.00065)	(0.00050)	(0.00110)	(0.00113)
Observations	7,452,302	7,452,302	7,452,302	7,452,302	7,452,302	7,452,302	7,452,302	7,452,302	7,452,302
	Alternate Measure of Broadband Access = 1 provider per 1 square mile (RMSE=1.39)								
Broadband Access	0.00359***	0.00226**	0.00299**	-0.00179*	0.00209*	0.00265***	0.00056	-0.00151	-0.00063
	(0.00118)	(0.00115)	(0.00140)	(0.00097)	(0.00120)	(0.00067)	(0.00054)	(0.00114)	(0.00114)
Observations	7,452,302	7,452,302	7,452,302	7,452,302	7,452,302	7,452,302	7,452,302	7,452,302	7,452,302

Notes: Standard errors adjusted for clustering at the zip code-level are in parentheses (* p<.10 ** p<.05 *** p<.01). Sample includes PSAT and SAT takers in the high school graduating cohorts of 2001-2008. Estimates are from separate regressions that include zip code and year fixed effects, as well as controls for gender, race, high school GPA, PSAT verbal and math sections, and time-varying zip code characteristics. RMSE refers to the fit of the measure with national trends in Internet usage from PEW.

Appendix Table 2: Alternative Measures of Zip Code-Level Broadband Access for Urban Subsample

	Applied to at least 5 Colleges	Applied to a 4-Year College	Applied to a College with an Avg. SAT of 1200	Applied to a College with an Avg. SAT of 1300	Applied to a Match College	Applied to Selective Liberal Arts College	Applied to a Top Liberal Arts College	Applied to the Instate Flagship	Applied to an Out-of-State College
	Urban Students - Measure of Broadband Access = 1 provider per 2700 people (RMSE=0.85)								
Broadband Access	0.00302***	0.00393***	0.00425***	0.00064	0.00345***	-0.00014	0.00196***	0.00077	0.00302***
	(0.00091)	(0.00087)	(0.00113)	(0.00080)	(0.00093)	(0.00054)	(0.00044)	(0.00087)	(0.00089)
Observations	6,564,715	6,564,715	6,564,715	6,564,715	6,564,715	6,564,715	6,564,715	6,564,715	6,564,715
	Measure of Broadband Access = 1 provider per 12 square miles (RMSE=2.89)								
Broadband Access	-0.00468***	-0.00680***	-0.00067	-0.00061	-0.00666***	-0.00292***	-0.00156***	0.00143	-0.00469***
	(0.00128)	(0.00130)	(0.00158)	(0.00110)	(0.00138)	(0.00075)	(0.00058)	(0.00130)	(0.00131)
Observations	6,564,715	6,564,715	6,564,715	6,564,715	6,564,715	6,564,715	6,564,715	6,564,715	6,564,715
	Measure of Broadband Access = 1 provider per 1 square mile (RMSE=1.45)								
Broadband Access	0.00255**	0.00214*	0.00330**	-0.00118	0.00189	0.00235***	0.00020	-0.00111	-0.00081
	(0.00119)	(0.00117)	(0.00143)	(0.00098)	(0.00122)	(0.00068)	(0.00055)	(0.00116)	(0.00115)
Observations	6,564,715	6,564,715	6,564,715	6,564,715	6,564,715	6,564,715	6,564,715	6,564,715	6,564,715

Notes: Standard errors adjusted for clustering at the zip code-level are in parentheses (* p<.10 ** p<.05 *** p<.01). Sample includes PSAT and SAT takers in the high school graduating cohorts of 2001-2008. Estimates are from separate regressions that include zip code and year fixed effects, as well as controls for gender, race, high school GPA, PSAT verbal and math sections, and time-varying zip code characteristics. RMSE refers to the fit of the measure with separate urban and rural national trends in usage from the CPS.

Appendix Table 3: Alternative Measures of Zip Code-Level Broadband Access for Rural Subsample

	Applied to at least 5 Colleges	Applied to a 4-Year College	Applied to a College with an Avg. SAT of 1200	Applied to a College with an Avg. SAT of 1300	Applied to a Match College	Applied to Selective Liberal Arts College	Applied to a Top Liberal Arts College	Applied to the Instate Flagship	Applied to an Out-of-State College
Alternate Measure of Broadband Access = 1 provider per 2700 people (RMSE=2.36)									
Broadband Access	-0.00390**	-0.00298	-0.00102	-0.00328**	-0.00343*	-0.00087	-0.00091	-0.00061	-0.00167
	(0.00180)	(0.00182)	(0.00200)	(0.00155)	(0.00194)	(0.00108)	(0.00084)	(0.00165)	(0.00189)
Observations	887,587	887,587	887,587	887,587	887,587	887,587	887,587	887,587	887,587
Main Measure of Broadband Access = 1 provider per 12 square miles (RMSE=0.50)									
Broadband Access	0.00286	-0.00176	-0.00034	-0.00117	-0.00333	-0.00178	-0.00038	0.00430**	0.00071
	(0.00212)	(0.00219)	(0.00231)	(0.00178)	(0.00228)	(0.00126)	(0.00101)	(0.00192)	(0.00218)
Observations	887,587	887,587	887,587	887,587	887,587	887,587	887,587	887,587	887,587
Alternate Measure of Broadband Access = 1 provider per 1 square miles (RMSE=2.06)									
Broadband Access	0.01195	0.00963	0.02461*	0.01168	0.01992	-0.00724	-0.00385	-0.00064	0.02048
	(0.01556)	(0.01351)	(0.01328)	(0.01343)	(0.01346)	(0.00872)	(0.00695)	(0.01141)	(0.01359)
Observations	887,587	887,587	887,587	887,587	887,587	887,587	887,587	887,587	887,587

Notes: Standard errors adjusted for clustering at the zip code-level are in parentheses (* p<.10 ** p<.05 *** p<.01). Sample includes PSAT and SAT takers in the high school graduating cohorts of 2001-2008. Estimates are from separate regressions that include zip code and year fixed effects, as well as controls for gender, race, high school GPA, PSAT verbal and math sections, and time-varying zip code characteristics. RMSE refers to the fit of the measure with separate urban and rural national trends in usage from the CPS.

Appendix Table 4: Effects of Broadband Internet Availability in Junior Year on Testing and Application Outcomes, Controlling for Self-Reported Parental Education and Parental Income

	SAT Score	Applied to at least 5 Colleges	Applied to a 4-Year College	Applied to a College with an Avg. SAT of 1200	Applied to a College with an Avg. SAT of 1300	Applied to a Match College	Applied to a Top Liberal Arts College	Applied to the Instate Flagship	Applied to the In-State Flagship	Applied to an Out-of-State College
Broadband Access	0.64913***	0.00289***	0.00302***	0.00325***	0.00035	0.00242***	0.00172***	0.00095	0.00095	0.00271***
	(0.16265)	(0.00083)	(0.00080)	(0.00103)	(0.00074)	(0.00086)	(0.00040)	(0.00080)	(0.00080)	(0.00082)
Observations	7,452,302	7,452,302	7,452,302	7,452,302	7,452,302	7,452,302	7,452,302	7,452,302	7,452,302	7,452,302

Notes: Standard errors adjusted for clustering at the zip code-level are in parentheses (* p<.10 ** p<.05 *** p<.01). Sample includes PSAT and SAT takers in the high school graduating cohorts of 2001-2008. Estimates are from separate regressions that include zip code and year fixed effects, as well as controls for gender, race, high school GPA, PSAT verbal and math sections, and time-varying zip code characteristics.