The Search: The Effect of the College Scorecard on Interest in Colleges

Nick Huntington-Klein[a]

November 16, 2016

Abstract

The College Scorecard is a website launched by the U.S. Department of Education in September 2015 that provides information about different colleges. This paper studies the effects of the website on interest in colleges, calculating both intent-to-treat and a local average treatment effect estimates of the effect of the College Scorecard. Interest is measured using Google search activity. The Scorecard led to more searches for keywords associated with high-earnings, high-graduation rate, and low-tuition colleges. However, the size of the effect is very small.

Keywords: college choice, policy evaluation, information
JEL: I22, I23, I28, D83

[a]Corresponding Author. Department of Economics, California State University Fullerton, 800 N. State College Blvd., Fullerton, CA 92831-3599. nhuntington-klein@fullerton.edu.

1 Introduction

When making decisions about higher education, students and their families must learn what colleges are out there and what they are like, before applying to and attending a given college. Ideally, they will accurately understand the consequences of their decision. However, a lot of information about a given college is unavailable except through a thorough investigation, likely above the heads of many high schoolers.

Attempting to address this problem is the College Scorecard.[1] Announced in 2013 and fully launched in September 2015, the College Scorecard is a website compiling detailed information on a long and nearly comprehensive list of American colleges and universities. The site lists commonly available information like each college's location and enrollment size. Using the federal government's financial aid database, the Scorecard adds information that was previously difficult or impossible to find, such as average net tuition for financial aid recipients, student loan repayment rates, and the earnings of graduates.

Part of the purpose of the College Scorecard is to assist in funding decisions for higher education, and indeed the project was originally launched as a replacement for the Obama administration's controversial plan to rank colleges. The provision of information to students and families is also a major stated goal of the website. As emphasized by the Secretary of Education Arne Duncan at the time the project was announced (Feb. 13, 2013), "We know students and families are often overwhelmed in the college search process but feel they lack the tools to sort through the information and decide which school is right for them [...] The College Scorecard provides a snapshot about an institutions cost and value to help families make smart decisions about where to enroll." The site offers a huge amount of data to potential students, was widely publicized at the time of launch, and is free to use.

The Scorecard is an intervention that attempts to change student behavior by providing information to students. Interventions of this sort have been tested a number of times in research contexts, and are often of interest because they promise behavioral response that, even if it is small, comes at a very low cost per treatment. The Scorecard may be the largest example of a real-world implementation of such a policy in education.

The goal of this study is to examine the degree to which the release of

[1] https://collegescorecard.ed.gov/

the College Scorecard data affected interest in different colleges. I examine whether colleges with high marks in the attributes prominently advertised by the Scorecard (post-graduation earnings, net tuition, and graduation rate) saw increases in their Google search activity as a result of the Scorecard.

I produce an estimate of the effect of the Scorecard on aggregate Google search behavior, as measured by Google Trends. Google Trends reports the intensity of search behavior on Google for given keywords over time, relative to the pool of all Google searches. There is a fair amount of work using Google Trends as a tool that allows for instantaneous predictions of current variables like the unemployment rate Choi and Varian (2012). However, Google Trends data also has the capacity to be useful in observational estimation of casual effects, when search activity is a reasonable measure of the variable of interest. Stephens-Davidowitz (2014) uses Google Trends data to estimate the casual effect of racial animus on Barack Obama's vote share. Racial animus is measured by Google Trends results for racial epithets, to get around the hesitancy to report racist attitudes in survey data.

Google search behavior acts as a proxy for "first-step" interest in colleges, and is capable of picking up additional curiosity about a given college even if students eventually decide not to apply for any reason. Using IPEDS data, I find that in past years, a Google Trends index one unit higher is associated with 57 more applications for the college, or 3.3 more applications when controlling for college and year fixed effects.

The Scorecard had statistically significant effects on Google search activity, and those effects were of the expected signs. Colleges with high post-graduation earnings, high graduation rates, and low tuition saw increases in their search behavior relative to other colleges as a result of the Scorecard. However, these effect sizes were extremely small. Colleges saw bumps in their search activity that was .5%, .8%, 5.6% of a standard deviation of search activity relative to other colleges with tuition $1,000 higher, median earnings $1,000 lower, or graduation rates 5% lower, respectively. Effects were stronger for colleges that primarily offer bachelor's degrees than for colleges that primarily offer associate's degrees.

These estimates of the College Scorecard are intent-to-treat estimates, and do not take into account that most people performing searches for colleges likely did not visit the College Scorecard (or may not be students). I apply a method for estimating the local average treatment effect (LATE) in aggregate data. This method uses the implementation of the College Scorecard policy as an instrument for the Google Trend Index for the Scorecard itself. The LATE estimate is also small. Even comparing colleges that are very different in terms of their Scorecard-reported attributes, the effect of adding of a single treated user on aggregate search results is on the order of one ten-millionth of a standard deviation of search activity.

This paper finds that the College Scorecard did have an effect on search behavior. However, this effect is small enough to be considered negligible. The policy may be justifiable given that its marginal costs of operation are likely low, and the information has meaningful uses for researchers and government policy makers. However, the site is unlikely to have large effects on student behavior as intended.

2 Literature

This study looks into how student interest in different colleges relates to the attributes of those colleges, and how that interest changes in response to information revelation. Both of these topics have long and thorough literatures generally. There has been a recent boom in studies examining the intersection between educational choice and the revelation of information, and I will focus on that intersection here.

The decisions of whether or not to go to college, which college to go to, and what to major in are important determinants of individual and aggregate human capital. Multiple considerations drive these decisions. Benefits include the immediate consumption value of education (Wiswall and Zafar, 2015a; Jacob et al., 2016) as well as long-term financial and non-financial benefits such as improved labor market performance, improved health outcomes, and improved marriage market performance (Oreopoulos and Salvanes, 2011; Oreopoulos and Petronijevic, 2013). The costs of higher education include tuition and the possibility of being burdened with student loans (Dynarski, 2003).

The literature on college choice is not short of evidence that each of these incentives play some part in college choice, although results vary on the weight each is given in the decision-making process.

A growing literature shows that students are not fully aware of exactly how large these costs and benefits are. Mostly following from Dominitz and Manski (1996), there has been considerable interest in eliciting student beliefs about the labor market outcomes associated with different education levels, different institutions, or different majors (among many others, Kodde, 1987;

Smith and Powell, 1990; Betts, 1996; Avery and Kane, 2004; Rouse, 2004; Botelho and Pinto, 2004; Attanasio and Kaufmann, 2009; Zafar, 2011; Attanasio and Kaufmann, 2012; Stinebrickner and Stinebrickner, 2014; Hastings et al., 2016; Huntington-Klein, 2015, 2016b) as well as student beliefs about tuition (Avery and Kane, 2004; Usher, 2005; Grodsky and Jones, 2007; Booij et al., 2012).

A general summation of this literature is that students have relatively noisy beliefs about both labor market outcomes and the costs associated with college, although beliefs are on average not wildly inaccurate. These beliefs are associated with student choice. Students choose educational options they expect to have higher labor market returns, and are more likely to go to college if they think tuition is low.

The inaccuracy of student beliefs about the costs and benefits of college in general, or different college majors in particular, is a cause for concern. Students necessarily must make decisions on the basis of their perceptions of costs and benefits, rather than the truth (Manski, 1993). Without accurate beliefs it seems unlikely that students would be able to make well-considered decisions about their education. Further, students from disadvantaged backgrounds tend to have less accurate beliefs (Avery and Kane, 2004; Rouse, 2004; Hastings et al., 2016; Huntington-Klein, 2016b). Disparities in educational attainment may have to do with differences in beliefs and informational access.

The upside of undesirable educational choices following from student beliefs is that it opens a potentially inexpensive policy lever for improving educational choices. Following from the literature on student beliefs has been a string of studies that attempt to alter student beliefs by presenting them with new information. Ideally, these altered beliefs will then change matriculation patterns.

A good example of this type of intervention is in Wiswall and Zafar (2015b). The authors surveyed undergraduate students at New York University. These students were asked to report their beliefs about the earnings of current workers in the labor force conditional on having completed particular college majors, as well as their expectations for what their own earnings would be having graduated with that same major. Then, students were provided with information about the actual earnings of current workers, with a goal of providing publicly available information that students might be able to access themselves. Finally, students reported their beliefs again, so any revisions could be observed.

Wiswall and Zafar (2015b) find that students do meaningfully revise their beliefs in response to the new information. There are large changes in beliefs reported after the interventions. Then, when beliefs are elicited again two years later, beliefs correlate more strongly with post-intervention beliefs than pre-intervention beliefs. However, despite the changes in beliefs, Wiswall and Zafar (2015a) find that the changes in actual behavior that arise from the changes in beliefs are small relative to other influences. This result can be interpreted as meaning that students do not weight earnings heavily in their decisions, or that the revision in reported beliefs does not translate into a revision in the actual beliefs that students base decisions on.

The result that informational interventions have relatively small behavioral effects, even when reported beliefs change, is not limited to work by Wiswall & Zafar. Many intervention studies that intend to give information to students as a means to change beliefs and thus behavior find behavioral effects that are small or localized within particular subgroups, even in cases where beliefs are changed (Oreopoulos and Dunn, 2013; Kerr et al., 2015; Hastings et al., 2015; Bergman et al., 2016; Fryer, 2017) Bettinger et al. (2012) perform a study in which they provide information and aid about the FAFSA to customers of tax preparation service H&R Block. They find that, while interventions offering direct assistance in filling out the FAFSA forms did have meaningful and large effects on submitting the FAFSA and matriculating in college, an intervention offering information alone did not have similar effects.

Information-based educational interventions have had some success. Nguyen (2008) and Jensen (2010) find considerable response to information about the returns to secondary education in Madagascar and the Dominican Republic, respectively, although the context and level of education are very different than in studies about college. Perhaps the biggest success story relating to college is in Hoxby and Turner (2013), in which the authors target high-achieving, low-income students with information about their ability to attend and pay for selective colleges. While they test several interventions, including some that make the application process easier, they find that even information alone has meaningful effects on applications and matriculation.[2]

[2]There is also the growing related literature on "nudges" from short messages that remind the recipient about certain tasks they should be doing, such as in Castleman and Page (2015), which often find meaningful effects on behavior, especially given the low cost of the interventions. However, with the exception of messages that intend to inform (as in Fryer, 2017), rather than remind, these interventions should not be understood as purely

These generally mixed results from informationally based interventions does not mean that students are not open to the use of information at all. One source of publicly available and well-known information about college quality is the U.S. News and World Report, which ranks colleges on several attributes and provides overviews to the public. There is evidence that information about college quality as portrayed in the U.S. News and World Report affects college choice, at least for some kinds of students or colleges (Parker and Summers, 1993; Monks and Ehrenberg, 1999; Buss et al., 2004). Changes in rankings from year to year, which may be likely to overstate actual changes in quality and thus may be taken as largely an informational effect, influence the applications that students send (Griffith and Rask, 2007).

The College Scorecard falls somewhere between the publicly available general college overview and ranking available in the U.S. News and World Report and the careful, directed information revelation of the experimental literature.

Like the U.S. News and World Report, it is publicly available. However, the Scorecard is less well-established and less well-known. A Google Trends comparison of the search terms "US News College Rankings" and "College Scorecard" shows that the U.S. News search term beats the Scorecard by about two to one since the launch of the Scorecard. This likely understates the difference in popularity, since "US News College Rankings" is a very specific search term. As such, we might expect the Scorecard to have less effect than the U.S. News and World Report.

Like the experimental studies, the College Scorecard focuses on information that is generally not publicly available, or is difficult to find (earnings in particular). Also, the site is made available with the express purpose of correcting student beliefs and improving decision-making. The mixed results of these studies suggest that it is unlikely the College Scorecard will have a major effect, especially since the Scorecard is exposed only to those who seek it out, as opposed to an experiment in which all subjects at least see the information.

These expectations are validated in Hurwitz and Smith (2016), which examines whether student behavior in sending SAT scores to particular schools changed after the introduction of the College Scorecard. They find that tuition and graduation rate information did not have an effect on student SAT-sending behavior. However, they find that a 10% increase in the earn-

informational.

ings associated with a college increases the number of SAT scores received by 2.4%, an increase driven by students with advantaged backgrounds.

This study adds to this literature on the effects of information revelation on student choice. Clearly, it is most similar to Hurwitz and Smith (2016), which examines the effects of the same policy. This paper measures the effect of the information on Google search activity. Compared to the SAT score sending in Hurwitz and Smith (2016), Google search activity is a less direct measure of student matriculation in several ways. Some students may not use Google to look up colleges, perhaps clicking links directly from the Scorecard itself every time they want to learn about the college. Additionally, people other than students can search for college terms (the author, for one), and so an unknown portion of the search activity represents non-students.

While Google search activity is an indirect measure of student interest, it also allows the effect of the Scorecard to be captured at colleges that do not request SAT scores, like most two-year colleges. The addition of two-year colleges expands the analysis to consider the effect on a wider range of institutions. Google searches also include responses from students whose interest is piqued by the new information but do not end up applying, because they found other aspects unappealing or because they do not believe they have any chance of being admitted. The effect in this paper should be understood as the effect of the Scorecard information on *interest* in colleges, which may be a more clean estimate of the upper bound of the effects of this sort of similarly well known publicly-available college information. This paper can be paired with Hurwitz and Smith (2016) which provides an estimate that is closer to the effects on matriculation.

3 Data

This study uses two sources of data: the data from the College Scorecard website on the college attributes reported to visitors, and data from Google Trends reporting the intensity of search in the United States for search terms associated with particular colleges.

College Scorecard data includes information on 3,595 two- and four-year colleges. Most of the Scorecard data are compiled from the Integrated Postsecondary Education Data System (IPEDS), and the rest comes from federal databases on financial aid recipients. The website itself includes a means of searching for colleges by name, region, or type.

Individual college pages include extensive information on each college, but the most prominent are the enrollment of the college, its location, whether it is public or private, and three variables that are more difficult to find from other sources: the graduation rate (a four-year graduation rate for primarily two-year colleges, and a six-year graduation rate for primarily four-year colleges), the average in-state net price among students who receive federal financial aid, and the average earnings ten years after graduation among students who receive federal financial aid. This information is displayed relative to national averages. Figure 1 shows how the data are presented to visitors.

This study focuses on these three prominently displayed variables, although much more information is available in dropdown menus.

For each college campus in the College Scorecard sample, I generate a list of keyword searches for each college campus. Keywords are first generated using an automatic algorithm that takes as inputs the name of the college and the college website URL as listed in the College Scorecard data. For example, one might be "University of California - Santa Cruz" and "www.ucsc.edu/" The first keyword is the name of the full name of the college, "University of California Santa Cruz"[3] and the second is the base URL of the college stripped of extraneous detail, "ucsc.edu". The third drops the domain type, which generally leaves a common nickname for the college, here "ucsc". Next, the algorithm generates abbreviations of the college name, alternately including and stripping terms like "of," "the," and "at" and alternately abbreviating terms like "university" as "u," "community college" as "cc," or "technical institute" as "tech" (so we get "U of California Santa Cruz" and "U California Santa Cruz"), and alternately including or excluding the word "university" or "college" when it follows "state" (so "Bismarck State College" generates "Bismarck State" as a search term). The algorithm also generates keywords taking into account several widely used abbreviations, such as "UC" for "University of California" or "PSU" for "Pennsylvania State University".

After the algorithm is run, the keyword list is expanded by hand, plugging in likely search terms that are be suggested by the algorithm, like "Texas Lutheran" for "Texas Lutheran University." All search terms are then checked by hand, plugging them into Google one at a time. Search terms that do not bring up the college in the first five results are dropped. For example, "kc" is a search term generated by the algorithm for Ketter-

[3]Google Trends is not case-sensitive.

ing College from the URL "kc.edu" but it is deleted because searching for "kc" produces only references to Kansas City in the first five results, and not Kettering College.

I use the Python package pyGTrends to download Google Trends index information for each search term. The data include a weekly Google Trends index number for each search term from the beginning of April 2013 to the end of March 2016. Several keyword or keyword comparison searches generate no data because the search terms are too rare. Google does not report Trends results for these terms or comparisons because of privacy concerns. This leads to some colleges being dropped because all of their keywords returned no results, and so there are 3,439 college campuses in the sample.

The resulting data set links Google Trends indices for each search term to the College Scorecard information reported for the related college. Google Trends data does not report direct search volume, but rather an index of the popularity of a given search term relative to all Google searches, and so are not directly comparable across search terms. Rather, they report the number of searches for a term in a given week relative to other weeks covered by the index and the total number of Google searches in that time period (Stephens-Davidowitz and Varian, 2014). As such, in analysis it is necessary to account for a time trend, since the number of other Google searches the index is relative to changes over time.

Accounting for the time trend allows indices to be compared. This does not, however, adjust for differences in scale between the indices. A one-unit change in the index for an extremely popular keyword might represent 10,000 searches, where that same change for an unpopular keyword might represent 10. Results should be interpreted in terms of relative popularity change, rather than changes in absolute search volume, without further assumptions.

The resulting sample of Google Trends indices is large, with over 1.5 million raw keyword-by-week observations, which are collapsed to over 125,000 college-months for most analyses, averaging over each keyword for a given college. Aggregating all the keywords for a given college into one observation ensures that colleges with more keywords are not overrepresented, and makes results easier to interpret. However, since the index for each term is, as above described, on a different scale, an average of the keywords is not the same thing as a true Google Trends index for the college itself. In Section 5.1 I use alternate means of aggregating keywords together.

Summary statistics for the sample are in Table 1. Except for the Google Trends index, each average is taken at the college level. On average, each

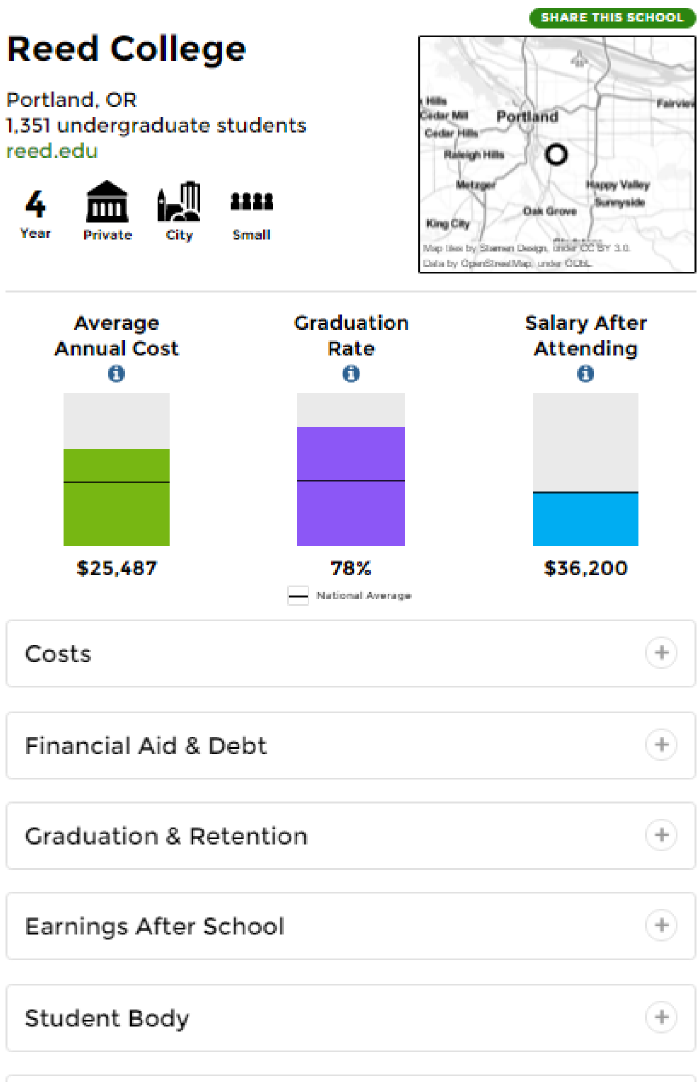
Table 1: Summary Statistics

Variable	Mean	(SD)	Min	Max
Earnings	\$37,601	(\$11,446)	\$11,600	\$166,200
Net Tuition	\$16,687	(\$7,833)	\$393	\$45,774
Graduation Rate	.437	(.209)	.022	1
Enrollment above 15,000	.067			
Enrollment 2,000 - 15,000	.385			
Enrollment below 2,000	.548			
Predominantly Associate's	.420			
Predominantly Bachelor's	.580			
Index	47.162	(21.497)	0	100
Keywords per College	2.901	(1.324)	1	9
Observation Level	N			
Colleges	3,439			
Keywords	9,976			
College-months	127,206			
Weekly Index Observations	1,561,739			

college has 2.9 keywords associated with it, although some campuses (in particular those that allow a lot of permutations of "University/U", "college"/omitted, and "of the"/omitted) have up to nine. Earnings are the median earnings of federal financial aid recipients ten years after graduation, and the mean over campuses of median earnings is \$37,601. Similarly, average net tuition for federal aid recipients is \$16,687. The average graduation rate is .437, which includes both the 58% of campuses that are primarily Bachelor's-granting institutions as well as the 42% that are primarily Associate's-granting. More than half of the colleges in the sample are small and have enrollments below 2,000.

These data can be used to estimate the effect of the college scorecard using a difference-in-difference specification that is capable of taking into account the context of the data.

Figure 1: College Scorecard Example



4 Estimation

In this section I describe the model used to estimate the effect of the College Scorecard on student interest in colleges, and the assumptions necessary to identify the estimate.

The model takes into account several features of the data and the intervention. First, the College Scorecard is implemented at a single point in time. So, the effect must be identified using cross-sectional variation in college attributes likely to interact with the effect of the Scorecard. Second, the effect of the Scorecard must be separated from long-term trends in the popularity of colleges on the basis of those attributes. Third, since popularity of a given search term in Google Trends data is given relative to popularity of the same term at other times, and not relative to other terms, care must be taken to scale inter-college comparisons appropriately.

To address all three of these issues, I use a difference-in-difference estimator with continuous college-attribute treatments, time trends, attribute-trend controls, and college fixed effects. The use of a difference-in-difference estimator with continuous college-attribute treatments identifies the model on the basis of college attribute differences in the effect of the Scorecard on interest in a particular college.

A college with low tuition, for example, might be expected to see a greater increase in Google activity relative to other colleges after the Scorecard is implemented. Attribute-trend controls address the possibility that these attributes have become more popular over time without the help of the Scorecard. If students have generally been more interested in low-tuition colleges over time, then the attribute-trend control allows the before/after effect of the scorecard to be separated from the general trend. Trend controls and attribute-trend controls also account for changes in the Google Trends data over time, since the pool of searches that the Google Trends indices are taken relative to also changes in size over time. College fixed effects allow for the Google Trends data to be compared across colleges.

In sum, the model of interest is

$$\begin{aligned} Index_{it} = \alpha_i + \beta_1 X_i \times Scorecard_t + \beta_2 X_i \times t + \beta_3 X_i \times t^2 \\ + \beta_4 t + \beta_5 t^2 + \beta_6 Scorecard_t + \varepsilon_{it} \end{aligned} \tag{1}$$

Where $Index_{it}$ is the aggregated Google Trends index for college i in period t, α_i is the college fixed effect, X_i is the attribute or attributes of interest, which could include tuition, earnings, or graduation rate, and $Scorecard_t$ is an indicator that the observation is from after the College Scorecard was implemented. β_2 and β_3 allow X_i to lead to additional interest over time regardless of the Scorecard. $\hat{\beta}_1$ is the estimate of interest, and shows how the introduction of the Scorecard led to additional interest in colleges with high levels of attribute X_i.

The trend controls make the observations comparable across time. Google Trends data provides the popularity of a given search term relative to the volume of other Google searches, and since the volume of Google searches changes over time, observations are not directly comparable. The time trend is intended to adjust for this. Results are robust to the interaction of the time trend with $Scorecard_t$, essentially forming a regression discontinuity design. The model is estimated using OLS with robust standard errors.

$Index_{it}$ combines the Google Trends data for multiple search keywords for the same college, so that colleges with more keywords are not overweighted. There are many ways to combine the searches for multiple terms into a single index, several of which will be covered later for robustness. The main method, however, uses equal weights, simply averaging together the indices for all keywords with enough search activity to not be omitted from the data to produce $Index_{it}$.

Time is taken at the month level, adjusted so that each month starts on the 12th, so that the policy launch date of September 12th, 2015, is not in the middle of a month. The weekly index score of each search term is averaged to create $Index_{it}$. Results are robust to the use of weekly data instead.

So, in the main model, the observation level is college-by-month, which is aggregated from keyword-by-week data.

The coefficient of interest in this model is $\hat{\beta}_1$, on the interaction term between an attribute and the post-Scorecard indicator. Given the nonstandard data source, interpretation is not obvious. $\hat{\beta}_1$ can be interpreted as the difference in how much that college's Google Trends index changed as a result of the Scorecard being introduced between two colleges with different levels of the given attribute. So if $\hat{\beta}_1 = .5$ and the attribute in question is Earnings in thousands of dollars, then we would say that the Scorecard increased the Google Trends index for a college with a \$20,000 median earnings by .5 more than it did for a college with a \$19,000 median earnings.

Additionally, the effect estimated here should be interpreted as the effect of the availability of information, not necessarily the effect of beliefs being changed. Without further data on beliefs there is no way to distinguish when the revelation of Scorecard information actually changes beliefs, as opposed to revealing something that everyone already knows.

$Index_{it}$ variable records aggregate search activity. This search activity includes both non-students searching for colleges, and students who have never seen the Scorecard. The estimate $\hat{\beta}_1$ should be properly understood as the intent-to-treat effect of the Scorecard on aggregate search behavior. In understanding the effect of the policy on individual behavior, a local average treatment effect (LATE) will be preferable to intent-to-treat.

To estimate the LATE, I use an instrumental variables method. This method is similar to Imbens and Angrist (1994), who show that in a randomized experiment with imperfect compliance, standard analysis identifies an intent-to-treat estimate, and the LATE can be identified using the randomization as an instrument for the receipt of the treatment. In Huntington-Klein (2016a) I modify the Imbens & Angrist method to apply to aggregate data in a difference-in-difference context, and use a version of that estimator here.

Treatment rate is not observed since there is no individual-level data. Instead of individual-level compliance data, I use an aggregate measure of interest in the policy. Specifically, data on Google searches for the College Scorecard. The main model then becomes

$$Index_{it} = \alpha_i + (\beta_1 N)X_i \times ScorecardIndex_t + \beta_2 t + \beta_3 t^2 + \varepsilon_{it} \quad (2)$$

Where $ScorecardIndex_t$ is the index of Google search activity for the College Scorecard in period t, which stands in for a variable indicating actual treatment, and N is the approximate number of College Scorecard users each unit of $ScorecardIndex_t$ represents, calculated using an observation of the number of website hits in April-May 2016 at the College Scorecard website.[4] The coefficient on $X_i \times ScorecardIndex_t$ is $\beta_1 N$, which can be adjusted by N to get β_1, interpretable as "one additional user of the Scorecard website leads to a $\beta_1 X_i$ increase in the search index for that college."

The use of the Scorecard and similar information sites, measured in $ScorecardIndex_t$, is of course endogenous, since people select themselves into treatment. As such, the $Scorecard_t$ variable is used as an instrument, under the assumption that the release of the Scorecard policy affected college

[4] https://analytics.usa.gov/ reports website hits in the past 30 days at time of viewing, but does not report old data on website hits.

search behavior only through the actual policy (or interest in similar information sources), and that Google searches for the policy is proportional to actual use. The first stage of the instrumental variables model is then

$$X_i \times ScorecardIndex_t = \alpha + \gamma_1 Scorecard_t + \gamma_2 t + \gamma_3 t^2 + \varepsilon_{it} \qquad (3)$$

The coefficient γ_1 captures only the increase in interest that occurs as a result of the Scorecard policy itself. That is, it isolates the compliers. The estimate of β_1 in Equation 2 is taken by dividing the coefficient on $X_i \times ScorecardInterest_t$ by N, and gives the average effect of the application of the policy to one additional treated person. Here, $\hat{\beta}_1$ is the LATE estimate, with the outcome variable measured on an aggregate level.

This method departs in several ways from Imbens and Angrist (1994). The identifying variation is not the result of a randomized controlled experiment, data is not at the individual level, and $ScorecardIndex_t$ is a measure of interest rather than a direct measure of the receipt of treatment.

5 Results

Table 2 displays the results of four different models of the form presented in Equation 1.

The coefficients on the Scorecard interaction terms can be taken as the aggregate treatment effect of the introduction of the College Scorecard on search activity for colleges with particular characteristics. So, for example, the .105 coefficient in model (2) indicates that if College A's graduates earn $1,000 more than College B's, then the introduction of the Scorecard raised the Google Trends index by .105 more for College A than for College B.

The first result that jumps out is the unintuitive sign on Tuition $\times$ Scorecard in Model (1). Higher tuition colleges saw more of a bump from the Scorecard than low-tuition colleges. However, this appears to be due to the correlation between tuition and the positive attributes of earnings and graduation rate. When all three are included in model (4), the sign becomes negative.

Earnings and graduation rate have intuitive and expected signs in all models they appear in. Colleges with higher earnings and higher graduation rates saw a larger increase in popularity from the Scorecard. The Scorecard had statistically significant impacts of largely the expected signs.

Table 2: Effects of the College Scorecard on Search Activity for Colleges

	Google Search Index			
Variable	(1)	(2)	(3)	(4)
Tuition ($1k) × Scorecard	.103*** (.015)			-.040** (.019)
Earnings ($1k) × Scorecard		.105*** (.010)		.066*** (.014)
Grad. Rate × Scorecard			9.879*** (.581)	9.453*** (.782)
College Fixed Effects	Yes	Yes	Yes	Yes
Quadratic Trend	Yes	Yes	Yes	Yes
Quadratic Trend × Attributes	Yes	Yes	Yes	Yes
N	121,582	117,142	113,664	107,115

*/**/*** indicates statistical significance at the 10%/5%/1% level.

However, while statistically significant, the coefficient sizes are extremely small. As shown in Table 1, the standard deviation across all index observations is 21.497. Even within colleges, the minimum, mean, and maximum standard deviation of the index is $3.033, 8.461$, and 27.041, respectively. As such, using model 4, these results suggest that comparing two otherwise similar colleges, one of which has tuition $1,000 lower, the low-tuition college would get an increase in their search index from the Scorecard that is $.040/8.461 = .005$ of a standard deviation of search activity. Similarly, a college with median earnings $1,000 higher would get a boost of $.066/8.461 = .008$, and a college with a graduation rate 5 percentage points higher would get a boost of $.05 \times 9.453/8.461 = .056$ of a standard deviation. While the Scorecard does have an effect on aggregate search behavior, that effect is meaningfully small. Meaningful differences in colleges translate to very small amounts of differential impact of the Scorecard.

The primary analysis here uses an aggregation of multiple keywords per college that equally weights each keyword, and treats each college similarly no matter its size. Alternate approaches are shown in 5.1. The analysis so far assumes that the average effect of the Scorecard is constant across all college types, which may be untrue, especially if those students who are likely to be interested in certain kinds of colleges are also more likely to use the Scorecard. I evaluate this possibility in Section 5.2. Additionally, while the effect of the Scorecard is meaningfully small, that the Scorecard has an effect at all may be considered an impressive feat given that only a small fraction of people searching for colleges are likely to have used it. This is addressed in Section 5.3.

5.1 Variable and Model Robustness

The Google Search Index for each college used for the models in Table 2 equally weights the Google Trends result from each keyword associated with the college, and treats a one-unit change in the index as equally meaningful across all colleges.[5]

In this section I show results for the same four models in 2 using three alternate calculations of the Google Search index. Using a comparative Google Trends index from December 2015 to March 2016, which allows up to five keywords to be compared in popularity against each other, I generate a popularity weight for each keyword based on how often it is used compared to other keywords for the same college.[6]

I use these popularity weights to create two alternate Google Search Index calculations. The first, "Weighted Google Search Index," constructs a weighted average of the keywords, using the popularity weights rather than weighting each keyword equally, and "Most Popular Keyword Search Index" uses only the Google Trends result for the single most popular keyword associated with each college.

These two alternate calculations still treat a one-unit change in the index as equally meaningful across all colleges. However, since each Google Trends result is indexed from 0 to 100, no matter how popular an individual keyword is, a single unit change on a very popular search term like "Harvard" might translate to a change of many more actual searches than a single unit change on a less popular term. To adjust for this, I take the most popular keyword for each college (as used for the Most Popular Keyword Search Index) and use Google Trends to construct a between-college ranking of the popularity of each college's most popular keyword.[7] I then weight the Most Popular

[5]While these results are not shown, results are also robust to the use of logarithms of earnings and tuition rather than levels, or using different standard specifications of the time trend.

[6]For colleges with more than five keywords, multiple comparative searches were performed, with one overlapping keyword so that all terms could be compared.

[7]Since Google Trends does not allow more than five keywords to be compared, I randomly create five-keyword sets and use Google Trends to rank the popularity of each

Keyword Search Index by the popularity ranking so that the resulting "Popularity Scaled Search Index" is still indexed from 0 to 100.

The Popularity Scaled Search Index, like the other search indices, does not offer enough information to say how many searches represent one unit on the scale. But it does allow me to test whether the original results arise only because of the difference in the meaning of the scale across colleges.

Table 3 repeats Table 2 using these three alternate outcome variables. While point estimates are sensitive to which variable is used, none of them contradict any of the qualitative findings of Table 2 in which estimates were not of an economically meaningful size, even though they were, with the exception of tuition, significant and of expected sign.

5.2 Results by College Type

Table 4 divides the sample in three different ways: first by enrollment size, then by primary degree awarded, and finally dividing colleges into terciles based on the median SAT of their incoming classes, or ACT if they do not request the SAT.[8] A large number of colleges do not report median incoming SAT or ACT in the College Scorecard data, and are dropped for this third analysis. All analyses mimic model (4) from Table 2, using the unweighted index and including all three characteristics at the same time.

The division by enrollment size displays some interesting differences. In particular, effects seem to be generally larger for the largest colleges, especially for the graduation rate effect. This is particularly interesting since these analyses use the equal-weight Google Search Index, and so a one-unit increase for a large college represents many more searches than would a one-unit increase at a small college. However, even at large colleges the effect size of each input, with perhaps the exception of graduation rate, is fairly small.

---

keyword in the set. I repeat this three times so each keyword is ranked in three different sets, for a total of $(5-1)\times 3 = 12$ "more popular/less popular" comparisons for each keyword. I then use these 12 comparisons per keyword to rank the popularity of all keywords. Because there is some noise in the Google Trends results due to sampling, this ranking is not exact, but is very close to exact.

[8]Median SAT is constructed by averaging the nonmissing median math, writing, and verbal scores (to allow for colleges that report math and verbal but not writing), and creating a tercile of the score. Then, the process is repeated for colleges that report ACT. The ACT tercile is used for colleges that do not report SAT.

Table 3: Alternate Google Search Index calculations

Weighted Google Search Index				
Variable	(1)	(2)	(3)	(4)
Tuition ($1k) ×	.053***			.003
Scorecard	(.009)			(.012)
Earnings ($1k) ×		.047***		.029***
Scorecard		(.006)		(.010)
Grad. Rate ×			4.496***	3.719***
Scorecard			(.349)	(.547)
N	121,286	116,809	113,442	106,967
Most Popular Keyword Search Index				
Variable	(1)	(2)	(3)	(4)
Tuition ($1k) ×	.135***			-.040
Scorecard	(.019)			(.024)
Earnings ($1k) ×		.149***		.111***
Scorecard		(.013)		(.018)
Grad. Rate ×			12.224***	1.462***
Scorecard			(.746)	(1.008)
N	120,139	115,736	112,443	106,079
Popularity Scaled Search Index				
Variable	(1)	(2)	(3)	(4)
Tuition ($1k) ×	.065***			.002
Scorecard	(.011)			(.014)
Earnings ($1k) ×		.065***		.051***
Scorecard		(.008)		(.012)
Grad. Rate ×			4.794***	3.484***
Scorecard			(.435)	(.627)
N	120,139	115,736	112,443	106,079

All models include college fixed effects, quadratic time trends, and quadratic trends interacted with all attributes. */**/*** indicates statistical significance at the 10%/5%/1% level.

Table 4: Effects of the College Scorecard by College Type

Enrollment	< 2,000	2,000-15,000	>15,000
Tuition ($1k) ×	-.039	.043	.023
Scorecard	(.029)	(.031)	(.085)
Earnings ($1k) ×	.030	.053**	.128*
Scorecard	(.020)	(.025)	(.073)
Grad. Rate ×	7.937***	7.156***	14.954***
Scorecard	(1.036)	(1.532)	(2.974)
Observations	52,096	46,694	8,325
Median SAT	Lowest	Middle	Highest
Tuition ($1k) ×	-.050	-.126**	-.124***
Scorecard	(.061)	(.057)	(.041)
Earnings ($1k) ×	.061	.033	.036
Scorecard	(.063)	(.049)	(.029)
Grad. Rate ×	5.137	7.833**	2.907
Scorecard	(3.289)	(3.203)	(2.578)
Observations	16,280	17,131	16,132
Pct. Receiving Pell	Lowest	Middle	Highest
Tuition ($1k) ×	.031	-.004	-.098**
Scorecard	(.029)	(.035)	(.041)
Earnings ($1k) ×	-.012	.102***	-.045
Scorecard	(.022)	(.032)	(.029)
Grad. Rate ×	11.820***	10.499***	1.527
Scorecard	(1.466)	(1.664)	(1.574)
Observations	38,332	37,666	31,117
Primary Degree	AA	BA	
Tuition ($1k) ×	-.019	-.129***	
Scorecard	(.030)	(.025)	
Earnings ($1k) ×	.012	-.032*	
Scorecard	(.035)	(.018)	
Grad. Rate ×	2.973**	11.618***	
Scorecard	(1.367)	(.999)	
N	45,288	61,827	

All models include college fixed effects, quadratic time trends, and quadratic trends interacted with all attributes. */**/*** indicates statistical significance at the 10%/5%/1% level.

The division of colleges by incoming SAT tercile shows that the graduation rate has the largest effect size for middle-tercile colleges, and the smallest effect for top-tercile colleges. Of course, these analyses only use a portion of the full sample, and while the point estimates are largest for the middle tercile, they are not statistically different from the effects for the lowest or highest tercile.

Earnings has a similar effect size as the main results, but is significant nowhere here, in contrast to Hurwitz and Smith (2016) who find that the earnings effect of the Scorecard was strongest for students with the highest SAT scores. I do not replicate that result here, and instead find no statistical difference. The difference may arise because they compare high and low SAT *students and high schools*, as opposed to high and low SAT *colleges*. If high-SAT students are a relatively small proportion of the sample, their response may not be visible in the aggregate college-level data.

Looking at colleges by the proportion of their students who receive Pell grants addresses an important finding in the prior literature. Previous studies typically find that students with low incomes are most responsive to new information, perhaps because they have the least information to begin with. However, the literature also finds that students with low incomes are the least likely to seek out information, such as the College Scorecard, in the first place.

While the tuition effect is concentrated among high-Pell recipient colleges, the graduation rate and earnings effects are much stronger at colleges with low and medium levels of Pell recipients. These results are rather surprising. The fact that tuition has its strongest effects among high-Pell colleges suggests that those interested in high-Pell colleges do seem to be finding and responding to information that those interested in lower-Pell colleges are not, which is unexpected. Given that there is some response to information, it is surprising again that the result is not replicated for the other attributes, when prior literature finds that low-income students respond more strongly to new information in general. Importantly, looking at the effects within groups of colleges based on the types of students they typically attract is not the same as looking at the effects within particular types of students, and this distinction may explain some of these unusual results.

The division in the Scorecard's effect by the primary degree awarded is perhaps the most stark. The effects of the Scorecard are be driven almost entirely by colleges that primarily award Bachelors' degrees (although the effect is small even for these colleges, and earnings has an unintuitive sign in this subgroup). These results make sense if those interested in community college are less likely to search for college information, or are more geographically constrained and thus have less reason to compare colleges.

The results in this section suggest that the effect of the Scorecard on aggregate search behavior is concentrated in BA-granting colleges, and colleges with medium to large enrollments. The particular college attributes that the Scorecard seems to boost differ based on the different types of colleges.

This analysis so far ignores selection between different types of colleges. If searchers respond to the Scorecard information by switching from AA-granting colleges to BA-granting colleges rather than to other AA-granting colleges with better attributes, this will not show up in Table 4. So, in Table 5, I show how the Scorecard information affected interest in colleges by their type, rather than by their attributes.

The first model shows how the Scorecard affects interest in colleges by college control and degree awarded. This distinguishes between six types of colleges: public, private non-profit, and for-profit, each in primarily-AA-granting and BA-granting varieties. Public AA-granting colleges are the reference group. The Scorecard does seem to have a significant impact on interest in different college types. In particular, the Scorecard seems to be directing interest away from AA-granting colleges and towards BA-granting colleges. This offers one potential explanation for the lack of a Scorecard effect among AA-granting colleges, if those who viewed the information searched instead for BA-granting institutions rather than choosing differently between AA-granting institutions, or if the Scorecard only incited search activity among those searching for BA colleges and had no effect on those looking at AA colleges. There is also evidence that the Scorecard seems to direct interest away towards not-for-profit BA colleges, relative to for-profit BA colleges. There is no significant effect for for-profit AA colleges relative to other AA colleges.

While these effects, like all the other effects of the Scorecard found in the paper, are meaningfully small, it is interesting that extending further information drives student interest away from for-profit BA colleges. Since the Scorecard does report for-profit status, it is plausible that this effect represents a small number of students who prefer not-for-profit colleges and learn from the Scorecard which colleges are for-profit.

The second model shows how interest in colleges by enrollment is affected, with large colleges as the reference group. The third divides colleges by test scores, with high-score colleges as the reference group. These models both

Table 5: Effects of the College Scorecard on College Type Interest

	(1)	(2)	(3)	(4)
Pub. BA-Granting	2.870***			
× Scorecard	(.231)			
Priv. AA-Granting	-.420			
× Scorecard	(.433)			
Priv. BA-Granting	1.241***			
× Scorecard	(.200)			
For-Profit AA	-.309			
× Scorecard	(.253)			
For-Profit BA	-.673*			
× Scorecard	(.357)			
Enrollment < 2,000		-3.678***		
× Scorecard		(.466)		
Enrollment 2,000-15,000		-2.455***		
× Scorecard		(.470)		
Lowest SAT Tercile			-1.002**	
× Scorecard			(.442)	
Medium SAT Tercile			-.622	
× Scorecard			(.423)	
Lowest Pell Tercile				4.115***
× Scorecard				(.285)
Medium Pell Tercile				3.338***
× Scorecard				(.297)
N	127,206	127,095	51,245	127,095

All models include college fixed effects and use an unweighted index.
*/**/*** indicates statistical significance at the 10%/5%/1% level.

show interest shifting towards high-enrollment and high-test-score colleges, respectively.

The fourth model shows how interest in colleges with different levels of Pell recipients was affected by the Scorecard. The Scorecard seems to have increased interest in colleges with low levels of Pell recipients relative to high levels. This could be interpreted as shifting attention towards low-Pell colleges, but also may reflect the type of person using the Scorecard, and the type of college they are likely to be interested in. People who would have been interested in low-Pell colleges being more likely to use the Scorecard, and searching more heavily in general as a result would also produce this result.

5.3 Local Average Treatment Effects

Table 6 shows the results of the instrumental variables regression described in Section 4. Each attribute is evaluated separately since there is only one instrument, with the note that the result for tuition should be considered with caution, given that it reverses sign when all three attributes are included in the same model in Table 1.

The first stage of the regression is strong in each case. Unsurprisingly, the release of the Scorecard itself significantly increases search intensity for "college scorecard." This jump in interest can be seen in Figure 2. The second stages have signs that match the individual-attribute regressions in Table 2. The magnitudes have changed, but keep in mind that these coefficients now represent the sum total effect of all compliers.

As mentioned in Section 4, an estimate of the LATE can be derived by dividing the coefficient on the second-stage interaction term (the total effect of all compliers) by N, the number of people each unit of $ScorecardInterest_t$ represents. I calculate N using website hits for the College Scorecard. From April 13-May 12, 2016, the College Scorecard was visited by 56,974 unique users. During this same period, the Google Trends index for "college scorecard" is 4, from a search window that covers the past three years.

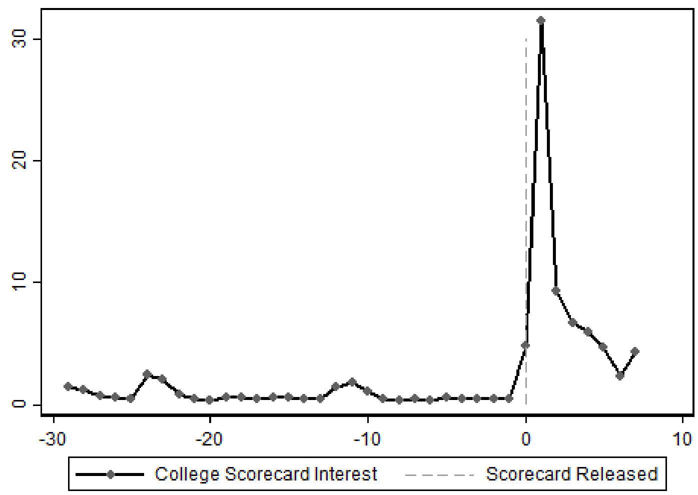
The Scorecard index is scaled so that the maximum and minimum search interest in a given period lie between 0 and 1. Making the simplifying assumption that the minimum search activity was a true zero, a single unit on the index translates into $56,974/4 = 14,243.5$ users. This is an inexact estimate, since the Google index reports only integers and not decimal points, and since scale may change somewhat over time as the total number of all

Table 6: Local Average Treatment Effects using IV

Variable	College Scorecard Interest
2SLS Second Stage for Tuition	
Tuition (\$1k) × $ScorecardInterest_t$	.004***
	(.001)
N	121,582
First Stage	
Scorecard	221.906***
	(1.175)
2SLS Second Stage for Earnings	
Earnings (\$1k) × $ScorecardInterest_t$	.002***
	(.000)
N	117,142
First Stage	
Scorecard	500.031***
	(2.468)
2SLS Second Stage for Grad. Rate	
Grad. Rate × $ScorecardInterest_t$	.184***
	(.021)
N	113,664
First Stage	
Scorecard	5.806***
	(.032)

All models include college fixed effects and a quadratic time trend. $ScorecardInterest_t$ is a Google Trends index for the term "college scorecard." */**/*** indicates statistical significance at the 10%/5%/1% level.

Figure 2: Google Trends Results for "College Scorecard"



Google searches changes Stephens-Davidowitz and Varian (2014). However, it is likely a reasonable approximation since most search activity for the Scorecard occurred over a relatively short recent window after its release.

I divide each of the coefficients by 14,243.5, focusing on Earnings and Graduation rate for these individual effect models. The coefficient on earnings then becomes 1.4×10^{-7}, and the coefficient on graduation rate becomes 1.3×10^{-5}. These are the estimates of β_1 in Equation 2. The treatment of one additional complier would increase the Google Trends index for a college by the product of β_1 and the attribute. One additional person treated because of the release of the Scorecard would change aggregate behavior in such a way that a college with $40,000 earnings would see an increase in their Google Trends index of $1.4 \times 10^{-7} \times (40 - 35) = 7 \times 10^{-7}$ more than a college with $35,000 earnings. Comparing a college with a graduation rate of 60% to one with 50%, the effect would be $1.3 \times 10^{-5} \times (.6 - .5) = 1.3 \times 10^{-6}$. In each case the effect is approximately one ten-millionth of a standard deviation of search behavior.

Again, it should be kept in mind that these figures rely on the estimate of N, the number of treated users, which is approximate. But under the current estimate, it would take an additional 1,000,000 users such that meaningfully different colleges would see an improvement of even .1 of a standard deviation. If N is overestimated by several orders of magnitude, this could be an impressive LATE. But the more N is overestimated, the farther away the Scorecard currently is from its goal. Overall, whatever the current number of users is, the user base would have to be increased by a factor of 176 in order for the Scorecard to have an effect of .1 of a standard deviation comparing colleges with \$5,000 income differences or 10% graduation rate differences.

6 Conclusion

This paper presents estimates of the effect of the College Scorecard on search behavior. The Scorecard led people to search more often for high-earnings, high-graduation rate, and low-tuition colleges, as intended. However, whether measured as the aggregate intent-to-treat effect of the Scorecard on search behavior, or as the local average treatment effect of one additional person using the website, the effect was extremely small. The Scorecard has only small effects on search patterns or aggregate search activity, and would need to recruit many more users in order to make a meaningful impact on aggregate search behavior and drive more interest towards high-performing colleges. These results largely match those of Hurwitz and Smith (2016), who study the effects of the Scorecard on student SAT-sending behavior.

It is worth emphasizing that affecting student behavior is not the only goal of the College Scorecard website, which also has uses for federal funding decisions and applications for researchers. Given the low marginal cost of providing information to students after these other needs have been met, the small effect sizes may be considered acceptable. Still, the policy is not having a major impact on its intended audience.

Evidence on other information-based interventions means the small effect of the Scorecard should have been predictable. Other studies of information-only interventions in college education, in which students are informed about the costs or benefits of different levels or types of education, tend to find small or null effect sizes.

The biggest successes of informational interventions in education tend to come when the policy targets low-income students or regions, where the information deficit may be largest (see Avery and Kane 2004; Rouse 2004; Hastings et al. 2016; Huntington-Klein 2016b on information deficiencies among low-income students, and Nguyen 2008; Jensen 2010; Hoxby and Turner 2013; Hastings et al. 2015 on informational interventions targeting low-income students or regions). However, the Scorecard, rather than informing students directly, makes information available on the internet, where it is likely that better-off students are more likely to access it. This may explain why Hurwitz and Smith (2016) find the largest effects of the Scorecard among students from privileged backgrounds, and I find that earnings and graduation rate have the strongest effects on search behavior for colleges with lower levels of Pell recipients.

In general, even among intensive interventions that directly provide information to students, we still do not know yet how to structure information-only interventions such that they all produce results as large as, say, Hoxby and Turner (2013), who manage to change student college application behavior by amounts as large as a half of a standard deviation. The Scorecard is a much less intensive intervention than even many of these only mildly successful studies, and so is fighting an uphill battle. It did not buck the disappointing trend.

However, the previous literature may provide some cause for optimism. Given that the Scorecard does work as intended, just not to the same degree as intended, increasing usage through marketing and, potentially, targeting usage, would increase the Scorecard's aggregate effect. With the current estimate of the LATE, the site would have to attract at least a hundred times as many users to have a meaningful impact. But, leaning on the literature's general finding that low-income students respond more strongly to information, the effect of each user may increase if the Scorecard can focus on recruiting low-income users. These students are likely to have more need of the information in the first place, and respond more strongly to the availability of the information if they know about the Scorecard and are encouraged to use it.

Finally, aside from any result about the College Scorecard in particular, this study outlines an approach to using Google Trends data in policy analysis more broadly. In any case where public interest or opinion is a policy-relevant variable, Google Trends may be used to capture that data in a way that does not rely on infrequent surveys with relatively small sample sizes. There are downsides to this approach as well; Google Trends indices can require work to make them comparable across time or across keywords, and the estimated effect size is not easily interpretable in absolute terms. But these are tradeoffs with the downsides that surveys themselves necessarily have. Just as it has previously been used in prediction, Google Trends data can also be used in the estimation of causal effects.

References

Attanasio, Orazio P., and Katja M. Kaufmann. 2009. Educational Choices, Subjective Expectations, and Credit Constraints.

———. 2012. Education Choices and Returns on the Labour and Marriage Markets: Evidence from Data on Subjective Expectations.

Avery, Christopher, and Thomas J. Kane. 2004. Student Perceptions of College Opportunities: The Boston COACH Program. In *College choices*, ed. Caroline Hoxby, 355–394. Chicago, IL: University of Chicago Press.

Bergman, Peter, Jeffrey T. Denning, and Dayanand Manoli. 2016. Is Information Enough ? Evidence from a Tax Credit Information Experiment with 1,000,000 Students.

Bettinger, Eric, Bridget T. Long, Philip Oreopoulos, and Lisa Sanbonmatsu. 2012. The Role of Simplification and Information in College Decisions: Results from the H&R Block FAFSA Experiment. *The Quarterly Journal of Economics* 127(3):1205–1242.

Betts, Julian R. 1996. What do Students Know about Wages? Evidence from a Survey of Undergraduates. *Journal of Human Resources* 31(1):27–56.

Booij, Adam S., Edwin Leuven, and Hessel Oosterbeek. 2012. The Role of Information in the Take-up of Student Loans. *Economics of Education Review* 31(1):33–44.

Botelho, Anabela, and Ligia C. Pinto. 2004. Students' Expectations of the Economic Returns to College Education: Results of a Controlled Experiment. *Economics of Education Review* 23(6):645–653.

Buss, Christian, Jeffrey Parker, and Jon Rivenburg. 2004. Cost, Quality and Enrollment Demand at Liberal Arts Colleges. *Economics of Education Review* 23(1):57–65.

Castleman, Benjamin L., and Lindsay C. Page. 2015. Summer Nudging: Can Personalized Text Messages and Peer Mentor Outreach Increase College Going Among Low-income High School Graduates? *Journal of Economic Behavior and Organization* 115:144–160.

Choi, Hyunyoung, and Hal Varian. 2012. Predicting the Present with Google Trends. *Economic Record* 88(SUPPL.1):2–9.

Dominitz, Jeff, and Charles F. Manski. 1996. Eliciting Student Expectations of the Returns to Schooling. *The Journal of Human Resources* 31(1):1–26.

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

Fryer, Roland G. 2017. Information, Non-Financial Incentives, and Student Achievement: Evidence from a Text Messaging Experiment. *Journal of Public Economics* (Forthcoming).

Griffith, Amanda, and Kevin Rask. 2007. The Influence of the US News and World Report Collegiate Rankings on the Matriculation Decision of High-ability Students: 1995-2004. *Economics of Education Review* 26(2): 244–255.

Grodsky, Eric, and Melanie T Jones. 2007. Real and imagined barriers to college entry: Perceptions of cost. *Social Science Research* 36(2):745–766.

Hastings, Justine S., Christopher A. Neilson, Anely Ramirez, and Seth D. Zimmerman. 2016. (Un)informed College and Major Choice: Evidence from Linked Survey and Administrative Data. *Economics of Education Review* (Forthcoming).

Hastings, Justine S., Christopher A. Neilson, and Seth D. Zimmerman. 2015. The Effects of Earnings Disclosure on College Enrollment Decisions.

Hoxby, Caroline, and Sarah Turner. 2013. Expanding College Opportunities for High-Achieving, Low Income Students.

Huntington-Klein, Nick. 2015. Subjective and Projected Returns to Education. *Journal of Economic Behavior & Organization* 117:10–25.

———. 2016a. A Method for Estimating Local Average Treatment Effects in Aggregate Data with Imperfect Assignment. *Applied Economics Letters* (Forthcoming).

———. 2016b. "(Un)informed College and Major Choice": Verification in an Alternate Setting. *Economics of Education Review* 53:159–163.

Hurwitz, Michael, and Jonathan Smith. 2016. Student Responsiveness to Earnings Data in the College Scorecard.

Imbens, Guido W., and Joshua D. Angrist. 1994. Identification and Estimation of Local Average Treatment Effects. *Econometrica* 62(2):467–475.

Jacob, Brian A., Brian McCall, and Kevin M. Stange. 2016. College as Country Club: Do Colleges Cater to Students' Preferences for Consumption? *Journal of Labor Economics* (Forthcoming).

Jensen, Robert. 2010. The (Perceived) Returns to Education and the Demand for Schooling. *The Quarterly Journal of Economics* 125(2):515–548.

Kerr, Sari Pekkala, Tuomas Pekkarinen, Matti Sarvimaki, and Roope Uusitalo. 2015. Post-Secondary Education and Information on Labor Market Prospects: A Randomized Field Experiment.

Kodde, David A. 1987. Unemployment Expectations and Human Capital Formation. *European Economic Review* 32(8):1645–1660.

Manski, Charles F. 1993. Adolescent Econometricians: How Do Youth Infer the Returns to Schooling? In *Studies of supply and demand in higher education*, ed. Charles T. Clotfelter and Michael Rothschild, chap. 2, 43–57. Chicago, IL: University of Chicago Press.

Monks, James, and Ronald G. Ehrenberg. 1999. The impact of US News & World Report College Rankings on Admissions Outcomes and Pricing Policies at Selective Private Institutions. *Change: The Magazine of Higher Learning* 31(6):42–51.

Nguyen, Trang. 2008. Information, Role Models and Perceived Returns to Education: Experimental Evidence from Madagascar.

Oreopoulos, Philip, and Ryan Dunn. 2013. Information and College Access: Evidence from a Randomized Field Experiment. *The Scandinavian Journal of Economics* 115(1):3–26.

Oreopoulos, Philip, and Uros Petronijevic. 2013. Making College Worth It: A Review of Research on the Returns to Higher Education. *The Future of Children* 23(1):41–65.

Oreopoulos, Philip, and Kjell G. Salvanes. 2011. Priceless: The Nonpecuniary Benefits of Schooling. *The Journal of Economic Perspectives* 25(1): 159–184.

Parker, Jeffrey, and Jeffrey Summers. 1993. Tuition and Enrollment Yield at Selective Liberal Arts Colleges. *Economics of Education Review* 12(4): 311–324.

Rouse, Cecilia Elena. 2004. Low-Income Students and College Attendance: An Exploration of Income Expectations. *Social Science Quarterly* 85(5): 1299–1317.

Smith, Herbert L., and Brian Powell. 1990. Great Expectations: Variations in Income Expectations among College Seniors. *Sociology of Education* 63(3):194–207.

Stephens-Davidowitz, Seth. 2014. The Cost of Racial Animus on a Black Candidate: Evidence using Google Search Data. *Journal of Public Economics* 118:26–40.

Stephens-Davidowitz, Seth, and Hal Varian. 2014. A Hands-on Guide to Google Data. Tech. Rep., Google.

Stinebrickner, Ralph, and Todd R. Stinebrickner. 2014. A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout. *Review of Economic Studies* 81(1):426–472.

Usher, Alex. 2005. A Little Knowledge is a Dangerous Thing: How Perceptions of Costs and Benefits Affect Access to Education. Tech. Rep., Toronto.

Wiswall, Matthew, and Basit Zafar. 2015a. Determinants of College Major Choice: Identification using an Information Experiment. *Review of Economic Studies* 82(2):791–824.

———. 2015b. How Do College Students Respond to Public Information about Earnings? *Journal of Human Capital* 9(2):117–169.

Zafar, Basit. 2011. How Do College Students Form Expectations? *Journal of Labor Economics* 29(2):301–348.