Digitization and Divergence: Online School Ratings and Segregation in America

Sharique Hasan
Duke University
Fuqua School of Business
Durham, NC
sharique.hasan@duke.edu

Anuj Kumar
University of Florida
Warrington College of Business
Gainesville, FL
akumar1@ufl.edu

November 26, 2019

Abstract

We analyze whether widespread online access to school-performance information affected economic and social segregation in America. We leverage the staged rollout of GreatSchools.org school ratings from 2006–2015 to answer this question. Across a range of outcomes and specifications, we find that the mass availability of school ratings has accelerated divergence in housing values, income distributions and education levels as well as the racial and ethnic composition across communities. Affluent and more educated families were better positioned to leverage this new information to capture educational opportunities in communities with the best schools. An unintended consequence of better information was less, rather than more, equity in education.

Keywords: Digitization, Ratings, Inequality, Education.

1 Introduction

Digitization and especially the internet are transforming social and economic life (Brynjolfsson and McAfee, 2014; DiMaggio and Bonikowski, 2008; Jorgenson, 2001). Because of these advances, individuals today can access extraordinary amounts of information to help them make important decisions. Job seekers, for example, can readily find ratings of workplaces; patients, ratings of hospitals; and parents, ratings of public schools. Although we know much more about how this information affects individual choices (e.g., Santos, Gravelle and Propper, 2017; Luca and Smith, 2013; Salganik, Dodds and Watts, 2006), our understanding of the broader social consequences of this mass availability of information remains limited.

Among the most critical decisions an American family makes is choosing where to live (Kane, Riegg and Staiger, 2006). For many, a crucial input to this decision is the quality of a community's public schools (Gibbons, Machin and Silva, 2013; Nguyen-Hoang and Yinger, 2011). Families have historically learned about school performance informally or inconsistently through social networks, real-estate agents, and other sources (Mikulecky and Christie, 2014; Figlio and Lucas, 2004). Because of this lack of consistent and accessible measures of school performance, families deciding where to live face substantial uncertainty. Despite such limited information, families' choices have still led neighborhoods to diverge economically (Owens, Reardon and Jencks, 2016). Today, however, parents have access to a substantial amount of school-performance information online. In this article, we ask whether this widespread access to school performance information has accelerated social and economic divergence.

We answer this question by leveraging zip code–year variation in the nationwide expansion of GreatSchools.org (hereafter, GS) ratings. GS, a nonprofit based in Oakland, California, provides detailed information about school performance for close to 100,000 public schools across America. The mission of GS is to empower parents to make better choices by providing detailed information about public schools. In 2003, GS expanded its ratings beyond its original state of California. Our data shows that coverage increased from 4,643 zip codes across five states in 2006 to 20,423 zip codes in 48 states plus Washington, DC in 2012. We use the rollout of school ratings to test whether this mass increase in school information accelerated or slowed inequality in home prices as well as the economic character of communities.

Across a range of specifications, we find that widespread access to school performance ratings accelerated divergence across zip codes. In our most conservative models, we find that housing prices for zip codes that are 1-sd apart in school performance diverge by an additional $2,283 after one year and $6,844 after three of rating availability. Further, we link rating availability for one and three years to an additional divergence of 0.21% and 0.66%, respectively, in the percentage of high-income earners for zip codes that are 1-sd apart in school performance. Finally, we find greater adjustment in the White and Asian population within communities in response to rating availability, with the proportion of these racial groups increasing in better school districts. We find no such effect for the Black population and a stronger negative relationship between school-performance and the Hispanic population when ratings are available online.

Our results speak to several streams of research in the social sciences. First, our study is one of the first to propose and test a novel mechanism for the increasing economic divergence across American communities (Reardon and Bischoff, 2011). We show that broader access to information increased segregation because high-income families could more readily leverage school ratings to move to neighborhoods with better schools. In this case, knowledge was indeed power, but only for the powerful. Second, our results speak to the growing literature on the social and economic impacts of digitization (Brynjolfsson and McAfee, 2014). Our research shows that the widespread availability of information enabled by the internet can have society-wide, and often unintended, effects. Finally, our results broaden the scope of the emerging research on online rankings by showing how they affect the outcomes of entire communities, not just individual consumers (Shore et al., 2015; Luca and Smith, 2013; Sauder and Espeland, 2009; Espeland and Sauder, 2007).

2 Literature Review

Today, an uneasy tension exists between two trends in American society. On the one hand, inequality is on the rise (Piketty and Saez, 2006, 2003). Growing inequities across individuals and families have led to both greater income segregation in American communities (Reardon and Bischoff, 2011) and divergence in access to economic opportunities (Owens, Reardon and Jencks, 2016). On the other hand, more Americans than ever have access to the internet and vast amounts of information to aid in their decision making (Pew, 2018). This widespread access promises to democratize knowledge and give all citizens the ability to find and take advantage of better opportunities (Brynjolfsson and McAfee, 2014).

One area where this tension is increasingly unfolding today is in the link between information, inequality, and access to an essential public good: schools. Although research indicates that access to good schools is highly unequal, parents today have unprecedented access to quantified school performance information online (Mikulecky and Christie, 2014). This information, some argue, gives parents—even lower-income ones—an essential tool for improving their child's education. At the same time, it may be the high-income families have greater access to this information and are those who can best leverage this information to find and capture the best opportunities (Reeves, 2017). A fundamental question is whether access to school performance information online has helped to slow economic and social divergence or to accelerate it.

2.1 Income Segregation and Inequality in Schooling

The rapid increase in top incomes, combined with the relative stagnation of wages for lower- and middle-income households, has reshaped many aspects of life in American communities. Reardon and Bischoff (2011), for instance, found that as income inequality increased from the 1970s into the 2000s, spatial segregation based on income grew as well. Furthermore, as high-income families became geographically concentrated, patterns of income and racial segregation further accelerated (Reardon, Fox and Townsend, 2015; Jargowsky, 1996).

A consequence of this rise in income segregation was its effect on American public education—a primary engine of economic opportunity (Coleman et al., 1966). Owens, Reardon and Jencks (2016), for instance, find that the effect of segregation led to a dramatic shift in the composition of school districts, with high-income families having disproportionate access to better-performing schools (see also Reeves, 2017). One implication of this segregation is that without access to good schools, the educational achievement of lower-income students is diminished (Ziol-Guest and Lee, 2016; Reardon, 2011; Mayer, 2002). As Quillian (2014) documents, income segregation negatively affected not only the high school graduation rates of poor students, but also their long-term college attendance and graduation rates (see also Mayer, 2002). In addition to the educational divergence between poor and non-poor households, income segregation appears to have exacerbated existing racial gaps in educational access and outcomes (Quillian, 2014; Logan, 2011; Sampson, Sharkey and Raudenbush, 2008; Rumberger and Palardy, 2005).

2.2 Information Availability and School Performance

A fundamental question asked by both policymakers and scholars is whether these performance gaps can be reduced. Some policymakers have taken the view that holding schools accountable for student outcomes may be one mechanism to improve school performance and reduce gaps (Harris and Herrington, 2006). Although interventions around the issue of accountability are somewhat diverse, two primary accountability mechanisms are widespread: (1) the administration of standardized testing via No Child Left Behind and (2) public availability of quantified performance measures for schools based on test results (e.g., Figlio and Rouse, 2006). Efforts at accountability are multifaceted and include both public and private initiatives. The State of Florida, for instance, assigned letter grades to schools in a performance “Report Card” (Figlio and Lucas, 2004). In addition to public efforts, organizations such as GS, city-data.com, and 50Can.org also collect performance data and publish measures of school-performance for use by parents and others (Mikulecky and Christie, 2014).

In theory, these quantified and widely accessible school-performance measures should serve as an important tool for parents in improving their child’s educational options. First, when parents have more information about the quality of their child’s schools, they can be more informed advocates for improving the schools their child attends.[1] Second, with more information about the performance of other schools outside their current district, parents can relocate to better neighborhoods with higher-performing public schools. Finally, rankings and ratings should ostensibly cause educational organizations to change in response to being evaluated (Shore et al., 2015; Espeland and Sauder, 2007; Sauder and Espeland, 2009; Espeland and Stevens, 1998).

A principal argument in this accountability narrative is that individuals make better choices with more information. The *information* mechanism has growing support in the literature. In a variety of settings, researchers have found that individuals do respond dramatically to performance information. They are significantly more likely to select better-rated options over lower-rated ones (Salganik and Watts, 2008; Salganik, Dodds and Watts, 2006; Chevalier and Mayzlin, 2006). In health

[1] GS, for instance, describes itself as “the leading national nonprofit empowering parents to unlock educational opportunities for their child.” (GS About Page) *Accessed: August 28, 2010 2:08pm EST.*

care, for instance, Santos, Gravelle and Propper (2017) find that public information on doctor quality led to an increase in demand for high-quality physicians. Varkevisser, van der Geest and Schut (2012) find similar results for patients selecting cardiologists. Similarly, Pope (2009) finds that hospitals that improved in "America's Best Hospitals" rankings saw significantly increased demand. Rankings also have a profound effect on the demand for educational institutions. Luca and Smith (2013), for instance, find that colleges ranked higher in *U.S. News and World Report College Rankings* received more applications.

Furthermore, research also suggests that this information is particularly valuable for disadvantaged students who may have gaps in their knowledge about where opportunities exist. Jensen (2010), for instance, finds that providing basic information about the financial returns from schooling increases educational persistence. Hoxby and Turner (2013) find that low-income students have limited knowledge about elite colleges, and simple mailers can dramatically increase their likelihood of applying to and attending these schools.

2.2.1 Differentiated Access and Use

The findings above, however, reflect "partial" treatment effects and ignore the systemic effects of providing greater information more broadly to both disadvantaged and higher-income families. In contrast to these predictions, it is possible that the benefits of school-performance information might have accrued to a limited subset of households; in particular, high-income ones. This differentiated effect of the mass availability of information can be a result of two related mechanisms. The first mechanism is that of differential access. A large literature on the "digital divide" suggests that access to the internet varied considerably by socioeconomic status. For instance, a large body of literature suggests that even basic access to the internet was unequal (Norris et al., 2001; Warschauer, 2004). Low-income families acquired access to the internet more slowly than high-income families, and also were limited to slower bandwidth. This differential access indicates that while this information was in theory widely accessible, some households may have had earlier and more widespread access to school-performance information. As a consequence, this differential access may have given higher-income families more information on which to base their school choice decisions.

The second, but related, mechanism is one of differentiated use. Although school-performance information might, in practice, be accessible to families in all socioeconomic strata, high-income households may be better able to take advantage of new information about school-performance. Figlio and Lucas (2004) in a study on several school districts in Florida, found that when schools were assigned performance "grades,"j housing prices adjusted to reflect school-performance. High-income families sorted into more expensive neighborhoods with better schools after ratings became available. While their study is valuable in that it can link the availability of performance measures to choices, substantial literature has long found that school-performance is capitalized in housing values (Gibbons, Machin and Silva, 2013; Fack and Grenet, 2010; Kane, Riegg and Staiger, 2006; Brasington and Haurin, 2006). High-income households are willing to pay a premium for homes with better-performing schools—approximately 4% more for a 1-sd better school-performance (for a review see Nguyen-Hoang and Yinger, 2011).[2]

Leveraging these two mechanisms described in prior work, we predict school performance information should affect the choices of high-income families more dramatically than lower-income families. Thus, in aggregate, the availability of school performance information should cause high-income families to leave zip codes with low-performing schools and move to communities with higher-performing schools.

[2]Several other authors find similar effect size. Brasington and Haurin (2006) find a 7.1% increase in housing prices for a 1-sd increase in school performance; Fack and Grenet (2010) in a study in France finds that housing values increase by 1.4 to 2.4% for a 1-sd increase in school performance.

3 Empirical Strategy

The gradual availability of online ratings of public schools by GS provides an unusual opportunity to estimate the effect of providing mass information about school performance on neighborhood composition and divergence across America. There is ample evidence that home prices and school performance are highly correlated (Nguyen-Hoang and Yinger, 2011).

As online ratings become available, we predict an upward shift in home prices for communities with better-performing schools and a downward shift for lower performing ones. This effect is reflected in the increased slope in the school performance and housing value relationship.

Furthermore, this shift in home values should also affect the economic and demographic composition of the affected communities. After rating availability, we predict based on prior research (e.g., Quillian, 2014) that higher-performing school districts should see an increase in higher-income and college-educated households, with more residents who are White or Asian, relative to those who are Black and Hispanic (Logan, 2011). Finally, we should expect higher rates of in-migration for communities where ratings are available for desirable, high-performing schools.

Our empirical analysis estimates the effect of the availability of school ratings via the internet (using the gradual availability of GS ratings as our proxy) on the changing economic and social character of American communities. Toward this goal, we combine several data sources. Our data are at the zip code–year level and include information on: (1) GS rating availability and average school performance; (2) housing prices; (3) proportion of high-income households; (4) racial and ethnic composition; (5) migration patterns; and (6) school ratings in pre-GS period from the Department of Education (hereafter, DOE) websites of 26 states plus Washington, DC. Below we describe our data sources, the construction of our variables, and estimated models.

Given that we use observational data to test our hypotheses, we consider several important alternative explanations for our findings in our robustness checks. In particular, our multipart empirical approach attempts to deal with issues of selection bias, omitted variables and reverse causality. These additional analyses are discussed after we present our primary results.

3.1 Data

GreatSchools.org: GreatSchools.org (hereafter, GS) is a national educational non-profit based in Oakland, California. It develops and disseminates quantitative ratings of thousands of American public schools based on the standardized test performance of their students. According to its website[3], GS provides:

> *...easy-to-understand information on K–12 schools, including ratings, information on school resources and student outcomes, and reviews.*

GS computes ratings using government-administered standardized test scores in subjects including mathematics, reading, and science. Although the actual test scores used to compute the GS ratings differ in content and measurement, GS normalizes these ratings into a decile scale, ranging from 1 through 10. The ratings are also color-coded to reflect quality differences, with green, orange, and red reflecting high, medium, and low performance, respectively. Figure 1 depicts examples of schools with ratings on the GS website and on the real estate website Zillow.com.

[Figure 1 about here.]

Our analysis leverages the school performance data available to GS beginning in 2006. At that time, the GS database included data on five states and 4,643 zip codes. By 2012, GS covered 48 states plus Washington, DC and about 20,423 zip codes in its database. From 2013 to 2015, GS maintained information on over 70,000

[3] https://www.greatschools.org/gk/about

schools. Table 1 presents the increase in coverage of GS data with respect to the number of states, zip codes, and schools from 2006 to 2012.

[Table 1 about here.]

Home Prices: Zillow: Zillow.com is an online real estate platform and database. We acquired zip code–level housing value data from Zillow.com's research database. Our primary dependent variable, *Housing Prices*, is derived from an aggregate measure of the value of all homes in a zip code called the Zillow Home Value Index (or ZHVI). The ZHVI, like the Case-Shiller Index, uses deed data for single-family homes, but also estimates sales prices for each home in a geographic area based on the characteristics of the home, tax assessment, sales transactions, and location using a hedonic approach (Dorsey et al., 2010). Prior research has found that the ZHVI is highly correlated to other standard home price indices (e.g., Case-Shiller Index), with $\rho = .96$ (Guerrieri, Hartley and Hurst, 2013), and has more comprehensive coverage (Damianov and Escobari, 2016). The measurements of the ZHVI are in dollars and are provided for each month beginning in 1997 until 2016, with scope increasing from 14,276 to 15,417 zip codes.

Household Adjusted Gross Income (AGI) Categories: Internal Revenue Service: The Internal Revenue Service (IRS) publishes an annual database of individual income tax statistics at the zip code level.[4] We use the tax statistics database compiled using the IRS data available through the National Bureau of Economic Research which includes information on the number of tax returns in each zip code, returns by AGI, exemptions, and other tax return items. Most relevant to our analysis are the number of households at each of the following 6 AGI levels: (1) \$1–under \$25,000, (2) \$25,000–under \$50,000, (3) \$50,000–under \$75,000, (4) \$75,000–under \$100,000, (5) \$100,000–under \$200,000, (6) \$200,000 or more. The IRS data cover

[4]This data can be found Here

the years beginning in 2005 until 2015, with zip code coverage ranging from 38,499 in 2005 to 27,680 in 2015.

Racial and Ethnic Composition, Education and Migration: American Community Survey: We use the American Community Survey (ACS) data product from the US Census Bureau to gather estimates of the racial and ethnic composition of zip codes. The ACS provides estimates of a zip code's total population, as well as population size by race and ethnicity (White, Black, Asian, and Hispanic). This data was obtained from ACS Demographic and Housing estimates available here on the American FactFinder product of US Census Bureau. We further obtained data on educational attainment of populations at the zip code level from the ACS Educational Attainment estimates available here and data on migration into the zip codes from the ACS Selected Social Characteristics estimates available at here.

3.2 Empirical Model

Our analysis examines whether the availability of GS ratings for a zip code i at time $t-1$ affected its economic and social composition at time t. Further, we hypothesize that the effect of online rating availability had an asymmetric effect, depending on the performance of the schools in that zip code i. When ratings went online for high-performing schools, home prices increased, the population of high income households increased, and additional White, Asian, and educated residents moved in. In contrast, when ratings became available for low-performing schools, home prices decreased, and high-income, White, and Asian families left. Below we describe the general specification of our empirical models, issues around the identification of our results, and the construction of the independent and dependent variables used to examine these effects.

Our basic model is of the following form:

$$
\begin{aligned}
Y_{it} = \beta_1 SP_{i(t-1)} + \beta_2 YearsAvail_{it} + \\
\beta_3 (SP_{i(t-1)} \times YearsAvail_{it}) + \alpha_x + \epsilon_{it}
\end{aligned}
\tag{1}
$$

Equation 1 is a panel model which exploits the variation in zip code–level characteristics over several years of our data. In this model, variable Y_{it} denotes the dependent variables including housing values, high-income household share, and ethnic and racial composition in zip code i in a year t. The variable $SP_{i(t-1)}$ reflects the standardized performance of schools in a zip code i for the prior year, $t-1$. Finally, the variable $YearsAvail_{it}$ denotes the number of years for which GS data have been available for the schools in zip code i by year t. We code the year in which the score is introduced as 0 with subsequent years iterating by +1. The main coefficient of interest in our model is β_3, which estimates the interaction effect of school performance and the exposure to rating availability.

Given the nature of our question and the need to use observational data to estimate the desired effects, we describe several potential issues with our estimation framework and how we deal with these in our models.

Reverse Causality: The first concern with our estimation strategy is reverse causality. It could be argued that higher house prices (reflecting higher-quality housing stock) in a zip code may result in greater demand from high-income households. A long stream of research in economics and sociology shows a strong intergenerational component to educational achievement—children from higher-income or better-educated families perform better in school. As a result, our findings may reflect a causal arrow going from housing prices to school performance, rather than the opposite. To account for this reverse relationship, all of our models use lagged measure of school performance $SP_{i(t-1)}$.

Selection: The next key issue in our estimation strategy is the problem of selection. It is possible that GS ratings are made available deferentially based on the demographic characteristics of communities, including based on prior demographics, income, housing values, and levels of education. To formally show that the characteristics of zip codes do not relate to *when* GS ratings were introduced, we estimate a Weibull hazard model. In this model, we regress the number of years from the start of our data (i.e., 2006) before the school scores of a zip code are introduced on GS on the percentage of households earning more than 100k per year, the average house prices (ZHVI), percentage of population of different races (White, Black, and Hispanic), and percentage of the population with college degrees. The standard errors are corrected at the state-level to account for the 'chunking' of GS ratings availability at the state-level (e.g., most zip codes in a state become available simultaneously). The results from this model are presented in table 2. Based on these results, GS does not appear to have a biased selection process for when it publishes ratings for a given zip code. Thus, we have evidence that variable $YearsAvail_{it}$ is unrelated to an important set of observable characteristics of zip codes.

[Table 2 about here.]

Omitted variables: Finally, our estimates are still susceptible to omitted variable bias, which can make β_3—the effect of ratings availability at each level of school performance—hard to interpret. For example, a variety of geography- and time-related omitted factors may simultaneously affect the housing prices and school performance in a community, thus biasing β_3, and our interpretation. In our models, we rigorously deal with these unobserved factors and demographic trends using several demanding fixed effects specifications, which are denoted by α_x in Equation 1. In our baseline specifications, we use zip code and year fixed effects. The zip code fixed effects capture differences in the scale of housing prices across communities.

The year fixed effects capture changes in house prices across the United States across years (e.g., during the period of financial crisis in 2008). However, the nature of the bias may be time-varying heterogeneity at the level of the community. Our models need to account for unobserved time-varying shocks such as changes in policies, investments, or business dynamics at a state or county level in a given year. For example, factors such as the entry or exit of a large employer, changes in taxation levels in a county, as well as other economic or social factors may bias our main coefficient. To account for these unobserved factors, we estimate more demanding specifications that include state-year (α_{st}) and then county-year fixed effects (α_{ct}) that capture non-parametric trends in housing values at the level of state-year and county-year, respectively.

In our robustness checks section, we conduct additional placebo tests to further rule out the possibility that omitted variable bias affects the interpretation of our results.

These checks and additional parameters in our regression allow us to deal with the main sources of specification error in our models. Below we describe our methodology and analysis.

3.3 Variables

Below we describe the construction of the independent and dependent variables used to examine these effects.

3.3.1 Independent Variables

To test our hypotheses, we construct two main independent variables, $YearsAvail_{it}$ and SP_{it-1}. We describe the construction and validation of these measures below.

Rating Availability: The primary treatment variable in our analysis is $YearsAvail_{it}$. This variable counts the number of years that ratings for at least one school has been available in zip code i by year t. For instance, if the first year that GS has data about a zip code i is $t = 2005$, we code $YearsAvail_{i,2005} = 0$, $YearsAvail_{i,2006} = 1$, $YearsAvail_{i,2007} = 2$, etc. Thus β_2 in Equation 1, the coefficient on $YearsAvail_{it}$ captures the change in the dependent variable (e.g., house prices in zip code i at time t) as a function of one additional year of rating availability.

School Performance: Our measure of school performance SP_{it-1} is at the zip code i and year $t-1$ levels. We construct our final variable using the mean score in a given school m for a given grade g in year t on the state administered Standardized Math Test, $SCORE_{mgt}$. The numeric score a school receives in the math subject test for a grade is the percentage of students who meet or exceed a given state's standards of passing performance. It is important to note that these are not "grades" on the exam. We standardize each score $SCORE_{mgt}$ into a Z-score for each school-grade-year observation: $ZSCORE_{mgt}$. This normalizes each school and grade's scores relative to all other schools whose students were tested using the *same* standardized test in that year.

Finally, we create an aggregate measure of school performance at the zip code level SP_{it} by calculating the mean of $ZSCORE_{mgt}$ for all schools in zip code i for a given year t. Thus, while the Z-score of a school captures its relative performance vis-à-vis other schools in the state taking the same standardized test in a year, the aggregated Z-score of all schools at a zip code level provides a normalized measure of relative school performance across geographies. As each state designs its standardized test and sets up the criteria for passing performance, the Z-scores at the zip code level may not correctly capture the relative school performance for zip codes across states. For example, a zip code in Minnesota with a similar Z score as a zip code in Florida may have a higher school performance. To mitigate this concern, we analyze within state-year and county-year variations in Z-scores at the zip code level using state-year and county-year fixed effects in our econometric specifications.

3.3.2 Dependent Variables

Home Values: Our primary measure of home values is the Zillow Home Value Index ($ZHVI_{it}$). The ZHVI is a seasonally adjusted measure of the average dollar value of a home in a zip code. Since this data is provided on a monthly level for each year, we use data for April, as it is the month with the most number of home sales nationally according to Zillow[5]. However, the correlation between monthly ZHVI indices across all months is $\rho > .99$, suggesting similar scaling for all months of the ZHVI data within a year. We use this data to examine the effect of rating availability on changes to home values. In our sample, the average home value is $217,843, with a minimum value of $13,600 (Earle, Arkansas) and a maximum of $5,442,900 (Atherton, California).

Percent Top Income: We use the Internal Revenue Service's Individual Income Tax Statistics database to construct a variable calculating the percentage of households with Adjusted Gross Income over $100,000 in a given zip code–year. We use the Urban Institute definition to define $100,000 and above as the threshold for the upper-middle class (Rose, 2016). In our sample, the average zip code had 13.6% of households earning more than $100,0000 per year. There were 3,650 zip codes with "0" households earning more than $100,000 per year.

Percentage White, Black, Asian, and Hispanic: We use the US Census Bureau's American Community Survey (ACS) data to construct our demographic variables. From 2010 to 2016, the Bureau publishes estimates for the number of White, Black, Asian, and Hispanic residents in a zip code. The average demographic

[5] Found here

of a zip code was 77.4% White, 7.4% Black, 8.8% Hispanic, and 1.9% Asian. Over this period, the % White population of all zip codes declined from 77.4% to 76.9%.

Percent College-Educated: We use the ACS data to calculate the proportion of college-educated residents in a zip code in a given year. Approximately 30% of residents in an average zip code had an associates degree or higher.

Migration: Finally, we use the ACS data to examine the degree of migration into a zip code for a given year. During our analysis period, we see that the average zip code had 12.1% in-migration, of which 9.9% were from within the state.

We present summary statistics in Table 3.

[Table 3 about here.]

4 Results

We begin our analysis by estimating Equation 1 using housing values, $log(ZHVI_{it})$, as our dependent variable.[6] We present these results in Table 4. Model 1 estimates the basic model without any fixed effects. The coefficient of interest is *School Performance* $*$ *Years Available*, which is positive and statistically significant.

The coefficient estimate $SchoolPerformance = 0.300$ suggests that home prices in a zip code with 1-sd better schools were 34.9% higher than that of a zip code with average schools. The coefficient estimate for $SchoolPerformance{*}YearsAvailable = 0.014$ indicates that the difference in house prices between such zip codes became 36.9% (an increase of 5.4%) and 40.8% (an increase of 16.6%) with one and three years of availability of the school ratings, respectively.[7]

[6] We take the logarithm of ZHVI to account for right skew in our data.

[7] Before school performance availability$\longrightarrow$ log(1-sd/average)ZHVI= 0.3 $\longrightarrow$ % Δ ZHVI = 100*(exp(0.3)–1)= 34.9%. After one year of school performance availability$\longrightarrow$ log(1-sd/average)ZHVI= 0.3 +1*0.014=0.314$\longrightarrow$ % Delta ZHVI = 100*(exp(0.314)–1)= 36.9%.

Figlio and Rouse (2006) found that the home values increase by 6.7 percent over a three-year period in areas of state assigned grade "A" schools to that of grade "B" school areas, and no difference in home values for grade "B" school areas and grade 'C' school areas over three years. Kane, Riegg and Staiger (2006) similarly found a 9.8% increase in home values with a 1-sd increase in school test scores than average in Mecklenburg County, North Carolina.

[Table 4 about here.]

In Model 2, we include zip code and year fixed effects separately to account for unobserved heterogeneity at the zip code level and yearly trends in housing prices. Again, we find a positive and significant coefficient for β_3. With these fixed effects, our coefficient drops to $\beta_3 = 0.008$. The next two models include state-year fixed effects and county-year fixed effects that account for different trends in housing values across states and counties, respectively. In these models, we formally account for the potential for each county (or state) to have differing trends in housing prices based on changes to tax policy, crime rates, or other factors that may have changed during this period. Again, we find consistent coefficient estimates for β_3. In Model 4, which is our most conservative specification, we find $\beta_1 = 0.266$ and $\beta_3 = 0.008$, which suggest that the house price in a zip code with 1-sd better schools was 30.5% higher than that of a zip code with average schools, but with availability of school performance information for one and three years, this gap widened to 31.5% and 33.6%, respectively.

Regarding dollar values, our estimates suggest that housing prices in a zip code with 1-sd better schools was \$66,384 higher than that of a zip code with average schools, but with availability of school performance information for one and three years, this gap widened to \$68,667 and \$73,228 (increased by \$2,283 and \$6,844), respectively.

The change in housing prices also signals a potential change in the underlying demographics of zip codes where ratings became available. Next, we estimate Equation 1 using the percentage of high-income households, *% 100k+*, as our dependent variable. These results are presented in Table 5.

[Table 5 about here.]

In Model 1, we present the results without any fixed effects. The coefficient estimate $SchoolPerformance = 5.754$ suggests that % 100k+ income households in a zip code with 1-sd better schools was greater by 5.75% than that in a zip code with average schools. The coefficient estimate $SchoolPerformance * YearsAvailable =$ 0.219 indicates that the difference in the percentage of %100k+ households between the two zip codes increased to 5.96% and 6.41% with one and three years of availability of the school ratings, respectively. With zip code and year fixed effects in Model 2, the coefficient is smaller at $\beta_3 = .054$, but still statistically significant.

In Models 3 and 4, we include state-year and county-year fixed effects. Our results remain within the general bounds of Models 1 and 2. Models 3 and 4 produce $\beta_3 = .237$ and $\beta_3 = .201$, respectively. Our results remain consistent with zip code fixed effects included with county-year and state-year effects that account for temporal heterogeneity across counties or states.

Overall, these results suggest a widening gap in the proportion of high-income households in zip codes with low-performing schools and those with high-performing schools. Regarding magnitudes based on model 4, The gap in the percentage of 100k+ income households between zip codes with 1-sd better schools and zip codes with average schools increased from 5.7% (over an average value of 13.6) without school performance availability to 5.9% and 6.3% with one and three years of school performance availability on the GS (i.e., an additional divergence of 0.2% and 0.6%, respectively).

4.1 Mechanisms Tests

Demographics: There is a strong correlation between income levels, race, and ethnicity in American society (Reardon, Fox and Townsend, 2015). In the models presented in Table 6, we estimate the impact of rating availability on the changing composition of communities. Again, we estimate Equation 1 with county-year fixed effects. We cluster our standard errors at the county-year level to account for intra-county correlation across time in policy and other shocks.

In these models, our dependent variables are the percentage of White, Black, Hispanic, and Asian residents in a zip code (Models 1 through 4). Furthermore, in Model 5, we present results for the proportion of individuals with associates degrees or higher. Broadly, we find the demographics of the communities in which ratings became available began to diverge. Quantitatively, we find that the differences between zip code with 1-sd better schools and zip codes with average schools widened with one year of school performance availability – increase in White population by 0.31% , Asian population by 0.16% , college educated residents by 0.42%.

It appears availability did not have a significant effect on the percentage of Black residents, but did reduce the percentage of Hispanic residents ($\beta = -0.69\%$).[8]

Like the prior results, these estimates suggest that when ratings become available, the racial and ethnic composition of communities shifts. For example, regarding magnitude, the difference in the percentage of White and Asian residents between zip codes with 1-sd better schools and zip codes with average schools widened within one year of school performance availability by 0.31% (over the mean value of 77.4) and 0.16 (over the mean value of 1.94), receptively. The gap in college-educated residents increases by a similar magnitude. This divergence is *in addition* to that caused by other factors beyond the impact of rating availability.

[8]Note that the dependent variables such as the percentage of Whites in a zip code are estimates that have a margin of error clearly given in ACS tables. Such measurement error in our dependent variable in OLS does not bias the coefficient estimates but increases the standard error of estimates. We adjust all standard errors in Table 6 for reported margins of error in the ACS tables.

[Table 6 about here.]

Migration: In our final set of models, we examine the effect of rating availability on migration into and out of the zip code. These models, like those above, are estimated using county-year fixed effects.

In Models 6 and 7, we examine the percentage of overall in-migration and the percentage of migration from within the state. We find that rating availability significantly affects migration into a zip code. A zip code with 1-sd above-average schools has overall higher in-migration ($\beta = 0.046\%$) with one year of rating availability, which appears to be largely driven by migrations within a state ($\beta = 0.029\%$).

To summarize, we find evidence that rating availability accelerated the divergence across American communities. Specifically, the gap between zip codes with high- and low-performing schools increased on several critical and related dimensions. First, housing prices began to diverge further, with zip codes containing better schools also having higher priced homes. Second, the ethnic composition of such communities also changed: White and Asian families increasingly moved into these communities, and the proportion of Hispanic residents declined. The change was also economic: Zip codes with the better-performing and more visible schools attracted college-educated residents with higher incomes. All these changes further widened the gap between the zip codes with low- and high-performing schools as identified in prior research.

Differential Access to the Internet and Divergence: As discussed earlier, the segregation in communities we observe, may be the result of two mechanisms: (a) differential access of rating information via unequal access to the internet and (b) differential ability to use the information (e.g., via differences in income or wealth). We attempt to account for the first mechanism by controlling for the internet penetration across different counties. If our main effects persist after controlling for differential access to the internet across communities, this should provide further support for the differential use mechanism.

To test for this mechanism, we collected data from Form 477 on internet penetration (access) at the county level from the FCC website.[9].

The data are coded to provide county-level information on the penetration of residential fixed high-speed internet; 1 for fewer than 200 households; 2 for 200–400 households; 3 for 400–600 households; 4 for 600–800 households; and 5 for more than 800 households having fixed high-speed internet connections per 1,000 households in a county. This data is filed twice (once in the month of June and once in December) every year. We utilize the data filed for the month of December from 2008 to 2015 in our current analysis. The mean value of internet penetration was 3.30 with the standard deviation of 0.90. The internet penetration values for counties at 0, 25, 50, 75, and 100 percentiles of distribution were 0, 3, 3, 4, and 5, respectively. To conduct this test, we estimate the following model in Equation 2. We examine whether inclusion of the $Internet_{ct}$ variable affects the sign, magnitude, or significance of β_3.

$$\begin{aligned} Y_{it} = \beta_1 SP_{i(t-1)} + \beta_2 YearsAvail_{it} + \\ \beta_3(SP_{i(t-1)} \times YearsAvail_{it}) + \beta_4 Internet_{ct} + \alpha_x + \epsilon_{it} \end{aligned} \tag{2}$$

Our results for this estimation, presented in Table 7, suggest that variation in access to the the internet does not significantly affect our primary results—particularly those concerning housing prices and household income. We continue to find a positive and significant coefficient estimate for β_3, the coefficient of the interaction term after controlling for the internet access at the county level. This estimation provides further corroboration for the differential use mechanism.

[Table 7 about here.]

[9]https://www.fcc.gov/general/form-477-county-data-internet-access-services

4.2 Robustness Checks for Alternative Explanations

In this section, we summarize several additional tests we conducted to account for alternative explanations for our findings.

Difference-in-Differences Estimation with DOE Data: A concern with our main specification is that we do not have pre-GS school performance information for a zip code in the models above. To mitigate this concern, we further collected the test scores from the department of education (DOE) website of various states where school performance data from before the entry of GS is available.

Unfortunately, states provide this information in various formats, often spread over many pages on their DOE website. Indeed, many states provide this information only at the school district level and not at the school level. After an extensive search process, we collected school-level standardized tests scores for schools in 26 states and Washington, DC before their availability on the GS website.[10] We summarize this data in Table 8. Overall, we have school performance data for 54,413 schools in 14,581 zip codes.

[Table 8 about here.]

With this data, we estimate the following model:

$$\begin{aligned} Y_{it} = \beta_1 SP_{i(t-1)} + \beta_2 YearsAvail_{it} + \\ \beta_3 (SP_{i(t-1)} \times YearsAvail_{it}) + \alpha_{ct} + \epsilon_{it} \end{aligned} \tag{3}$$

In this model, $YearsAvail_{it}$ is the number of years since the introduction of school scores on the GS website. Therefore, the $YearsAvail_{it}$ variable will be equal to 0 for all years up to the year of its introduction on the GS website. Like our

[10]The states in our sample are AZ, CA, DC, FL, GA, IL, IN, KY, MA, MD, ME, MI, MN, MO, MT, NC, NH, NJ, NY, OH, OR, PA, TN, TX, VA, WA, and WI.

previous models, $SP_{i(t-1)}$ is our measure of standardized school performance for a given zip code–year observation. We again estimate this model using the various dependent variables in Equation 1. β_3 is our coefficient of interest. Table 9 presents these estimations.

[Table 9 about here.]

Even in these models, we find that the coefficient of interest—the interaction between $SP_{i(t-1)}$ and $YearsAvail_{it}$—remains similar in sign and statistical significance for all dependent variables. We further find that even the magnitude of coefficient estimates are also similar to our findings with the full sample of zip codes in Tables 4, 5 and 6.These estimates provide further support for our main finding that the mass availability of online school ratings led to the divergence in house prices and the concentration of high-income households in zip codes with higher performing schools.

[Figure 2 about here.]

In Figure 2 we provide a lead-lag plot of the effect of GS rating availability on housing prices in nominal dollars. Before the availability of online ratings, zip codes with equivalent performing schools differed little in home values. However, after ratings became available for some zip codes, home prices began to diverge. Figures 3a and 3b more clearly show this dynamic. In this graph, we can see that by year three, the difference in home prices for treatment (GS available) and control (GS not available) increases.

[Figure 3 about here.]

Placebo Test: To further address the concerns that our estimated coefficients could be due to a variety of unobserved factors instead of the availability of rating information on GS, we conduct a placebo test. In this test, we take the DOE data prior to the availability of GS, and introduce a placebo treatment of rating availability at a randomly chosen year for different zip codes. If our results are not attributable to the GS rating information, then we may find a significant estimate for the placebo treatments, all else equal. We examine this point with the following model:

$$
\begin{aligned}
Y_{it} = \beta_1 SP_{i(t-1)} + \beta_2 PlaceboGS_{it} + \\
\beta_3(SP_{i(t-1)} \times PlaceboGS_{it}) + \alpha_{ct} + \epsilon_{it}
\end{aligned} \tag{4}
$$

In this model, $PlaceboGS_{it}$ is an indicator variable equal to one if the placebo treatment is assigned in zip code i at time t, and zero otherwise. Like our previous models, $SP_{i(t-1)}$ is our measure of standardized school performance for a given zip code–year observation. We estimate this model for home values and percentage of 00k+ households as dependent variables. β_3 is our coefficient of interest. Table 10 presents these estimations. We find that the coefficient of interest—the interaction between $SP_{i(t-1)}$ and $PlaceboGS_{it}$— is insignificant for both dependent variables, which supports that our estimated results in Tables 4 and 5 are indeed due to the availability of GS rating information.

[Table 10 about here.]

5 Conclusion

Can greater access to online information help to bridge the rising inequality in American society? Using the gradual availability of online school ratings provided by GS, we ask whether the widespread access to quantified school performance information available today has minimized or accelerated this divergence. Across a range of specifications, we find that access to school performance ratings appeared to accelerate, rather than reduce, economic divergence across zip codes in the United

States.

Regarding effect size, we find that the difference in home values between 1-sd better and an average zip codes would diverge from 30.5 percent with no ratings availability to 33.6 percent with three years of rating availability. This significant change in housing prices is also related to economic and demographic divergence across zip codes. In most of our specifications, we find that neighborhoods with lower-performing schools lose high-income and college-educated residents as well as White and Asian residents. We find an asymmetric effect for high-performing zip codes. These results broadly support the thesis that widespread access to quantified school performance information accelerated, rather than minimized, social and economic divergence across American communities.

Our findings speak to several streams of emerging research. First, we propose and test a novel mechanism for the increasing economic divergence across American communities that has been documented by many scholars (Piketty and Saez, 2003; Reardon and Bischoff, 2011). We show that broader access to information increased this divergence because high-income families could more readily leverage school ratings to move to neighborhoods with better schools. Second, our results speak to the growing literature on the social and economic impacts of digitization (DiMaggio and Bonikowski, 2008; DiMaggio et al., 2004). Our research shows that widespread availability of information enabled by the internet can have society-wide, and often unintended, effects. Finally, our results broaden the scope of the emerging research on online rankings by showing how they affect the outcomes of entire communities, not just individual consumers (Luca and Smith, 2013; Pope, 2009).

We also acknowledge several limitations of our approach. First, ours is an observational study that uses the time-varying availability of online ratings across communities. As a result, given that rating availability is not random, our estimates may still have some degree of bias. However, we can account for many possible sources of selection bias in our models using various fixed effects specifications. Furthermore, it is possible that our effect sizes have a potentially conservative bias. That is, if rating availability is related to the ease of access to the data for GS, then it is likely that this school-performance information should have already been priced into homes, as school performance information can be accessed from other sources. Nevertheless, we believe that this issue should still temper the interpretation of our results.

Another limitation of our analysis is that we focus on the effect of rating availability on community characteristics, but do not address the equally important question of how rating availability affects school performance. Prior research has shown that such accountability measures often affect school performance, but only through the types of composition effects we identify in this paper (e.g., Figlio and Lucas, 2004). Therefore, more research needs to be conducted on how parents use this information to influence schools and what rating availability means for individual student outcomes.

Finally, we have conducted our analysis at the zip code level. This approach allowed us to analyze the effect of rating availability on many outcomes at that level of analysis. However, this approach also introduces noise in our estimates because zip codes often, but not always, define the geographic units delineating school boundaries. Moreover, analyzing outcomes at such an aggregate level limits our ability to identify the effect of rating availability on the choices of individual households, and thus our ability to more neatly understand mechanisms.

We hope these results encourage new research on how large-scale access to information and resources through the internet are affecting critical social dynamics (Hargittai and Hinnant, 2008; DiMaggio et al., 2004; Cotten, Anderson and Tufekci, 2009). Research exploring the value of online informational interventions—and how to most effectively design them—has potential to inform policy and practice, especially as more individuals are using the internet to make important decisions about their economic and social well-being.

References

Brasington, David and Donald R Haurin. 2006. "Educational outcomes and house values: a test of the value added approach." *Journal of Regional Science* 46(2):245–268.

Brynjolfsson, Erik and Andrew McAfee. 2014. *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company.

Chevalier, Judith A and Dina Mayzlin. 2006. "The effect of word of mouth on sales: Online book reviews." *Journal of Marketing Research* 43(3):345–354.

Coleman, James S et al. 1966. "Equality of educational opportunity.".

Cotten, Shelia R, William A Anderson and Zeynep Tufekci. 2009. "Old wine in a new technology, or a different type of digital divide?" *New Media & Society* 11(7):1163–1186.

Damianov, Damian S and Diego Escobari. 2016. "Long-run equilibrium shift and short-run dynamics of US home price tiers during the housing bubble." *Journal of Real Estate Finance and Economics* 53(1):1–28.

DiMaggio, Paul and Bart Bonikowski. 2008. "Make money surfing the web? The impact of Internet use on the earnings of US workers." *American Sociological Review* 73(2):227–250.

DiMaggio, Paul, Eszter Hargittai, Coral Celeste and Steven Shafer. 2004. "From unequal access to differentiated use: A literature review and agenda for research on digital inequality." *Social Inequality* pp. 355–400.

Dorsey, Robert E, Haixin Hu, Walter J Mayer and Hui-chen Wang. 2010. "Hedonic versus repeat-sales housing price indexes for measuring the recent boom-bust cycle." *Journal of Housing Economics* 19(2):75–93.

Espeland, Wendy Nelson and Michael Sauder. 2007. "Rankings and reactivity: How public measures recreate social worlds." *American Journal of Sociology* 113(1):1–40.

Espeland, Wendy Nelson and Mitchell L Stevens. 1998. "Commensuration as a social process." *Annual Review of Sociology* 24(1):313–343.

Fack, Gabrielle and Julien Grenet. 2010. "When do better schools raise housing prices? Evidence from Paris public and private schools." *Journal of Public Economics* 94(1-2):59–77.

Figlio, David N and Cecilia Elena Rouse. 2006. "Do accountability and voucher threats improve low-performing schools?" *Journal of Public Economics* 90(1-2):239–255.

Figlio, David N and Maurice E Lucas. 2004. "What's in a grade? School report cards and the housing market." *American Economic Review* 94(3):591–604.

Gibbons, Stephen, Stephen Machin and Olmo Silva. 2013. "Valuing school quality using boundary discontinuities." *Journal of Urban Economics* 75:15–28.

Guerrieri, Veronica, Daniel Hartley and Erik Hurst. 2013. "Endogenous gentrification and housing price dynamics." *Journal of Public Economics* 100:45–60.

Hargittai, Eszter and Amanda Hinnant. 2008. "Digital inequality: Differences in young adults' use of the Internet." *Communication Research* 35(5):602–621.

Harris, Douglas N and Carolyn D Herrington. 2006. "Accountability, standards, and the growing achievement gap: Lessons from the past half-century." *American Journal of Education* 112(2):209–238.

Hoxby, Caroline and Sarah Turner. 2013. "Expanding college opportunities for high-achieving, low-income students." *Stanford Institute for Economic Policy Research Discussion Paper* (12-014).

Jargowsky, PA. 1996. "Take the money and run: Economic segregation in US metropolitan areas." *American Sociological Review* 61(6):984–98.

Jensen, Robert. 2010. "The (perceived) returns to education and the demand for schooling." *Quarterly Journal of Economics* 125(2):515–548.

Jorgenson, Dale W. 2001. "Information technology and the US economy." *American Economic Review* 91(1):1–32.

Kane, Thomas J, Stephanie K Riegg and Douglas O Staiger. 2006. "School quality, neighborhoods, and housing prices." *American Law and Economics Review* 8(2):183–212.

Logan, John R. 2011. "Separate and unequal: The neighborhood gap for blacks, Hispanics and Asians in Metropolitan America." *Project US2010 Report* pp. 1–22.

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

Mayer, Susan E. 2002. "How economic segregation affects children's educational attainment." *Social Forces* 81(1):153–176.

Mikulecky, Marga and Kathy Christie. 2014. "Rating states, grading schools: What parents and experts say states should consider to make school accountability systems meaningful." *Education Commission of the States* .

Nguyen-Hoang, Phuong and John Yinger. 2011. "The capitalization of school quality into house values: A review." *Journal of Housing Economics* 20(1):30–48.

Norris, Pippa et al. 2001. *Digital divide: Civic engagement, information poverty, and the Internet worldwide.* Cambridge University Press.

Owens, Ann, Sean F Reardon and Christopher Jencks. 2016. "Income segregation between schools and school districts." *American Educational Research Journal* 53(4):1159–1197.

Piketty, Thomas and Emmanuel Saez. 2003. "Income inequality in the United States, 1913–1998." *Quarterly Journal of Economics* 118(1):1–41.

Piketty, Thomas and Emmanuel Saez. 2006. "The evolution of top incomes: a historical and international perspective." *American Economic Review* 96(2):200–205.

Pope, Devin G. 2009. "Reacting to rankings: evidence from "America's Best Hospitals"." *Journal of Health Economics* 28(6):1154–1165.

Quillian, Lincoln. 2014. "Does segregation create winners and losers? Residential segregation and inequality in educational attainment." *Social Problems* 61(3):402–426.

Reardon, Sean F. 2011. "The widening academic achievement gap between the rich and the poor: New evidence and possible explanations." *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances* pp. 91–116.

Reardon, Sean F and Kendra Bischoff. 2011. "Income inequality and income segregation." *American Journal of Sociology* 116(4):1092–1153.

Reardon, Sean F, Lindsay Fox and Joseph Townsend. 2015. "Neighborhood income composition by household race and income, 1990–2009." *Annals of the American Academy of Political and Social Science* 660(1):78–97.

Reeves, Richard V. 2017. *Dream hoarders: How the American upper middle class is leaving everyone else in the dust, why that is a problem, and what to do about it.* Brookings Institution Press.

Rose, Stephen. 2016. "The growing size and incomes of the upper middle class." *Urban Institute* .

Rumberger, Russell W and Gregory J Palardy. 2005. "Does segregation still matter? The impact of student composition on academic achievement in high school." *Teachers College Record* 107(9):1999.

Salganik, Matthew J and Duncan J Watts. 2008. "Leading the herd astray: An experimental study of self-fulfilling prophecies in an artificial cultural market." *Social Psychology Quarterly* 71(4):338–355.

Salganik, Matthew J, Peter Sheridan Dodds and Duncan J Watts. 2006. "Experimental study of inequality and unpredictability in an artificial cultural market." *Science* 311(5762):854–856.

Sampson, Robert J, Patrick Sharkey and Stephen W Raudenbush. 2008. "Durable effects of concentrated disadvantage on verbal ability among African-American children." *Proceedings of the National Academy of Sciences* 105(3):845–852.

Santos, Rita, Hugh Gravelle and Carol Propper. 2017. "Does quality affect patients' choice of doctor? Evidence from England." *Economic Journal* 127(600):445–494.

Sauder, Michael and Wendy Nelson Espeland. 2009. "The discipline of rankings: Tight coupling and organizational change." *American Sociological Review* 74(1):63–82.

Shore, Cris, Susan Wright, Vered Amit, Judy Brown, Casper Bruun Jensen, Mark Maguire, Sally Engle Merry, Michael Sauder, Wendy Espeland, Gavin Smith et al. 2015. "Audit culture revisited: Rankings, ratings, and the reassembling of society." *Current Anthropology* 56(3):431–432.

Varkevisser, Marco, Stephanie A van der Geest and Frederik T Schut. 2012. "Do patients choose hospitals with high quality ratings? Empirical evidence from the market for angioplasty in the Netherlands." *Journal of Health Economics* 31(2):371–378.

Warschauer, Mark. 2004. *Technology and social inclusion: Rethinking the digital divide*. MIT press.

Ziol-Guest, Kathleen M and Kenneth TH Lee. 2016. "Parent income–based gaps in schooling: Cross-cohort trends in the NLSYs and the PSID." *AERA Open* 2(2):2332858416645834.

Figure 1: Example of GreatSchools ratings on main website (above) and through the Zillow website (below).

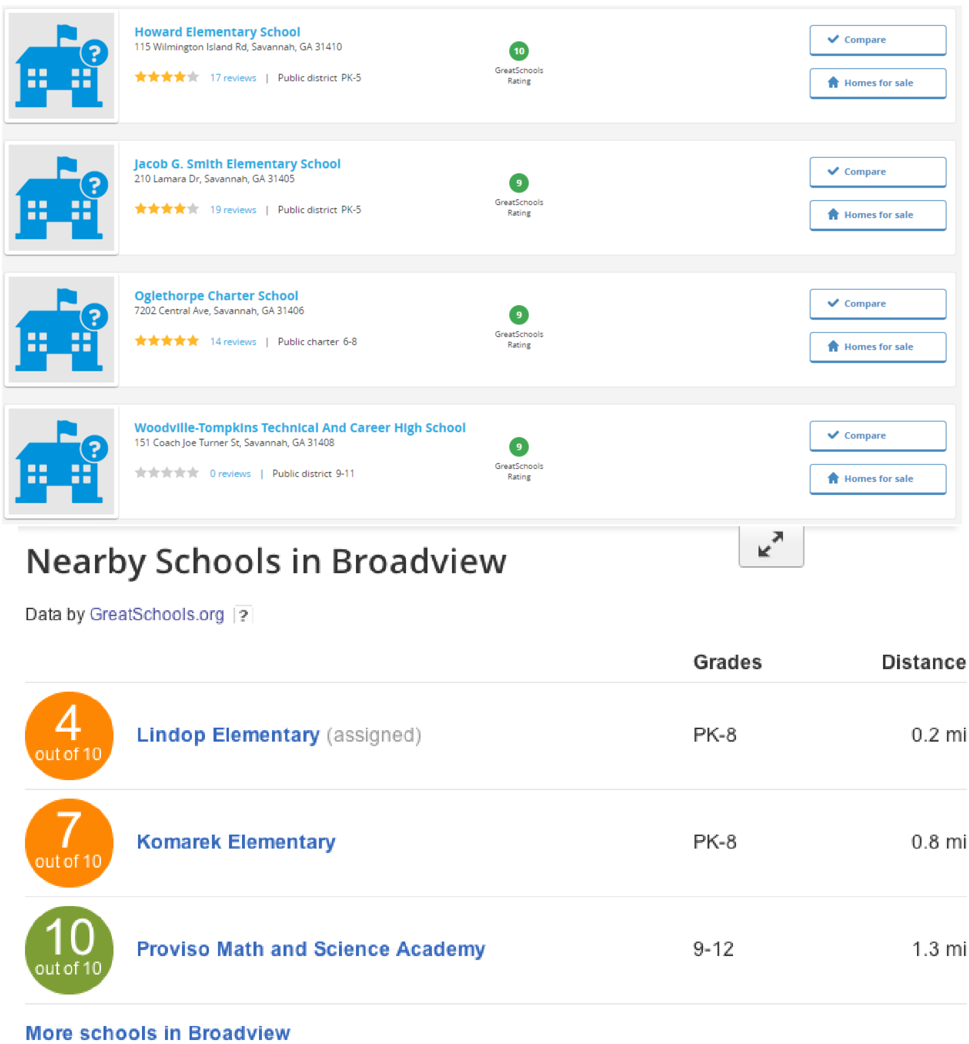
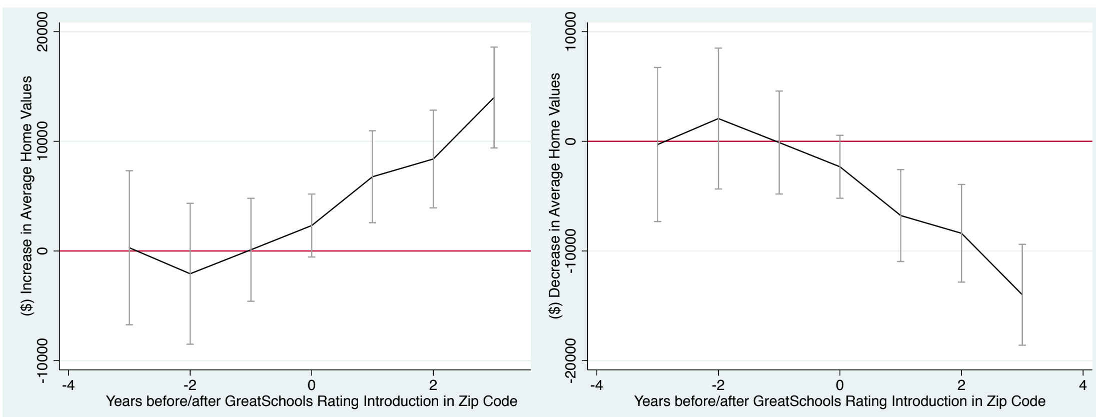


Figure 2: Plot of the difference in housing values between zip codes whose schools do/do not have GreatSchools ratings.[11]



(a) Effect of GreatSchools rating availability for above median schools

(b) Effect of GreatSchools rating availability for below median schools

Figure 3: Plot of the difference in housing values for zip codes with and without GreatSchools ratings for zip codes above- and below-median test scores.

Year	Schools	zip codes	States
2006	20298	4643	5
2007	20637	4658	5
2008	24763	5819	7
2009	25830	6151	8
2010	31548	7908	14
2011	41741	10175	22
2012	73740	20423	49

Table 1: Coverage of Greatschools.org data from 2006 to 2012.

Table 2: Weibull estimates examining the relationship between zip code characteristics and the timing of GS rating availability.

	(1)
% 100k+ Households	-0.014
	(0.013)
Log(ZHVI+1)	0.152
	(0.317)
% White	0.023
	(0.018)
% Black	0.028
	(0.021)
% Hispanic	0.022
	(0.018)
% College Degree	0.007
	(0.005)
Constant	-21.158***
	(2.459)
Observations	8435
R^2	

Standard errors in parentheses
* $p < .1$, ** $p < .05$, *** $p < .01$

Table 3: Summary statistics for main variables used in our analysis.

	count	mean	sd	min	max
School Performance	126282	0.025	0.812	-7.335	4.610
Years Available	147787	3.376	2.691	0.000	10.000
Log(ZHVI+1)	165549	12.048	0.658	9.518	15.510
$ ZHVI	165549	217842.590	198219.159	13600.000	5442900.000
% 100k+ Households	279720	13.621	10.358	0.000	84.507
% White	231840	77.370	26.350	0.000	100.000
% Black	231840	7.402	15.748	0.000	100.000
% Hispanic	231840	8.800	16.377	0.000	100.000
% Asian	231840	1.937	5.161	0.000	100.000
% College Degree	195734	30.468	16.805	0.000	100.000
% Migriation	195961	12.077	9.558	0.000	100.000
% In-State Migration	195961	9.916	7.977	0.000	100.000
% No Migration	195961	87.555	9.880	0.000	100.000
Observations	349041				

Table 4: The effect of school rating availability on the relationship between school performance and housing prices.

	(1)	(2)	(3)	(4)
School Performance	0.300***	-0.031***	0.300***	0.266***
	(0.009)	(0.003)	(0.016)	(0.007)
Years Since	0.012***	-0.010***	-0.002	-0.020***
	(0.001)	(0.001)	(0.009)	(0.005)
School Performance * Years Available	0.014***	0.008***	0.013***	0.008***
	(0.002)	(0.001)	(0.004)	(0.002)
Constant	11.940***	12.182***	11.996***	12.076***
	(0.006)	(0.003)	(0.034)	(0.021)
Observations	75454	75454	75454	75454
R^2	0.155	0.986	0.567	0.837
Zip FE	No	Yes	No	No
Year FE	No	Yes	No	No
State-Year FE	No	No	Yes	No
County-Year FE	No	No	No	Yes

Standard errors are cluster corrected at the level of the geographic and time fixed effects.
* $p < .1$, ** $p < .05$, *** $p < .01$

Table 5: The effect of school rating availability on the relationship between school performance and % of $100,000 or higher income households in zip code.

	(1)	(2)	(3)	(4)
School Performance	5.754***	-0.048	5.548***	5.700***
	(0.112)	(0.035)	(0.477)	(0.228)
Years Since	0.065***	-0.339***	0.169	-0.151
	(0.017)	(0.006)	(0.138)	(0.108)
School Performance * Years Available	0.219***	0.054***	0.237**	0.201***
	(0.030)	(0.007)	(0.119)	(0.060)
Constant	13.572***	17.317***	13.220***	14.325***
	(0.068)	(0.033)	(0.479)	(0.372)
Observations	93994	93994	93994	93994
R^2	0.247	0.987	0.399	0.624
Zip FE	No	Yes	No	No
Year FE	No	Yes	No	No
State-Year FE	No	No	Yes	No
County-Year FE	No	No	No	Yes

Standard errors are cluster corrected at the level of the geographic and time fixed effects.
* $p < .05$, ** $p < .01$, *** $p < .001$

Table 6: The effect of school rating availability on the relationship between school performance and zip code demographics and migration patterns.

	% White	% Black	% Hisp	% Asian	% College	% In Migr.	% State Migr.
School Performance	9.449***	-6.781***	-1.435***	0.190**	6.251***	-1.295***	-1.364***
	(0.402)	(0.396)	(0.296)	(0.097)	(0.263)	(0.102)	(0.085)
Years Since	-0.820***	0.133	0.589***	0.099**	0.099	0.106	0.216***
	(0.179)	(0.138)	(0.143)	(0.042)	(0.135)	(0.083)	(0.069)
School Performance * Years Available	0.305***	0.059	-0.697***	0.162***	0.422***	0.046**	0.029*
	(0.090)	(0.093)	(0.078)	(0.027)	(0.066)	(0.019)	(0.016)
Constant	76.173***	8.656***	9.335***	2.309***	32.047***	12.293***	9.841***
	(0.743)	(0.574)	(0.594)	(0.174)	(0.568)	(0.349)	(0.292)
Observations	100860	100860	100860	100860	94371	94378	94378
R^2	0.741	0.636	0.723	0.558	0.617	0.320	0.329
County-year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses
* $p < .05$, ** $p < .01$, *** $p < .001$

Table 7: The effect of school rating availability on the relationship between school performance and primary outcomes, accounting for internet availability by county. (1) Log(ZHVI+1); (2) % 100k+; (3) % White; (4) % Black; (5) % Hisp; (6) % Asian; (7) % College;(8) % In Migr.; (9) % State Migr.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
School Performance	0.257***	4.958***	10.243***	-7.207***	-1.351***	0.184	5.783***	-1.216***	-1.216***	-1.254***
	(0.010)	(0.235)	(0.632)	(0.602)	(0.357)	(0.114)	(0.397)	(0.159)	(0.159)	(0.140)
Years Since	-0.023***	0.032	-0.303**	-0.160*	0.551***	0.007	-0.055	-0.096***	-0.096***	-0.019
	(0.004)	(0.042)	(0.131)	(0.093)	(0.086)	(0.038)	(0.092)	(0.029)	(0.029)	(0.026)
School Performance * Years Available	0.008**	0.222***	0.130	0.144	-0.662***	0.147***	0.351***	0.029	0.029	0.012
	(0.003)	(0.072)	(0.183)	(0.178)	(0.132)	(0.045)	(0.127)	(0.032)	(0.032)	(0.028)
Broadband availability	0.155***	1.925***	-0.720	-0.471	0.434	0.589***	4.853***	0.647***	0.647***	0.307***
	(0.017)	(0.208)	(0.560)	(0.413)	(0.366)	(0.094)	(0.338)	(0.114)	(0.114)	(0.101)
Constant	11.422***	6.037***	76.693***	11.698***	8.133***	0.354	13.160***	10.576***	10.576***	9.642***
	(0.060)	(0.736)	(2.004)	(1.482)	(1.297)	(0.397)	(1.209)	(0.418)	(0.418)	(0.374)
Observations	60296	89628	82692	82692	82692	82692	76190	76196	76196	76196
R^2	0.745	0.512	0.622	0.464	0.640	0.513	0.520	0.219	0.219	0.225

Standard errors in parentheses
* $p < .1$, ** $p < .05$, *** $p < .01$

Table 8: Description for State Department of Education Data Prior to GreatSchools introduction.

State	First year	Zip codes	Schools	GS year
AZ	2010	337	1809	2012
CA	2001	1225	8110	2006
DC	2007	19	113	2011
FL	2005	796	3098	2012
GA	2004	471	1613	2011
IL	2001	862	3035	2006
IN	2006	508	1442	2012
KY	2007	393	1135	2009
MA	2008	427	1797	2012
MD	2003	306	1158	2012
ME	2007	292	464	2010
MI	2002	687	2219	2008
MN	2005	548	1843	2010
MO	2010	638	2006	2012
MT	2007	226	734	2011
NC	2002	594	1914	2006
NH	2009	129	222	2011
NJ	2003	514	1776	2012
NY	2007	1039	3260	2011
OH	2006	830	2773	2012
OR	2005	1	1	2006
PA	2006	896	2752	2012
TN	2010	414	1124	2012
TX	2003	1441	6015	2006
WA	2006	452	2046	2013
WI	2006	536	1815	2012

Table 9: The effect of school rating availability on the relationship between school performance and socio-economic variables in zip code. Estimates use data from 26 states and Washington DC for which we have school performance information prior to GreatSchools rating availability. All models include county-year fixed effects and standard errors cluster corrected at the county-year level. (1) Log(ZHVI+1); (2) % 100k+; (3) % White; (4) % Black; (5) % Hisp; (6) % Asian; (7) % College;(8) % In Migr.; (9) % State Migr.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
School Performance	0.210***	5.421***	9.645***	-6.810***	-2.266***	0.402***	6.921***	-1.533***	-1.597***
	(0.004)	(0.106)	(0.278)	(0.285)	(0.191)	(0.060)	(0.194)	(0.082)	(0.069)
Years Rating Present	-0.037***	-0.524***	-1.643***	0.656**	1.078***	0.105*	-0.533**	0.114	0.210**
	(0.007)	(0.164)	(0.277)	(0.261)	(0.239)	(0.055)	(0.230)	(0.115)	(0.098)
School Performance * Years Available	0.021***	0.349***	0.507***	-0.135	-0.673***	0.173***	0.425***	0.045**	0.028
	(0.002)	(0.040)	(0.094)	(0.098)	(0.073)	(0.026)	(0.068)	(0.020)	(0.017)
Constant	12.112***	15.823***	78.908***	6.948***	8.361***	2.585***	34.405***	12.310***	10.154***
	(0.012)	(0.327)	(0.800)	(0.756)	(0.686)	(0.157)	(0.726)	(0.364)	(0.309)
Observations	100377	126082	81406	81406	81406	81406	68667	68667	68667
R^2	0.843	0.652	0.737	0.580	0.717	0.535	0.615	0.332	0.345

Standard errors in parentheses
* $p < .1$, ** $p < .05$, *** $p < .01$

Table 10: Placebo test using randomly generated GS availability year.

	(1)	(2)
School Performance	0.192***	4.993***
	(0.008)	(0.194)
Placebo Rating Present	-0.003	-0.246**
	(0.005)	(0.122)
School Quality * Placebo Rating	0.006	0.186
	(0.008)	(0.188)
Constant	12.092***	15.636***
	(0.004)	(0.112)
Observations	49761	56730
R^2	0.850	0.676

Standard errors in parentheses
* $p < .1$, ** $p < .05$, *** $p < .01$