Sorting schools:

A computational analysis of charter school identities and stratification

By Jaren Haber
Dept. of Sociology
University of California, Berkeley

Date: March 9, 2020

Forthcoming in *Sociology of Education* (an ASA journal)

Abstract

Research shows charter schools are more segregated by race and class than traditional public schools. I investigate an under-examined mechanism for this segregation: Charter schools project identities corresponding to parents' race- and class-specific parenting styles and educational values. I use computational text analysis to detect the emphasis on inquiry-based learning in the websites of all charter schools operating in 2015-16. I then estimate mixed linear regression models to test the relationships between ideological emphasis and school- and district-level poverty and ethnicity. I thereby transcend methodological problems in scholarship on charter school identities by collecting contemporary, population-wide data, and by blending text analysis with hypothesis testing. Findings suggest charter school identities are both race- and class-specific, outlining a new mechanism by which school choice may consolidate parents by race and class—and paving the way for behavioral and longitudinal studies. This project contributes to literatures on school choice and educational stratification.

Copyright 2020. Jaren Haber. All rights reserved.
This paper is for the reader's personal use only.

ACKNOWLEDGMENTS

I am especially grateful to Heather Haveman for her constructive criticism, which after many drafts have greatly improved this paper. I also give thanks to the UC Berkeley Data-Intensive Social Science Lab (D-Lab) community for teaching me to code and to embrace not knowing. I acknowledge also Sam Lucas, Calvin Morrill, Bruce Fuller, David Bamman, Ben Gebre-Medhin, and Caroline Le Pennec-Caldichoury for their feedback and insightful comments; Aaron Culich, Carl Mason, and the Cloud Working Group for help with web data collection and computing infrastructure; and my family and wife for their encouragement and support. This complex project wouldn't have been possible without the contributions of 38 research assistants from the Undergraduate Research Apprentice Program and Data Science Discovery Program: Kanika Ahluwalia, Brad Afzali, Akcan Balkir, Muying Chen, Siyuan Chen, Yitong Chen, Kaan Dogusoy, Saabhir Gill, Harshayu Girase, Akshat Gokhale, Yoon Sung Hong, Jennifer Huang, Elaine Huynh, Krutika Ingale, Jiyoon Jeong, James Jung, Inderpal Kaur, Francis Kumar, Yong Jin Kweon, Ariel Langer, Brian Yimin Lei, Xueyong Liu, Haley Miller, Anna Nguyen, Thao Nguyen, Tina Nguyen, Madeleine Peng, Emily Qian, Samyukta Raman, Ji Shi, Sarah Solieman, Arjun Srinivasan, Prianka Subrahmanyam, Frank Wang, Violet Yao, George Wu, Max Yuan, and Jiahua Zou.

Previous versions of this paper and its methods were presented at the Berkeley Institute for Data Science's 2018 Text Across Domains (TextXD) symposium; the D-Lab's Computational Text Analysis Working Group in 2017-18; the 2018 Making Text Research-Ready symposium; the 2018 Graduate School of Education Research Day; and the American Sociology Association's Sociology of Education Section in 2017 and 2018. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National

Science Foundation grant number ACI-1548562, as well as the Berkeley Demography Lab cloud computing facility. Financial support was provided by the UC Berkeley Dissertation Completion Fellowship and the Bridge Lowenthal Fellowship. I declare no conflict of interest in doing this research.

As the bastion of school choice, the educational reform of charter schools has won bipartisan support and grown tremendously in recent decades (Berends 2015). Charter schools are publicly funded yet free from many regulations (around labor, content, finances, etc.) facing traditional public schools (e.g., Orfield, Gumus-Dawes, and Luce 2013), granting them autonomy to innovate in pursuit of greater performance (e.g., Lubienski 2003). Recent years have seen charter schools more than triple in number and a nearly nine-fold increase in the number of students served, climbing from 1,542 schools with 349,714 students in 1999-2000 to reach 6,992 schools serving 3,008,106 students in 2016-17 (National Alliance for Public Charter Schools 2018). Thus, in under two decades, charter schools have grown from 0.74% of the U.S. K-12 student population in the fall of 1999 to 5.95% in fall 2016 (National Center for Education Statistics 2018b). Given the mounting prevalence of charter schools in U.S. society, this project examines the relationships between charter schools and their local communities.

All schools depend for survival on local funding and pupils, but as schools of choice, this is especially true for charter schools. Charter schools must be sufficiently attractive that local parents opt not for their neighborhood traditional public school but instead for a charter school. Yet presenting an appealing picture to parents is no simple matter—it depends on local parents' positions in broader social hierarchies of socioeconomic status and race, which structure their goals and interests. That is, what parents want for their children depends on cultural and social factors linked to their social positions and resources: parents and schools prioritize different habits and skills for different social strata (e.g., Bowles and Gintis 1976; Lareau 2011), and parents also want their children to be around peers "of quality" (e.g., Abdulkadiroglu et al. 2017; Rothstein 2006) and socially similar to them in dimensions like race and social status (e.g., Holme 2002; Roda and Wells 2013). Thus, each charter school must present a strategic *identity*—a cultural conception projected to external audiences, namely parents, teachers, and oversight agencies—that resonates with local parents' socially embedded inclinations. Given their relative lack of direct district control and need to carefully cultivate a locally attractive image in order to secure resources, charter schools offer an opportunity to analyze how schools signal identities to navigate local instantiations of broader social hierarchies.

I investigate the role that distinct organizational forms play in parents' self-sorting into segregated schools and schools' self-sorting into segregated neighborhoods. I argue that this self-sorting is driven partly by alignment between charter schools' projected identities and parents' race- and class-specific parenting styles and educational values. Focusing on the 'supply' side of this dynamic, I propose that charter schools present different identities to communities that are advantaged—relatively affluent and white—compared to those that are disadvantaged—relatively poor or people of color. To test this proposition, I analyze an educational *ideology* (a set of beliefs about the social world that motivates moral action; Pettigrew 1979:575) common in schools: *inquiry-based learning* (IBL).

As a central ideology of progressive-style educational programs (e.g., Dewey 1938), IBL entails student observation of phenomena, first-hand scientific inference, and student-centered construction of knowledge (e.g., Bruner 1961; Steffe and Gale 1995). IBL is widespread in charter as well as traditional public schools (Waite 2019) and has inspired broad policy initiatives including the Common Core (e.g., Watanabe 2007) and Next Generation Science Standards (Achieve 2010), efforts fueled by lively scholarship in the learning sciences.[1] In practice, IBL-based approaches favor the privileged: Arts; Classical; International; Science, Technology, Engineering, and Math (STEM); or Progressive models generally enroll more white students and fewer low-income students or students of color than do nearby traditional public schools (McShane and Hatfield 2015; Waite 2019). This tendency of IBL to segregate social groups makes it an especially sharp tool for studying how ideologies separate race and class factions among charter schools.

But little research has explored how IBL or other educational ideologies are invoked in race- and class-differentiated settings (but see McShane and Hatfield 2015). I fill this gap through fine-grained, socially embedded analysis of educational discourse. Specifically, I address the research question: What is the relationship between a charter school's race and class composition and its emphasis on IBL? I predict that charter schools present themselves to affluent and white communities in ways emphasizing IBL.

To operationalize this ideology and study its connections with race and class distributions in charter schools, I capitalize on new computational tools for measuring culture inductively (e.g., Bail 2014; Nelson 2017) alongside the deductive method of mixed-effects linear regression. Moreover, I redress methodological and theoretical oversights in the school choice literature by collecting contemporary, valid, and population-wide charter school data and analyzing it with complementary text-analytic and statistical methods.

Specifically, I gathered rich, nationally comprehensive data on identities for all U.S. charter schools drawn from their websites. Websites are the collective hubs of many modern forms of organization, enabling collective action, information sharing, and new forms of interaction (e.g., Bennett and Segerberg 2013). Moreover, websites are rich, ecologically valid sources of cultural information targeting multiple audiences and revealing the organization's identity and goals (Powell, Horvath, and Brandtner 2016). As such, parents commonly rely on websites for signals on school quality and "fit" with their children—and in some cases, website information influences parents' perceptions of school quality more than do test scores (Yettick 2016).

In the rest of this paper, I situate my study within the literatures on organizational identities, school choice, inequality, and parents' educational preferences. I then offer my hypothesis, outline methods, and discuss results and implications.

THEORY

Charter school identities and stratification

As organizations, the survival of charter schools depends on attaining both legitimacy—that is, inclusion in established categories with proven results—on one hand, and uniqueness or innovation relative to peers, on the other. As such, charter schools seek to satisfy expectations for both uniformity—specifically, by clustering into recognizable identities (King, Clemens, and Fry 2011) or generalizing the language of their mission statements (Renzulli, Barr, and Paino 2015)—and uniqueness—by recombining standard elements (King et al. 2011) or narrowing the focus of their mission statements (Renzulli et al. 2015). Indeed, to fit a sociodemographic niche, charter schools not only deploy ritualized symbols of effectiveness like academic proficiency scores (Meyer and Rowan 1977), but also their identity claims reflect the social backgrounds of those to whom they wish to appeal (e.g., Fuller 2009; Lauen, Fuller, and Dauter 2015). Specifically, charter schools' identities correspond to the educational ideologies and parenting styles of race- and class-specific audiences—a mechanism for segregation by race and class and the focus of this study.

School choice and inequality

Impact on racial and socioeconomic divisions is particularly concerning for charter schools, which were originally conceived (Shanker 1988) and are continually justified (e.g., Roth et al. 2017) as a bottom-up organizational means to equalize educational opportunity for ethnic minorities and the poor. Thus, concerns for inequality have driven research into charter school practices and outcomes, examining whether they select high-achieving students (e.g., Lacireno-Paquet et al. 2002), negatively affect resources and outcomes in nearby school districts (e.g., Preston et al. 2012), restrict access based on information and social connections (e.g., Yettick 2016), increase educational instability for the under-privileged (e.g., Paino, Boylan, and Renzulli 2017), or sort disadvantaged students into less effective programs (e.g., Golann 2015).

In principle, publicly funded charter schools accept strict accountability to content and performance standards enforced by high-stakes assessments (e.g., Ladd 1996) in exchange (e.g., Paino et al. 2017) for market-like, innovation-boosting autonomy from the labor regulations, content restrictions, and financial oversight that traditional public schools face. According to school-choice advocates (e.g., Chubb and Moe 1990), the regulatory independence of schools of choice sidesteps the pedagogical and organizational constraints on traditional public schools, thereby improving student outcomes by promoting organizational competition and a range of parental options. Thus, performance and innovation are the twin objectives of charter schools; to survive, they must balance these goals.

However, current research does not affirm that charter schools have achieved their mission of effective and innovative education—which is especially important for the poor students and students of color increasingly enrolling in them (Berends 2015; Wang, Rathbun, and Musu 2019). Studies show that charter schools' academic performance is heterogeneous and not consistently superior to that of comparable conventional public schools (Berends 2015; Wang et al. 2019). In addition, both charter and local public schools share the same set of "innovative" practices at the classroom (e.g., individualized instruction, cooperative learning) and organizational levels (e.g., small class sizes, teacher merit pay; Lubienski 2003), and the same characteristics that raise achievement in charters are also effective for their local counterparts (e.g., increased instructional time, high academic and behavioral expectations, teacher coaching, use of data; Berends 2015; Furgeson et al. 2012; Gleason 2017). Furthermore, when compared to traditional public schools in their vicinity, charters are generally not the sole adopters of administrative innovations in the areas of academic support services (e.g., after-school tutoring), staffing policies (e.g., merit pay), organizational structures (e.g., block scheduling), and governance practices (teacher/parent influence on hiring; Preston et al. 2012). In sum, neither performance nor discrete innovations in pedagogy or organization appear to differentiate charter schools from traditional public schools.

Furthermore, as schools of choice, charter schools are especially vulnerable to parents' self-sorting by race and class—a key social mechanism of segregation. This is well-evidenced by studies documenting high-status groups like whites escaping to schools of choice to avoid integration with low-status groups like people of color (e.g., Renzulli and Evans 2005; Saporito 2003). This has worked against integration efforts for decades (for a review, see Reardon and Owens 2014)—as well as people of color opting for schools where they are over-represented (e.g., Frankenberg et al. 2019). Also evident is parents' tendency to select schools with peers similar to their own children in race and class (e.g., Holme 2002; Roda and Wells 2013).

As a result of these trends, charter schools tend to have student bodies that are more homogeneous by race and class than traditional public schools (Malkus 2016; Monarrez, Kisida, and Chingos 2019). In 2014-15, 17% of charter schools—compared with 4.5% of traditional public schools—were "racially isolated" (had enrollments that were at least 99% students of color; Moreno 2017). Such segregation poses several risks: It may deepen inequalities by excluding disadvantaged students from the resources and benefits of integrated schools (e.g., Frankenberg et al. 2019; Hanushek, Kain, and Rivkin 2009), isolate youths from civic engagement opportunities (e.g., Levinson 2012), and lead to ethnic fragmentation, undermining both the "common schools" ideal and democracy itself (e.g., Asante and Ravitch 1991). Thus, it is especially important to understand the mechanisms of educational segregation by race and class in charter schools, the leading edge in a growing trend (e.g., Owens, Reardon, and Jencks 2016; Reardon and Owens 2014).

Inequality and parenting styles

Research has discovered two primary, class-distinct parenting styles (Lareau 2000, 2011): the middle-class approach, 'concerted cultivation', is characterized by development of individual talents and rich vocabulary, packed schedules, and parental intervention in schooling. In contrast, the poor and working-class approach, 'natural growth', is characterized by emphasis on meeting basic needs, strict discipline, parent/child separation, social free time, and parental deference to school personnel. Children raised in the middle-class, concerted cultivation style are advantaged in turning interactions to their interests, winning accommodation from authorities, and navigating institutions—such as getting extra help from teachers (Calarco 2011) or custom treatment by doctors (Lareau 2011). This is evidence not of the superior quality of concerted cultivation, but that it provides a means for privileged parents to secure their children's futures by developing their "cultural capital"—those habits, skills, and knowledge that impart cultural and educational distinction through alignment with dominant class tastes, styles, and institutions (Bourdieu 1977; Lareau 2011).

In sum, my theoretical account suggests that parents' self-sorting by race and class into charter schools reflects their attraction to particular charter school identities (e.g., Lauen et al. 2015; McShane and Hatfield 2015) and racially hued perceptions of school quality (e.g., Holme 2002; Roda and Wells 2013), which in turn are driven by class-specific parenting styles—

specifically, 'concerted cultivation' and 'natural growth' (Lareau 2000, 2011). These values and styles (the 'demand' side) are an under-explored influence on charter school segregation, and no prior research has analyzed how these correspond with charter schools' identities (the 'supply' side). This study is an important first step in documenting this theory, by providing meso-level, organizational evidence of the linkage between educational ideology and sociodemographic factors.

Hypothesis

IBL shares with the white, middle-class 'concerted cultivation' style a focus on individual skills and capacities, especially critical thinking; questioning of authority and outside knowledge; and strategic—rather than purely directive—adult guidance of child-centered activities. As such, I predict that:

> Hypothesis: Emphasis on IBL is negatively associated with charter school enrollments of (a) low-income students or (b) students of color.

In order to account for alternative influences on IBL and charter school enrollments, I build two sets of models: one predicting IBL emphasis, and the other predicting proportions of low-income students and (separately) students of color.

Regarding the first set of models, the social context within which charter schools seek to secure resources is not confined to those already enrolled in the school. They must appeal also to potential 'clients' within the school district—the political and administrative arena in which policies are enacted (e.g., Finnigan 2007), parents and others exert influence (e.g., Preston et al. 2012), and schools compete for students and favor (e.g., Arum 1996). Indeed, just as sociodemographic factors drive parents to sort themselves into schools, so do they drive schools to sort themselves into districts—an influence all the more pronounced given persistent residential segregation by race and income (e.g., Massey, Rothwell, and Domina 2009; Reardon, Townsend, and Fox 2017). Thus, social context in the school district—rather than the school—represents an alternative mechanism for my hypothesis. Accordingly, I analyze the relationships between IBL and (as independent variables) both school and school district socioeconomic and ethnic composition.

Regarding the second set of models, school choice scholars often assume that parents respond rationally to their educational options, choosing the highest quality schools—that is, those that perform best on standardized tests—available for their children (e.g., Epple, Figlio, and Romano 2004; Hanushek et al. 2007). Moreover, academic quality is generally reported at the top of parents' educational values—with school safety, extracurriculars, and moral instruction taking precedence only in select circumstances (for reviews, see Erickson 2017; Posey-Maddox, Kimelberg, and Cucchiara 2014). But I argue that parents' school choices are not driven solely by objective signals of academic quality; rather, educational ideology plays a role in shaping enrollment patterns. To support this claim, I analyze the relationships between school socioeconomic and ethnic composition and (as independent variables) IBL and academic quality.

Recent empirical research supports this challenge to the primacy of academic quality. Scholars may have exaggerated the influence of objective academic quality on parents' school preferences, which are better explained by peer quality (the performance of the existing student body; Abdulkadiroglu et al. 2017; Rothstein 2006). Indeed, much research shows that "school quality" is not an objective signal, but rather is socially constructed in ways that reflect hierarchies of race and class. Even high-status parents—who may possess better information on school composition, achievement, etc. than do low-status parents (e.g., Teske, Fitzpatrick, and Kaplan 2006; Yettick 2016)—assess school quality through their social networks, rather than relying on objective test score data or first-hand observation (Holme 2002; Roda and Wells 2013).

RESEARCH METHODS

Previous attempts to classify charter school identities have relied on hand-coding limited samples (e.g., McShane and Hatfield 2015; Renzulli et al. 2015), resulting in several incongruent categorizations—e.g., a set of 13 categories such as No Excuses, international, and arts (McShane and Hatfield 2015) versus a very different set of 11 categories such as values, homeschool, and special education (Renzulli et al. 2015). Prior research (e.g., King et al. 2011; Renzulli et al. 2015) has distinguished four elements that charter school mission statements provide to demonstrate innovation: curriculum (e.g., Montessori or college-oriented), thematic focus (e.g., STEM or marine biology), target population (e.g., gifted or at-risk students), and resources and services (e.g., arts facilities or full-day kindergarten). Other studies have compared charter schools using simple categorical schemes of market "niches" such as district-affiliated vs. not, start-up vs. conversion (Lauen et al. 2015) or critically examined the racial implications of marketing materials from a few large charter management organizations (CMOs)[2] (Hernández 2016). While such studies are illustrative, they significantly reduce the complexity of social contexts, obscuring charter schools' embeddedness in communities of varying demographic characteristics. Indeed, research designs that sort charter schools into single, uniform categories (e.g., McShane and Hatfield 2015; Renzulli et al. 2015) forestall attempts to capture both isomorphism and differentiation in charter school identities (Huerta and Zuckerman 2009). I overcome this limitation through fine-grained, culturally embedded linguistic measures of how schools signal their identities and appeal to specific sociodemographic niches.

Moreover, the scale and depth of previous studies has been limited either by a small number of research sites (e.g., Oakland school district: Jha and Beckman 2017; or Arizona state: King et al. 2011) or the superficiality of the cultural information examined (e.g., by sorting schools into preconceived categories: McShane and Hatfield 2015; or by relying on third-party summaries of charter school missions: Renzulli et al. 2015). The difficulty of collecting comprehensive, valid data and the sensitivity of measurement to geography and history impede effective, theoretically grounded analysis of charter school identities. Building on groundbreaking advances in computational social science (e.g., Mohr, Wagner-Pacifici, and Breiger 2015; Nelson et al. 2018), I overcome these methodological obstacles by collecting detailed, comprehensive, ecologically valid data on charter schools and their social contexts and applying flexible, reliable text-analytic methods.

Data and measures

Schools' websites appeal to parents and school district authorities, connect staff, and detail instructional design choices: the skills and traits it develops, the behaviors it promotes and restrains, its mission and values, its view of the learning process, etc. While websites are ubiquitous and culturally relevant, however, user experience research shows that most readers scan pages quickly, retaining at most 28% of their text content (Nielsen and Morkes 1997). And the text people do read is not taken for granted; an organization's self-descriptive claims made in "About Us" pages tend to be checked against third-party sites (Kaley and Nielsen 2019)—a pattern exacerbated by the prevalence of review sites (e.g., greatschools.org, schooldigger.com) in the top, most-viewed search engine results. Because I operationalize the concept of organizational identity as website self-descriptions, the above conditions make it less likely that identity (as measured here) influences parents' school choices. Thus, this study amounts to a conservative test of my hypothesis.

I used web-crawling in Python 3 (Van Rossum and Drake 2011) to gather data on organizational identities from the websites of all 6,872 charter schools open in 2015-16 (National Center for Education Statistics 2018a), about 92% of which had websites when crawled in June of 2018 (author's calculations). See Appendix A for detail on my web-crawling workflow. My code and URL lists for charter schools and CMOs are available online.[3]

Web-crawling yielded data on 6,300 websites, or 91.7% of all open charter schools. Most of these websites have a significant amount of information: 87.5% include up to 100 web pages, and 88.0% have more than 200 words. However, 7.6% of websites have less than 10 words—a weak information source; I remove these to strengthen my measure of school ideology, reducing the sample to 5,806 schools. I address the possible effects of other outliers (e.g., the 12.5% of schools with more than 100 pages) with robustness checks in Appendix E.

To handle other missing data for those 5,806 schools whose websites I successfully captured, I implemented multiple imputation (Rubin 1976) using the `mi` package in Stata 15 (StataCorp 2017) with 100 imputations. Multiple imputation uses predictive modeling to compute multiple sets of plausible values (here, 100) to replace missing data. Each imputation is then analyzed separately, and their estimates are pooled into a single result—with standard errors reflecting the sampling variability between imputations. Thus, multiple imputation is more efficient and representative than listwise deletion and more precise than single imputation. Accordingly, I dropped only 22 cases missing information on school size, demographics, or grade range served, yielding 5,784 schools in my models.

My first dependent variable is the degree of emphasis on IBL, measured as the percentage of IBL terms on the school's website (see "Dictionaries" section for detail). My second and third dependent variables are the school's percentage of students of color (black, Hispanic, Native American, Asian, Pacific Islander, or multiracial), and the percentage of students receiving free- or reduced-price lunch (FRPL, a proxy for poverty), at the school level. Each of these also serves as independent variables in some analyses. My other independent variables are school district demographics, specifically the percentage of residents of color and the percentage of families below the poverty level; and school academic performance, measured as proficiency rates on standardized state assessments of reading/language arts and mathematics.

To capture these variables, I match web data to school data in the 2015-16 Public School Universe Survey (PSUS) of the National Center for Educational Statistics (NCES 2018a). The PSUS data include ethnicity and FRPL, plus the grade range (as dummy variables: primary, middle, high, and other/ungraded), operating status (used to calculate each school's age), number of students (in hundreds; excludes adult education), urban locale status (whether a school is in a central city of at least 50K residents), and latitude/longitude (which I use to geo-locate charter schools into school districts). I also match to data on school districts in the 2012-16 5-year estimates of the American Community Survey (ACS; U.S. Census Bureau 2018), which includes metrics on ethnicity and poverty. Lastly, I match with the EdFacts 2014-15 school-level proficiency scores in reading/ language arts and mathematics maintained by the U.S. Department of Education (USDE 2018)—high-quality, comprehensive data widely used in education research.

Table 1 provides descriptive statistics on these variables.

Analytic strategy

To predict schools' emphases on IBL, I use count-based dictionary methods and neural-net word embedding models in Python using the Extreme Science and Engineering Discovery Environment (XSEDE; Towns et al. 2014).[4] Below, I outline these methods before detailing how I built the IBL dictionary. I then describe the mixed linear regression models I ran in Stata using the Berkeley Demography Lab cloud computing facility (see https://lab.demog.berkeley.edu/).

Dictionaries. By counting the frequency of terms in the IBL *dictionary*—that is, a list of terms in an overarching category (Stone, Dunphy, and Smith 1966)—I measure the emphasis on IBL within each school's identity. In dictionary methods, researchers develop a list of words connected to a category or concept of interest, then count instances of these words in a sample of texts for purposes of categorization (Grimmer and Stewart 2013).

Dictionary approaches rely on the classic linguistic assumption that language reflects culture: frequent words reflect the cognitive categories most on the author's mind, while rare words are cognitively peripheral or alien (Whorf 1940). Thus, analyzing word frequencies on websites reveal how central is an ideology to a charter school's identity: to the extent that a website uses the concepts in an ideology's dictionary, the school identifies with that ideology.

I operationalize the emphasis on the IBL ideology as the ratio of the number of times a concept from the dictionary appears on a given website divided by the total number of words on that website (to account for varying lengths of websites). Ideological emphasis thus has a range of [0,1], where 0 indicates that *none* of the website's words are concepts from the ideology and 1 indicates that *all* of the website's words are concepts from the ideology (see Appendix B for examples of charter school websites with high and low IBL emphasis). Given that term counts for both IBL and word totals are skewed right (see Table 1) to a degree that could change with website length, I account for possible bias by taking the log of each measure. I thus calculate emphasis as follows:[5]

$$Emphasis = \frac{\log(\#\text{ inquiry terms})}{\log(\#\text{ total terms})}$$

A dictionary is restricted in application to the ecological context in which it was generated and validated (Grimmer and Stewart 2013; Nelson et al. 2018). To my knowledge, no dictionary has been developed for a context like school websites; accordingly, I create and validate an IBL dictionary specific to my web corpus. Specifically, I construct word embedding models from charter school websites, and then I use these to iteratively expand a set of seed terms into a longer dictionary of words and phrases that have similar meaning as the core concepts (see below for explanation of this method). An emerging trend in the social sciences, this workflow represents a modified form of computational grounded theory (Nelson 2017): it begins with content knowledge (seed terms), detects patterns with unsupervised computational methods (finding similar terms with word embeddings), and confirms sociological patterns by applying dictionaries of different sizes—through counts or cosine distances—to specific corpora (e.g., movie reviews or social media posts: Garten et al. 2018; Sivak and Smirnov 2019).

Word embeddings. Word embeddings map words onto a high-dimensional vector space and represent semantic relations between words as geometric relations in space (Mikolov et al. 2013). Word embeddings are becoming more common in the social sciences; for example, to analyze associations between basic cultural categories such as gender, race, and status/wealth (Kozlowski, Taddy, and Evans 2019).

A word's position in vector space is based on the context that it shares with other words in the focal text. Words that share many contexts (i.e., words that are frequently collocated with the same other words) are positioned near each other in vector space, and words that have very different contexts (i.e., words that are collocated with different other words) are positioned far apart. In other words, words that are positioned near each other in vector space share similar meanings, so vector space can be understood as semantic space. Importantly, this relational mapping captures commonalities in words' local contexts, rather than collocation alone. This large-scale mapping of contexts encodes word embeddings with underlying cultural meanings, rather than strictly on-the-ground observable patterns.

Mechanically, word embeddings assess word associations using 'word context windows', indicating the number of words (typically 5-12) on either side of focal word w to consider as connected to w. Formally, for a series of training words w_1, w_2, w_3, … , w_T, the goal of word embeddings to maximize the average log probability of predicting w_{t+j} given w_t (Mikolov et al. 2013:2):

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{-c\leq j\leq c, j\neq 0}\log p\left(w_{t+j}\middle|w_t\right) \tag{1}$$

where w indicates a word in the sentence, t is the iterator over T training words in the sample sentence, c is the word context window size, and j is a number between $-c$ and c (excluding zero) that indicates the distance in words from focal word w_t to a word within its context, w_{t+j}.

The high number of dimensions in the vector space (typically 100-300) means that Euclidean or straight-line distances between vectors cannot be calculated (Kozlowski et al. 2019:9). Instead, the distances between word vectors are established via cosine similarity, which measures the angle between vectors such that a score of 0 indicates perfect independence (orthogonality or 90 degrees between vectors) and 1 indicates perfect similarity (parallelity or 0 degrees between vectors). See Appendix D for additional technical notes on word embeddings.

Development of IBL dictionary. To borrow from the rich literature on IBL in educational psychology, I began with five seed terms taken directly from the subtitle of a seminal article: "The failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching" (Kirschner, Sweller, and Clark 2006). Thus, my five seed terms are: "inquiry-based", "problem-based", "discovery-based", "experiential", and "constructivist".[6] Indeed, though conceptually distinct (e.g., Steffe and Gale 1995), empirical research generally treats these terms as synonyms. My word embeddings support this claim: the average cosine similarity of these five terms is high at 0.70.

In a blended data- and hand-driven procedure (Nelson 2017; Schwartz and Ungar 2015), I used the `most_similar()` method in Python's `word2vec` module (Řehůřek and Sojka 2010) to identify word vectors positioned in the vector space near the seed terms, indicating semantic similarity. I then applied two manual filters to focus the dictionary. First, I removed terms conceptually distinct (in my eyes) from IBL. For example, I removed the terms "strengths-based", "small-group", "explorations", and "self-motivated" because these teaching methods and desirable traits can be taught without IBL—regardless of their cosine similarities to the IBL seed terms (0.76, 0.59, 0.56, and 0.51, respectively). Second, I removed terms that occurred rarely in the corpus, such as "sustained inquiry" (7 counts) and "multimodal" (14 counts), but kept any obvious synonyms to the seed terms, such as "problem-centered" (12 counts, but identical to "problem-based"). These filters ensured that new terms were theoretically relevant; for this reason, manual checks are a common means of validating dictionaries built with word embeddings (e.g., Sivak and Smirnov 2019; see Appendix E for more on dictionary validation). For visualization of the full dictionary, see Figure 1; for per-term counts and cosine similarities (between a given term and the seed terms), see Appendix C.

[FIGURE 1 ABOUT HERE]

Mixed-effects models. I next estimate mixed-effects linear models (which have both fixed and random effects) in STATA 15 for dependent variables at the school level. Given that state policy (e.g., Bodine et al. 2008; Finnigan 2007), school district practices (e.g., Paino 2018), and CMO structure (e.g., Furgeson et al. 2012) have significant, convergent impacts on charter school practices and outcomes, ignoring the nesting of charter schools (level-1) in states, school districts, and CMOs (level-2) by pooling all observations would violate the assumption of independence required for ordinary least squares (OLS) regression and thus bias the results.[7]

Furthermore, these are cross-classified data in that each CMO could have schools in multiple states, and each state could be home to multiple CMOs. These data are also hierarchically clustered (Goldstein 1987) in that school districts are nested within states. To accommodate this data structure, my mixed-effects models could estimate crossed random effects for state and CMO as well as random effects for school districts nested in states. Such models take the following form:

$$y_{isdc} = \beta_0 + \beta' x_{isdc} + \zeta_s + \zeta_d + \zeta_c + \varepsilon_{isdc} \tag{2}$$

where *i* represents schools, *s* represents states, *d* represents school districts, *c* represents CMOs, y_{isdc} is the emphasis on IBL, β_0 is the intercept, x_{isdc} is a vector of explanatory and control variables, ζ_s is the random effect for state *s*, ζ_d is the random effect for school district *d,* ζ_c is the random effect for CMO *c*, and ε_{isdc} is the error term. The first random effect captures unobserved factors that might shape each state's effect on school ideology; the second captures those factors that might affect ideology within each school district; and the third captures those that might affect ideology in each CMO. Thus, in this full model, controls at the levels of state, school district, and CMO are not necessary.

In addition, I assessed nestedness by measuring the Intraclass Correlation Coefficient (ICC) or *rho* (ρ) for each level through preliminary models. By measuring the proportion of residual variance in the outcome explained by the non-independence (nesting) of units in each level, the ICC indicates whether a random effect for that level is necessary. Accordingly, I nested my IBL models by CMO, my school poverty models by school district, and my school ethnicity models by state and school district. I report revised, model-specific ICC values below.

I include school-level controls for grade range (dummies indicating whether a primary, middle, or high school), age of school (in 2015-16 school year; logged), number of students (logged), and an indicator for urban locale status (1 if a school is located in a principal city and urbanized area, 0 otherwise). In models with IBL, I also include a control for the % of a website pages that are PDFs (e.g., student handbooks, charter applications), which tend to be longer and more procedural in tone than other web pages and thus influence the language used. Lastly, in models with academic proficiency rates, I include measures of data "blurring" by the U.S. Department of Education to protect the identities of (especially small) student groups. These measures scale with data precision: possible values are 1 (percentiles reported, e.g. 95% academic proficiency for a school; the most precise), 5 (quintiles reported, e.g. 90-95%), 10 (deciles reported, e.g. 80-90%), 20 (ventiles reported, e.g. 60-80%), and 50 (medians reported, e.g. 50-100%; the least precise).

The first set of mixed linear models feature a lagged dependent variable: IBL was captured via web-crawling in June 2018, while the sociodemographic predictors were measured in 2015-16 (at school level) or 2012-16 (at school district level). This time lapse between measures strengthens the argument that sociodemographics influences educational ideology, because this means potential founders have over two years to establish new schools or retune their educational approach to respond to local demands. In predicting school sociodemographics, the second set of models similarly uses lagged academic proficiency scores (measured in 2014-15), allowing parents and resource providers time to observe and react to objective signals of school quality. However, the second set also includes educational ideology, a covariate measured several years prior to the outcome (June 2018 compared to 2015-16)—making the second set of models a more conservative test of my hypothesis.

RESULTS

Correlations

The correlations in Table 1 support my analytic approach. IBL correlates with demographic contexts in the directions predicted: IBL is negatively correlated with school percentage poverty and percentage students of color and school district percentage poverty and percentage people of color (-0.18, -0.17, -0.11, and -0.06, respectively), and all of these correlations are significant at the $p<0.05$ level. Moreover, academic proficiency is negatively correlated with poverty and percent students of color at both the school and district levels, suggesting that academic quality may be an alternative explanation for charter school enrollment patterns. These correlations tentatively support my hypothesis—though they are subject to confounding and do not disentangle causal relationships.

In addition, the correlations between the demographic variables are positive and strong, ranging from 0.27 (between school district percentage people of color and school percentage poverty) and 0.62 (between school district percentage people of color and school percentage students of color). Due to this high correlation among sociodemographic variables, the direct relationship of each with the outcome would be muddled if all were included in a single model; thus, I estimate the effect of each of these independent variables separately.

[TABLE 1 ABOUT HERE]

Mixed-effects models

The mixed-effects linear models regressing IBL emphasis on school and school district poverty and race—plus school controls—are shown in Table 2. These findings support my hypothesis, for school and district percentage poverty and percentage nonwhite all have statistically significant, negative relationships with IBL emphasis in their respective models (see Figure 2). Moreover, the effect sizes are considerable: each standard deviation increase in school poverty, school percentage students of color, district poverty, or district percentage people of color is associated in their respective models with a change in IBL emphasis of -1.48, -1.63, -1.15, and -0.418 standard deviations, respectively. Thus, the effect of school poverty and ethnic composition on IBL greater than for school district poverty and race. This difference is most dramatic for school district race, which unlike school district poverty has a small beta coefficient as well (-0.03 vs. -0.21). As such, these findings suggest that sociodemographics—and especially poverty—at the level of school, rather than the school district, have the stronger, more negative relationship with IBL emphasis.

Furthermore, the ICC indicates that CMO membership explains an impressive 34% or so of the variation in IBL emphasis, and the conservative χ^2 test confirms the need in the model for random effects. Together, these findings suggest that schools sharing a CMO are significantly more alike in their IBL emphasis than are schools not sharing a CMO.

[FIGURE 2 ABOUT HERE]

[TABLE 2 ABOUT HERE]

The mixed-effects linear models regressing school poverty and school ethnicity on IBL emphasis and academic quality—plus school controls—are shown in Table 3. These findings support my hypothesis, for IBL emphasis has statistically significant, positive relationships with school poverty and ethnicity in their respective models. Furthermore, this effect does not disappear when academic proficiency is accounted for (see Figure 3)—though it does decrease in size, the effects of proficiency in reading/language arts and math also dip slightly. In the full models, each standard deviation increase in IBL emphasis is associated with a change in school poverty and percent students of color of -2.32 and -3.18, respectively. In contrast, the same unit change for reading/language arts proficiency is -4.60 and -4.45, and for math proficiency it is -0.741 and -0.901. Thus, the effect of IBL emphasis is weaker than that of reading proficiency but greater than for math proficiency, which is only marginally significant.

In these models, ICC measures indicate that school district membership explains about 37% of the variation in both school poverty and school ethnicity. And the conservative χ^2 test confirms the need for random effects, which are statistically significant ($p < 0.01$) at the CMO and school district levels. Together, these findings underscore the sociodemographic similarity of schools sharing a school district.

[FIGURE 3 ABOUT HERE]

[TABLE 3 ABOUT HERE]

To review, not only are demographic factors correlated with IBL emphasis in the predicted directions, but also these same directions hold with statistical significance in each model. School poverty and ethnicity and school district poverty all have strong relationships with IBL emphasis. Moreover, the significant relationship between IBL emphasis and school sociodemographics is robust to traditional measures of school quality—and to various filters and alternative specifications (see Appendix E). This effectively discounts the alternative explanation that charter school identity has no link with enrollment patterns once academic quality is accounted for. Finally, both ideology and sociodemographics are embedded in nested organizational contexts: significant variation in IBL is explained by CMO membership, while substantial variation in school class and race is explained by school district and state membership.

DISCUSSION AND CONCLUSIONS

These results support my argument that charter schools' identities are associated with class- and race-differentiated social contexts. That is, charter schools present themselves differently—by virtue of their explicit educational ideologies—to different race and class niches (Carroll 1985; Lauen et al. 2015). Specifically, I find that charter schools present themselves to affluent and (especially) white communities in ways emphasizing IBL, and that schools emphasizing IBL have student enrollments that are more affluent and white—independent of objective measures of school quality. This provides initial support for my theory that schools' self-presentation strategies—in particular, their educational ideologies—respond to race- and class-specific educational values and expectations (e.g., Erickson 2017; Posey-Maddox et al. 2014) and culturally distinct parenting styles (Lareau 2000, 2011). And this relationship may be driven as much by parents selecting schools (school selection effects), evidenced by the relationship between school race and poverty and the educational ideology of IBL; as by schools selecting school districts (neighborhood selection effects), evidenced by the relationship between district race and poverty and IBL.

My research helps resolve the pressing ambiguity over the relationship between charter schools, on one hand, and persistent racial and socioeconomic divisions in U.S. society, on the other. And yet, no research to date has examined how heterogeneous organizational identities in charter schools (e.g., King et al. 2011; Renzulli et al. 2015)—the most widespread school choice reform today (Berends 2015)—interact with the community social forces underlying parents' educational preferences to create charter school segregation. Underlying this gap is a dearth of research at the nexus of organizations and education (Renzulli 2014).

In beginning to fill this gap, my findings suggest (but cannot prove) that the intentional ideological differentiation of charter schools reinforces social inequalities—rather than alleviates them, as educational reformers claim (e.g., Roth et al. 2017). The present study initiates a research program to examine evidence that organizationally differentiated identities attract socially differentiated 'clients', a mechanism by which parents and schools may self-sort along dimensions of inequality such as race and class (Holme 2002; Roda and Wells 2013). Multiple methods and studies are required to demonstrate that such self-sorting is facilitated by resonance between race- and class-specific socialization and parenting styles (Lareau 2000, 2011), on one hand, and the educational ideologies constructed by charter schools seeking to attract parents (e.g., Jha and Beckman 2017), on the other.

Indeed, the present large-scale, observational study lays the groundwork for direct observational and longitudinal studies. The web-based data of the present study require grounded observation through interviews or experiments to confirm the impacts of schools' heterogeneous identity claims (a process called "triangulation"; Powell et al. 2016), given that websites themselves may represent a form of "myth and ceremony" (Meyer and Rowan 1977). For unlike the present cross-sectional study, behavioral studies are capable of directly evidencing individual-level mechanisms. Similarly, with longitudinal data scholars could disentangle the causal influence of sociodemographics and educational ideology through parents' and schools' self-sorting.

Such insights have important implications for education policy, which has offered incentives and support for school choice programs including charter schools, vouchers, and home schooling since the federal No Child Left Behind law of 2001, continuing with President Obama's Race to the Top initiative and Education Secretary Betsy DeVos. Similarly, 44 states and D.C. now have charter laws to provide "innovative" education. At the same time, the U.S. Supreme Court has placed increasing legal limits on mandatory school district desegregation plans since *Brown v. Board of Education* (1954), resulting in the dismissal of roughly half of such plans by 2007 (Lutz 2011)—and in turn, gradual resegregation by race and income (Reardon et al. 2012; Reardon and Owens 2014).

The upshot of this political and legal legacy is stratified school choice systems in which race and class groups congregate in schools with others like them (e.g., Frankenberg and Siegel-Hawley 2013; Paino et al. 2017)—and in particular, relatively affluent, mobile white families concentrate (e.g., Holme 2002) in schools and districts out of reach of the less privileged (e.g., Renzulli and Evans 2005; Saporito 2003). In such a system, schools of choice too are necessarily exclusionary: seeking to differentiate themselves within educational "markets" segregated by race and class (e.g., Fuller 2009; Lauen et al. 2015), such schools face pressure from authorizers and clients to cater to sociodemographic niches through location and admission decisions. This study (and those to follow) suggests a consequence of this organizational pressure: Charter schools select the best performers (e.g., Abdulkadiroglu et al. 2017; Lacireno-Paquet et al. 2002) not universally, but among the race- and class-specific niche they seek to attract.

Many political and organizational solutions have been offered for segregated schools, including equitable access to information on school quality (e.g., Yettick 2016), accountability and fairness in charter schools' admissions procedures (e.g., Abdulkadiroglu et al. 2017; Holme 2002), or opening racially diverse schools (e.g., Roda and Wells 2013). Indeed, by illuminating an additional mechanism by which schools of choice participate in self-perpetuating inequalities, this work underscores the importance of incorporating into school choice programs stronger oversight of enrollment practices and encouragement for diverse schools. Specifically, to obtain and reauthorize their charters, schools could be required to share their marketing strategies, student demographics, and evidence of attempts to recruit diverse student bodies. So long as such integrating organization is absent, centrifugal social dynamics from organizational differentiation to neighborhood segregation to parents' racially charged perceptions of school quality will likely continue to pull students and families into charter schools where they meet fewer peers of race and class backgrounds different from their own—potentially undermining ethnic integration and our civic ideals.

DATA ACCESSIBILITY

Replication data and code are available at http://bit.ly/sorting-schools-2019 and a methods pre-registration is available at http://bit.ly/OSF-prereg-SocEd. To access the original web-crawled data, please contact the author; these data will not be publicly released pursuant to the copyrights on some charter schools' web content. Data access has been granted to research assistants only for the duration of their position.

RESEARCH ETHICS

The research reported here used only publicly available, organization-level data and did not involve human subjects. It was thus not subject to review by the institutional review board for the protection of human subjects. Web-crawling speeds were throttled to prevent server overload.

ENDNOTES

[1] The pedagogy of IBL (e.g., "What guidance is essential for student learning?", "To what extent should students 'act like scientists'?"), its efficacy (e.g., "How much learning do students retain in long-term memory?"), and even its terminology (e.g., "What is 'minimally guided instruction?'"; Kirschner, Sweller, and Clark 2006) have been rigorously debated for decades (for a recent review, see Lazonder and Harmsen 2016).

[2] I define a CMO broadly as any private organization that manages two or more charter schools—so long as this organization's website lists all schools it manages, thus encapsulating them within a larger "brand" identity. CMOs manage a large and growing segment of the charter sector (45% in 2014-15 according to some estimates; National Alliance for Public Charter Schools 2018), and some boast impressive networks of schools; e.g., the KIPP network includes 242 schools as of spring 2020 (see https://www.kipp.org/).

[3] See my code for collecting URLs, crawling web pages, and parsing text—as well as a complete list of charter school and CMO URLs—at http://bit.ly/web_crawl_tools.

[4] In particular, I used the Jetstream resource at Indiana University through allocation TG-SES170018.

[5] To avoid dropping from my models schools with no IBL terms on their websites—otherwise producing an undefined numerator from `log(0)`—I add the number 1 to both the # IBL terms and the # total terms when implementing this formula.

[6] I changed the word "discovery" to "discovery-based" because unlike the celebratory former, in my web corpus the latter is reliably pedagogical in meaning.

[7] While modeling with fixed effects (FE) controls for level 2 influences (e.g., state policy, CMO size) and thus eliminates unobserved sources of heterogeneity (Schneider et al. 2007), pure FE models give poor estimates for low-sample groups and control away all across-state, across-district, and across-CMO variation in outcomes. Mixed-effects models resolve these concerns through partial pooling—estimates for states and CMOs with fewer observations are based partially on the estimates for larger such groups.

REFERENCES

Abdulkadiroglu, Atila, Parag A. Pathak, Jonathan Schellenberg, and Christopher R. Walters. 2019. "Do Parents Value School Effectiveness? " American Economic Review (forthcoming).

Achieve. 2010. *International Science Benchmarking Report. Taking the Lead in Science Education: Forging Next-Generation Science Standards*. Achieve, Inc.

Brown v. Board of Education. 1954. 347 U.S. 483.

Arum, Richard. 1996. "Do Private Schools Force Public Schools to Compete?" *American Sociological Review* 61(1):29–46.

Asante, Molefi Kete, and Diane Ravitch. 1991. "Multiculturalism: An Exchange." *The American Scholar* 60(2):267–76.

Bail, Christopher A. 2014. "The Cultural Environment: Measuring Culture with Big Data." *Theory and Society* 43(3–4):465–82.

Bennett, W. Lance, and Alexandra Segerberg. 2013. *The Logic of Connective Action: Digital Media and the Personalization of Contentious Politics*. New York, NY: Cambridge University Press.

Berends, Mark. 2015. "Sociology and School Choice: What We Know After Two Decades of Charter Schools." *Annual Review of Sociology* 41(1):159–80.

Bodine, Edward, Bruce Fuller, María-Fernanda González, Luis Huerta, Sandra Naughton, Sandra Park, and Laik Woon Teh. 2008. "Disparities in Charter School Resources—the Influence of State Policy and Community." *Journal of Education Policy* 23(1):1–33.

Bourdieu, Pierre. 1977. *Outline of a Theory of Practice*. Cambridge University Press.

Bowles, Samuel, and Herbert Gintis. 1976. *Schooling in Capitalist America*. New York, NY: Basic Books.

Bruner, Jerome S. 1961. "The Act of Discovery." *Harvard Educational Review* 31:21–32.

Calarco, Jessica McCrory. 2011. "'I Need Help!' Social Class and Children's Help-Seeking in Elementary School." *American Sociological Review* 76(6):862–882.

Carroll, Glenn R. 1985. "Concentration and Specialization: Dynamics of Niche Width in Populations of Organizations." *American Journal of Sociology* 90:1262–83.

Chubb, John E., and Terry M. Moe. 1990. *Politics, Markets, and America's Schools*. Washington, D.C.: Brookings Institution Press.

Dewey, John. 1938. *Experience and Education*. Indianapolis, IN: Kappa Delta Pi.

Epple, Dennis, David Figlio, and Richard Romano. 2004. "Competition between Private and Public Schools: Testing Stratification and Pricing Predictions." *Journal of Public Economics* 88(7):1215–45.

Erickson, Heidi Holmes. 2017. "How Do Parents Choose Schools, and What Schools Do They Choose? A Literature Review of Private School Choice Programs in the United States." *Journal of School Choice* 11(4):491–506.

Finnigan, Kara S. 2007. "Charter School Autonomy: The Mismatch between Theory and Practice." *Educational Policy* 21(3):503–26.

Frankenberg, Erica, Jongyeon Ee, Jennifer B. Ayscue, and Gary Orfield. 2019. *Harming Our Common Future: America's Segregated Schools 65 Years after Brown*. Los Angeles, CA: The Civil Rights Project.

Frankenberg, Erica, and Genevieve Siegel-Hawley. 2013. "A Segregating Choice?: An Overview of Charter School Policy, Enrollment Trends, and Segregation." Pp. 129–44 in *Educational Delusions?: Why Choice Can Deepen Inequality and How to Make Schools Fair*. University of California Press.

Fuller, Bruce. 2009. "Policy and Place: Learning from Decentralized Reforms." Pp. 855–875 in *Handbook of education policy research*, edited by G. Sykes, B. Schneider, and D. Plank. New York: Routledge.

Furgeson, Joshua, Brian Gill, Joshua Haimson, Alexandra Killewald, Moira McCullough, Ira Nichols-Barrer, Bing-ru Teh, Natalya Verbitsky-Savitz, Melissa Bowen, Allison Demeritt, Paul Hill, and Robin Lake. 2012. *Charter-School Management Organizations: Diverse Strategies and Diverse Student Impacts*. Cambridge, MA: Mathematica Policy Research, Inc./Center on Reinventing Public Education.

Garten, Justin, Joe Hoover, Kate M. Johnson, Reihane Boghrati, Carol Iskiwitch, and Morteza Dehghani. 2018. "Dictionaries and Distributions: Combining Expert Knowledge and Large Scale Textual Data Content Analysis." *Behavior Research Methods* 50(1):344–61.

Gleason, Philip M. 2017. "What's the Secret Ingredient? Searching for Policies and Practices That Make Charter Schools Successful." *Journal of School Choice* 11(4):559–84.

Golann, Joanne W. 2015. "The Paradox of Success at a No-Excuses School." *Sociology of Education* 88(2):103–19.

Goldstein, Harvey. 1987. "Multilevel Covariance Component Models." *Biometrika* 74(2):430–431.

Grimmer, Justin, and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21(3):267–97.

Hanushek, Eric A., John F. Kain, and Steven G. Rivkin. 2009. "New Evidence about Brown v. Board of Education: The Complex Effects of School Racial Composition on Achievement." *Journal of Labor Economics* 27(3):349–83.

Hanushek, Eric A., John F. Kain, Steven G. Rivkin, and Gregory F. Branch. 2007. "Charter School Quality and Parental Decision Making with School Choice." *Journal of Public Economics* 91(5):823–48.

Hernández, Laura E. 2016. "Race and Racelessness in CMO Marketing: Exploring Charter Management Organizations' Racial Construction and Its Implications." *Peabody Journal of Education* 91(1):47–63.

Holme, Jennifer Jellison. 2002. "Buying Homes, Buying Schools: School Choice and the Social Construction of School Quality." *Harvard Educational Review* 72(2):177–206.

Huerta, Luis A., and Andrew Zuckerman. 2009. "An Institutional Theory Analysis of Charter Schools: Addressing Institutional Challenges to Scale." *Peabody Journal of Education* 84(3):414–31.

Jha, Harsh K., and Christine M. Beckman. 2017. "A Patchwork of Identities: Emergence of Charter Schools as a New Organizational Form." Pp. 69–107 in *Emergence*. Emerald Publishing Limited.

Kaley, Anna, and Jakob Nielsen. 2019. "'About Us' Information on Corporate Websites." *Nielsen Norman Group*. Retrieved September 20, 2019 (https://www.nngroup.com/articles/about-us-information-on-websites/).

King, Brayden G., Elisabeth S. Clemens, and Melissa Fry. 2011. "Identity Realization and Organizational Forms: Differentiation and Consolidation of Identities Among Arizona's Charter Schools." *Organization Science* 22(3):554–72.

Kirschner, Paul A., John Sweller, and Richard E. Clark. 2006. "Why Minimal Guidance during Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching." *Educational Psychologist* 41(2):75–86.

Kozlowski, Austin C., Matt Taddy, and James A. Evans. 2019. "The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings." *American Sociological Review* 84(5):905–49.
Lacireno-Paquet, Natalie, Thomas T. Holyoke, Michele Moser, and Jeffrey R. Henig. 2002. "Creaming Versus Cropping: Charter School Enrollment Practices in Response to Market Incentives." *Educational Evaluation and Policy Analysis* 24(2):145–58.
Ladd, Helen F., ed. 1996. *Holding Schools Accountable: Performance-Based Reform in Education*. Washington, D.C.: Brookings Institution Press.
Lareau, Annette. 2000. *Home Advantage: Social Class and Parental Intervention in Elementary Education*. 2nd ed. Rowman & Littlefield Publishers.
Lareau, Annette. 2011. *Unequal Childhoods: Class, Race, and Family Life*. 2nd ed. Berkeley: University of California Press.
Lauen, Douglas Lee, Bruce Fuller, and Luke Dauter. 2015. "Positioning Charter Schools in Los Angeles: Diversity of Form and Homogeneity of Effects." *American Journal of Education* 121(2):213–39.
Lazonder, Ard W., and Ruth Harmsen. 2016. "Meta-Analysis of Inquiry-Based Learning: Effects of Guidance." *Review of Educational Research* 86(3):681–718.
Levinson, Meira. 2012. *No Citizen Left Behind*. Harvard University Press.
Lubienski, Christopher. 2003. "Innovation in Education Markets: Theory and Evidence on the Impact of Competition and Choice in Charter Schools." *American Educational Research Journal* 40(2):395–443.
Lutz, Byron. 2011. "The End of Court-Ordered Desegregation." *American Economic Journal: Economic Policy* 3(2):130–68.
Maaten, Laurens van der, and Geoffrey Hinton. 2008. "Visualizing Data Using T-SNE." *Journal of Machine Learning Research* 9(Nov):2579–2605.
Malkus, Nat. 2016. "Seeing Charters Differently: A New Approach to National Comparisons of Charter and Traditional Public Schools." *Journal of School Choice* 10(4):479–94.
Massey, Douglas S., Jonathan Rothwell, and Thurston Domina. 2009. "The Changing Bases of Segregation in the United States." *The Annals of the American Academy of Political and Social Science* 626(1).
McShane, Michael Q., and Jenn Hatfield. 2015. *Measuring Diversity in Charter School Offerings*. Washington, D.C.: American Enterprise Institute (AEI).
Meyer, John W., and Brian Rowan. 1977. "Institutionalized Organizations: Formal Structure as Myth and Ceremony." *The American Journal of Sociology* 83(2):340–63.
Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. 2013. "Distributed Representations of Words and Phrases and Their Compositionality." *Advances in Neural Information Processing Systems* 3111–3119.
Mohr, John W., Robin Wagner-Pacifici, and Ronald L. Breiger. 2015. "Toward a Computational Hermeneutics." *Big Data & Society* 2(2):2053951715613809.
Monarrez, Tomas, Brian Kisida, and Matthew Chingos. 2019. *When Is a School Segregated? Making Sense of Segregation 65 Years after Brown v. Board of Education*. Urban Institute.
Moreno, Ivan. 2017. "US Charter Schools Put Growing Numbers in Racial Isolation." *AP News*. Retrieved January 10, 2018 (https://apnews.com/e9c25534dfd44851a5e56bd57454b4f5).
National Alliance for Public Charter Schools. 2018. "Data Dashboard." Retrieved October 1, 2018 (https://data.publiccharters.org/).

National Center for Education Statistics. 2018a. *Common Core of Data: Public Elementary/Secondary School Universe Survey*. Washington, D.C.: U.S. Department of Education.

National Center for Education Statistics. 2018b. "Digest of Education Statistics, 2017." *Enrollment and Percentage Distribution of Enrollment in Public Elementary and Secondary Schools, by Race/Ethnicity and Level of Education: Fall 1999 through Fall 2027*. Retrieved October 1, 2018 (https://nces.ed.gov/programs/digest/d17/tables/dt17_203.60.asp?).

Nelson, Laura K. 2017. "Computational Grounded Theory: A Methodological Framework." *Sociological Methods & Research* 1–40.

Nelson, Laura K., Derek Burk, Marcel Knudsen, and Leslie McCall. 2018. "The Future of Coding: A Comparison of Hand-Coding and Three Types of Computer-Assisted Text Analysis Methods." *Sociological Methods & Research* 0049124118769114.

Nielsen, Jakob, and John Morkes. 1997. "Concise, SCANNABLE, and Objective: How to Write for the Web." *Nielsen Norman Group*. Retrieved September 20, 2019 (https://www.nngroup.com/articles/concise-scannable-and-objective-how-to-write-for-the-web/).

Orfield, Myron, Baris Gumus-Dawes, and Thomas Luce. 2013. "The State of Public Schools in Post-Katrina New Orleans: The Challenge of Creating Equal Opportunity." Pp. 159–84 in *Educational Delusions?: Why Choice Can Deepen Inequality and How to Make Schools Fair*. University of California Press.

Owens, Ann, Sean F. Reardon, and Christopher Jencks. 2016. "Income Segregation Between Schools and School Districts." *American Educational Research Journal* 53(4):1159–97.

Paino, Maria. 2018. "From Policies to Principals: Tiered Influences on School-Level Coupling." *Social Forces* 96(3):1119–54.

Paino, Maria, Rebecca L. Boylan, and Linda A. Renzulli. 2017. "The Closing Door: The Effect of Race on Charter School Closures." *Sociological Perspectives* 60(4):747–67.

Pettigrew, Andrew M. 1979. "On Studying Organizational Cultures." *Administrative Science Quarterly* 24(4):570–81.

Posey-Maddox, Linn, Shelley McDonough Kimelberg, and Maia Cucchiara. 2014. "Middle-Class Parents and Urban Public Schools: Current Research and Future Directions." *Sociology Compass* 8(4):446–56.

Powell, Walter W., Aaron Horvath, and Christof Brandtner. 2016. "Click and Mortar: Organizations on the Web." *Research in Organizational Behavior* 36:101–120.

Preston, Courtney, Ellen Goldring, Mark Berends, and Marisa Cannata. 2012. "School Innovation in District Context: Comparing Traditional Public Schools and Charter Schools." *Economics of Education Review* 31(2):318–30.

Reardon, Sean F., Elena Tej Grewal, Demetra Kalogrides, and Erica Greenberg. 2012. "Brown Fades: The End of Court-Ordered School Desegregation and the Resegregation of American Public Schools." *Journal of Policy Analysis and Management* 31(4):876–904.

Reardon, Sean F., and Ann Owens. 2014. "60 Years After Brown: Trends and Consequences of School Segregation." *Annual Review of Sociology* 40(1):199–218.

Reardon, Sean F., Joseph Townsend, and Lindsay Fox. 2017. "A Continuous Measure of the Joint Distribution of Race and Income Among Neighborhoods." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 3(2):34–62.

Řehůřek, Radim, and Petr Sojka. 2010. "Software Framework for Topic Modelling with Large Corpora." Pp. 45–50 in *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*.

Renzulli, Linda. 2014. "Educational Transformations and Why Sociology Should Care." *Social Currents* 1(2):149–56.

Renzulli, Linda A., Ashley B. Barr, and Maria Paino. 2015. "Innovative Education? A Test of Specialist Mimicry or Generalist Assimilation in Trends in Charter School Specialization Over Time." *Sociology of Education* 88(1):83–102.

Renzulli, Linda A., and Lorraine Evans. 2005. "School Choice, Charter Schools, and White Flight." *Social Problems* 52(3):398–418.

Roda, Allison, and Amy Stuart Wells. 2013. "School Choice Policies and Racial Segregation: Where White Parents' Good Intentions, Anxiety, and Privilege Collide." *American Journal of Education* 119(2):261–93.

Roth, Erin, Abel McDaniels, Catherine Brown, and Neil Campbell. 2017. "The Progressive Case for Charter Schools." *Center for American Progress*. Retrieved (https://www.americanprogress.org/issues/education-k-12/news/2017/10/24/440833/the-progressive-case-for-charter-schools/).

Rothstein, Jesse M. 2006. "Good Principals or Good Peers? Parental Valuation of School Characteristics, Tiebout Equilibrium, and the Incentive Effects of Competition among Jurisdictions." *American Economic Review* 96(4):1333–50.

Rubin, Donald B. 1976. "Inference and Missing Data." *Biometrika* 63(3):581–592.

Saporito, Salvatore. 2003. "Private Choices, Public Consequences: Magnet School Choice and Segregation by Race and Poverty." *Social Problems* 50(2):181–203.

Schneider, Barbara, Martin Carnoy, Jeremy Kilpatrick, William H. Schmidt, and Richard J. Shavelson. 2007. *Estimating Causal Effects Using Experimental and Observational Designs*. *Think Tank White Paper*. Washington, D.C.: American Educational Research Association.

Schwartz, H. Andrew, and Lyle H. Ungar. 2015. "Data-Driven Content Analysis of Social Media: A Systematic Overview of Automated Methods." *The ANNALS of the American Academy of Political and Social Science* 659(1):78–94.

Shanker, Albert. 1988. "Restructuring Our Schools." *Peabody Journal of Education* 65(3):88–100.

Sivak, Elizaveta, and Ivan Smirnov. 2019. "Parents Mention Sons More Often than Daughters on Social Media." *Proceedings of the National Academy of Sciences* 116(6):2039–41.

StataCorp. 2017. *Stata 15 Multiple Imputation Reference Manual*. College Station, TX: Stata Press.

Steffe, Leslie P., and Jerry Edward Gale. 1995. *Constructivism in Education*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Stone, Philip J., Dexter C. Dunphy, and Marshall S. Smith. 1966. *The General Inquirer: A Computer Approach to Content Analysis*. Oxford, England: M.I.T. Press.

Teske, Paul, Jody Fitzpatrick, and Gabriel Kaplan. 2006. "The Information Gap?" *Review of Policy Research* 23(5):969–81.

Towns, John, Timothy Cockerill, Maytal Dahan, Ian Foster, Kelly Gaither, Andrew Grimshaw, Victor Hazlewood, Scott Lathrop, Dave Lifka, Gregory D. Peterson, Ralph Roskies, J. Ray Scott, and Nancy Wilkins-Diehr. 2014. "XSEDE: Accelerating Scientific Discovery." *Computing in Science & Engineering* 16(5):62–74.

U.S. Census Bureau. 2018. *American Community Survey, 2012-16 Summary File Data*. Washington, D.C.

U.S. Department of Education. 2018. "EDFacts Data Files." Retrieved September 15, 2016 (http://www2.ed.gov/about/inits/ed/edfacts/data-files/index.html).

Van Rossum, Guido, and Fred L. Drake. 2011. *Python Language Reference Manual*. United Kingdom: Network Theory Limited.

Waite, Chelsea. 2019. *A View from the Canopy: Building Collective Knowledge on School Innovation*. Christensen Institute.

Wang, Ke, Amy Rathbun, and Lauren Musu. 2019. *School Choice in the United States: 2019*. Washington, D.C.: National Center for Education Statistics.

Watanabe, Maika. 2007. "Displaced Teacher and State Priorities in a High-Stakes Accountability Context." *Educational Policy* 21(2):311–68.

Whorf, Benjamin Lee. 1940. "Science and Linguistics." *Technology Review* 42(6):229–31, 247–48.

Yettick, Holly. 2016. "Information Is Bliss: Information Use by School Choice Participants in Denver." *Urban Education* 51(8):859–90.

Table 1. Descriptive Statistics and Correlations.

	IBL	RLA	math	pov_ schl.	SOC	pov_ SD	POC _SD	pri- mary	mid- dle	high	other	age (log)	size (log)	urb- an	% PDF	# wor- ds (K)	RLA blur[b]	math blur[b]
Mean	.133	.483	.402	.555	.650	.147	.335	.467	.099	.219	.215	1.73	5.64	.571	.005	23.8	6.97	7.33
Median	.129	.470	.370	.597	.739	.141	.311	.000	.000	.000	.000	1.95	5.75	1.00	.000	8.81	5.00	5.00
Standard error	.120	.231	.245	.325	.324	.075	.200	.499	.299	.413	.411	.969	.968	.495	.047	142	10.4	10.8
IBL emphasis	1																	
% proficiency RLA	.10*	1																
% proficiency math	.05*	.80*	1															
% student poverty	-.18*	-.35*	-.34*	1														
% students of color	-.17*	-.28*	-.17*	.53*	1													
% district poverty	-.11*	-.13*	-.11*	.42*	.47*	1												
% district POC	-.06*	-.18*	-.13*	.27*	.62*	.56*	1											
Primary school [a]	.06*	.02	.11*	-.04*	.05*	.01	.07*	1										
Middle school [a]	-.06*	-.02	.01	.10*	.12*	.02	.07*	-.31*	1									
High school [a]	-.06*	.00	-.09*	.01	.00	.02	-.05*	-.50*	-.18*	1								
Other grade range	.04*	-.01	-.05*	-.04*	-.14*	-.04*	-.08*	-.49*	-.17*	-.28*	1							
Years open (log)	.01	.06*	.01	-.02	-.11*	.00	-.07*	.02	-.10*	.03*	.03*	1						
# students (log)	.09*	.17*	.14*	-.03*	.16*	.06*	.18*	.08*	-.07*	-.20*	.16*	.17*	1					
Urban locale	-.03*	-.07*	-.01	.23*	.46*	.38*	.43*	.03	.08*	.01	-.10*	-.06*	.08*	1				
% PDF web pages	.05*	.01	.00	-.02	-.01	-.02	-.02	-.02	.03*	.00	.01	.02	.01	-.03*	1			
# words	.20*	.03*	.04*	-.04*	-.02	-.03*	.00	-.02	-.01	.01	.02	-.03*	-.01	-.01	.06*	1		
RLA blurring	-.06*	-.09*	-.11*	.00	-.09*	-.03*	-.10*	-.13*	-.10*	.30*	-.07*	-.11*	-.56*	-.05*	-.01	-.02	1	
Math blurring	-.08*	-.08*	-.08*	.01	-.08*	-.02	-.11*	-.15*	-.11*	.34*	-.08*	-.11*	-.56*	-.05*	-.02	-.02	.93*	1

Sources: American Community Survey 2012-16 (U.S. Census Bureau 2018), Common Core of Data 2015-16 (NCES 2018a), EdFacts Achievement Results for State Assessments (USDE 2018), and the author's data collection and calculations.

NOTES: [a] Binary indicators of grade range served; the baseline is "other" (incl. ungraded). [b] This indicates the degree of data "blurring" by USDE to protect student groups' privacy. Higher blurring reflects less precise data.

* $p<0.05$

Acronyms: IBL = inquiry-based learning; RLA = reading/language arts; SOC = students of color; POC = people of color; SD = school district; pov = poverty.

Table 2. Mixed Linear Models Predicting IBL Emphasis with School and School District Poverty and Race.

Outcome: IBL emphasis Independent variable	Model 1a: Controls only	Model 1b: School poverty	Model 1c: School race	Model 1d: School district poverty	Model 1e: School district race
Poverty & race					
% students in poverty		-0.000617***			
		(5.00e-05)			
% students of color			-0.0733***		
			(0.00539)		
% district in poverty				-0.212***	
				(0.0221)	
% district people of color					-0.0302***
					(0.00869)
School controls					
Primary school[a]	0.000856	0.000188	0.00602	0.000415	0.00162
	(0.00394)	(0.00389)	(0.00390)	(0.00391)	(0.00395)
Middle school[a]	-0.0177**	-0.0155**	-0.00918	-0.0183**	-0.0164**
	(0.00590)	(0.00582)	(0.00584)	(0.00585)	(0.00590)
High school[a]	-0.0134**	-0.0137**	-0.00784	-0.0125**	-0.0129**
	(0.00472)	(0.00466)	(0.00467)	(0.00469)	(0.00472)
Years open (log)	-0.00396*	-0.00370*	-0.00607***	-0.00376*	-0.00441**
	(0.00163)	(0.00161)	(0.00161)	(0.00162)	(0.00163)
Number students (log)	0.00916***	0.00801***	0.0124***	0.00953***	0.0102***
	(0.00170)	(0.00168)	(0.00169)	(0.00169)	(0.00172)
Urban locale (binary)	0.000541	0.00850**	0.0188***	0.0117***	0.00503
	(0.00309)	(0.00312)	(0.00333)	(0.00328)	(0.00335)
% PDF web pages	0.115***	0.112***	0.114***	0.113***	0.114***
	(0.0319)	(0.0315)	(0.0314)	(0.0316)	(0.0319)
Model parameters					
Constant	0.0661***	0.106***	0.0908***	0.0905***	0.0692***
	(0.0116)	(0.0120)	(0.0116)	(0.0118)	(0.0116)
Variance (σ^2) between CMOs	0.00583	0.00592	0.00570	0.00592	0.00581
Residual variance (σ^2)	0.0117	0.0113	0.0113	0.0115	0.0117
CMO ICC[b] (ρ)	0.343	0.335	0.337	0.340	0.333
Number of CMOs	377	377	377	377	377
Number of observations	5,784	5,784	5,784	5,784	5,784

Model fit					
Log likelihood	4456	4536	4547	4502	4462
Degrees of freedom	10	11	11	11	11
AIC	-8892	-9050	-9072	-8982	-8902
BIC	-8825	-8976	-8999	-8909	-8829
Wald χ^2 [c]	78.1***	233***	265***	173***	90.3***
RE test: χ^2 [d]	666***	663***	643***	696***	654***

Sources: American Community Survey 2012-16 (U.S. Census Bureau 2018), Common Core of Data 2015-16 (NCES 2018a), and the author's data collection and calculations.
Notes: [a] Binary indicators of grade range served; the baseline is "other" (incl. ungraded). [b] The Intraclass Correlation Coefficient (ICC) or *rho* (ρ) here measures nesting of the outcome in charter management organizations (CMOs). [c] The Wald χ^2 statistic tests the null hypothesis that all regression coefficients are zero. [d] This conservative χ^2 test assesses the null hypothesis that all random effects (RE) in the model are zero; thus, significance supports the model. Standard errors in parentheses.
*** p<0.001, ** p<0.01, * p<0.05

Table 3. Mixed Linear Models Predicting School Poverty and Race with IBL Emphasis and Academic Achievement.

	Number of students in poverty				Number of students of color			
Independent variable	M2a: Controls only	M2b: IBL emphasis	M2c: Academic achievement	M2d: Fully specified	M3a: Controls only	M3b: IBL emphasis	M3c: Academic achievemen t	M3d: Fully specified
Ideology & academic quality								
IBL emphasis		-0.298***		-0.204***		-0.292***		-0.217***
		(0.0306)		(0.0286)		(0.0233)		(0.0221)
RLA proficiency[a]			-0.450***	-0.430***			-0.335***	-0.314***
			(0.0302)	(0.0304)			(0.0229)	(0.0229)
Math proficiency			-0.0587*	-0.0678*			-0.0561*	-0.0639**
			(0.0298)	(0.0297)			(0.0232)	(0.0230)
School controls								
Primary school[b]	-0.00429	-0.00318	0.00303	0.00391	0.0448***	0.0455***	0.0518***	0.0522***
	(0.00953)	(0.00946)	(0.00901)	(0.00897)	(0.00724)	(0.00714)	(0.00682)	(0.00675)
Middle school[b]	0.0336*	0.0297*	0.0396**	0.0368**	0.0704***	0.0663***	0.0783***	0.0747***
	(0.0141)	(0.0140)	(0.0134)	(0.0133)	(0.0107)	(0.0105)	(0.0101)	(0.0100)
High school[b]	-0.0110	-0.0135	0.00416	0.00231	0.0566***	0.0541***	0.0659***	0.0638***
	(0.0113)	(0.0112)	(0.0108)	(0.0108)	(0.00860)	(0.00848)	(0.00824)	(0.00816)
Years open (log)	0.00167	0.000751	0.00871*	0.00780*	-0.0159***	-0.0167***	-0.00907**	- 0.00992***
	(0.00399)	(0.00397)	(0.00376)	(0.00375)	(0.00306)	(0.00302)	(0.00289)	(0.00286)
Number students (log)	-0.0165***	-0.0131**	-0.000526	0.00120	0.00481	0.00805*	0.0222***	0.0237***
	(0.00446)	(0.00443)	(0.00512)	(0.00511)	(0.00337)	(0.00333)	(0.00390)	(0.00386)
Urban locale	0.0687***	0.0710***	0.0627***	0.0646***	0.107***	0.110***	0.0990***	0.101***
	(0.0110)	(0.0109)	(0.0102)	(0.0101)	(0.00919)	(0.00908)	(0.00859)	(0.00852)
% PDF web pages		0.0590		0.0469		0.104		0.101
		(0.0777)		(0.0726)		(0.0601)		(0.0565)
RLA blurring[c]			-3.86e-05	0.000124			0.000832	0.000962
			(0.000941)	(0.000938)			(0.000768)	(0.000757)
Math blurring[c]			-0.000430	-0.000633			-0.000530	-0.000706
			(0.000929)	(0.000927)			(0.000740)	(0.000730)
Model parameters								
Constant	0.568***	0.591***	0.716***	0.730***	0.432***	0.457***	0.507***	0.523***
	(0.0266)	(0.0265)	(0.0327)	(0.0327)	(0.0353)	(0.0349)	(0.0374)	(0.0371)

Variance (σ^2) between states					0.0304	0.0296	0.0292	0.0287
Variance (σ^2) between school districts	0.0364	0.0347	0.0305	0.0294	0.0400	0.0397	0.0356	0.0356
Residual variance (σ^2)	0.0571	0.0563	0.0482	0.0480	0.0332	0.0322	0.0284	0.0278
State ICC[d] (ρ)					0.294	0.292	0.317	0.315
School district ICC[d] (ρ)	0.382	0.373	0.380	0.372	0.386	0.391	0.345	0.345
Number of states					43	43	43	43
Number of school districts	1,481	1,481	1,481	1,481	1,481	1,481	1,481	1,481
Number of observations	5,784	5,784	5,784	5,784	5,784	5,784	5,784	5,784
Model fit								
Log likelihood	-594	-543	-114	-87.5	638	716	1076	1123
Degrees of freedom	9	11	13	15	10	12	14	16
AIC	1206	1108	255	205	-1255	-1408	-2124	-2215
BIC	1266	1181	341	305	-1189	-1328	-2030	-2108
Wald χ^2 [e]	67.2***	164***	1157***	1219***	234***	397***	1225***	1344***
RE test: χ^2 [f]	1892***	1832***	2038***	1956***	2992***	2944***	3313***	3250***

Sources: American Community Survey 2012-16 (U.S. Census Bureau 2018), Common Core of Data 2015-16 (NCES 2018a), EdFacts Achievement Results for State Assessments (USDE 2018), and the author's data collection and calculations.

Notes: [a] School proficiency rates in state assessments of reading/language arts (RLA). [b] Binary indicators of grade range served; the baseline is "other" (incl. ungraded). [c] This indicates the degree of data "blurring" by USDE to protect student groups' privacy. This score ranges from 1 (percentiles reported, e.g. 95% academic proficiency for a school) to 50 (medians reported, e.g. 50-100% proficiency), with higher blurring reflecting less precise data. [d] The Intraclass Correlation Coefficient (ICC) or *rho* (ρ) here measures nesting of the outcome in states and school districts. [e] The Wald χ^2 statistic tests the null hypothesis that all regression coefficients are zero. [f] This conservative χ^2 test assesses the null hypothesis that all random effects (RE) in the model are zero; thus, significance supports the model. Standard errors in parentheses.

*** p<0.001, ** p<0.01, * p<0.05

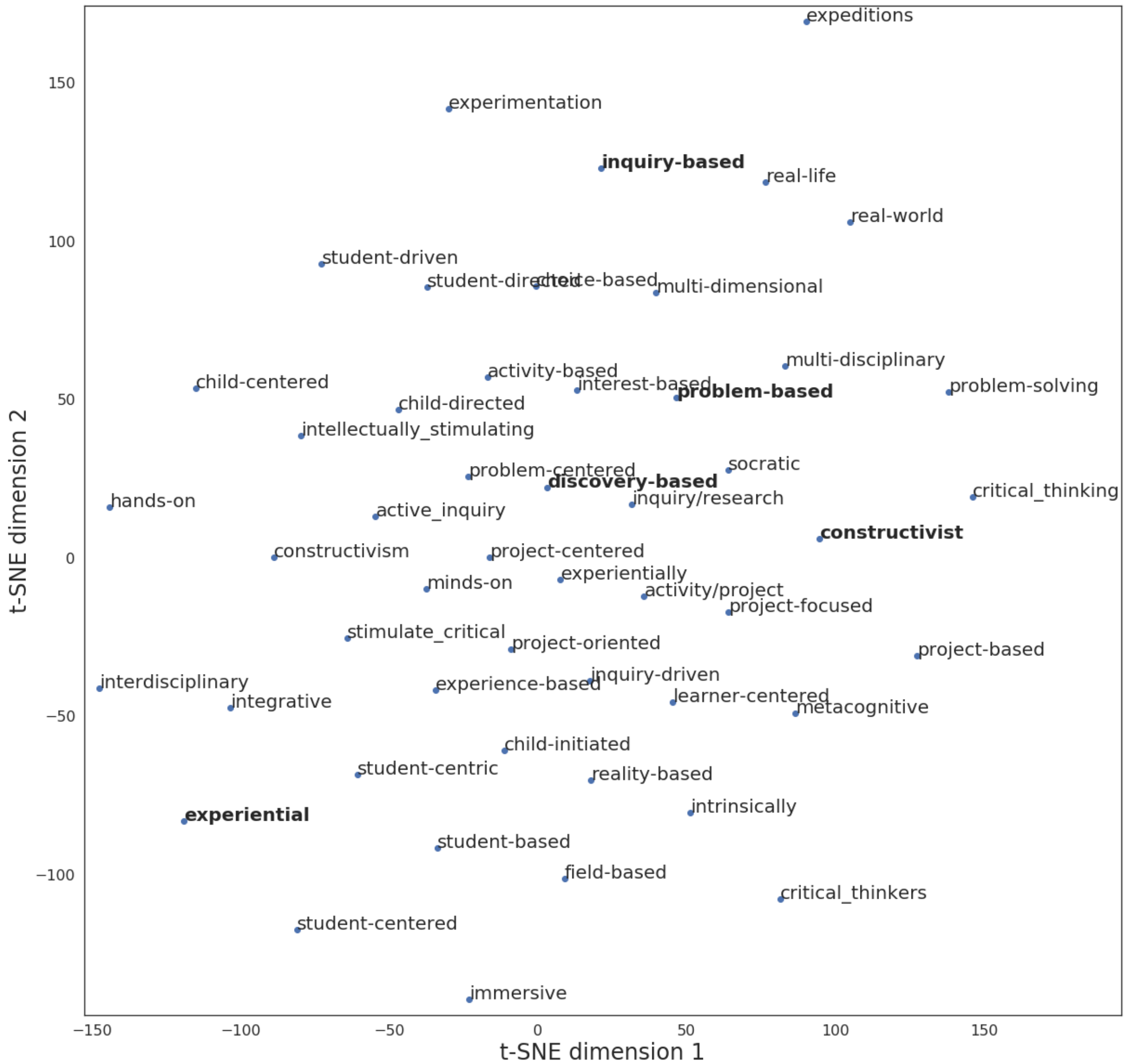

Figure 1. *Visualization of IBL dictionary in vector space*. This is a projection of the 300-dimension word embedding vector space into two dimensions through t-Distributed Stochastic Neighbor Embedding (t-SNE), a non-parametric means for visualizing high-dimensional data (Maaten and Hinton 2008). The 50 points represent word vectors from the dictionary for inquiry-based learning (IBL), with the five seed terms in **bold.** Two terms overlap above the graph's center: these are "student-directed" with "choice-based". Pointwise positions are based on cosine distances and preserve local structure, such that points close in high-dimensional space remain close in low-dimensional space. The axes are an artifact of dimensionality reduction and are not directly interpretable. (See text for explanation of cosine similarity scores and dictionary development.)

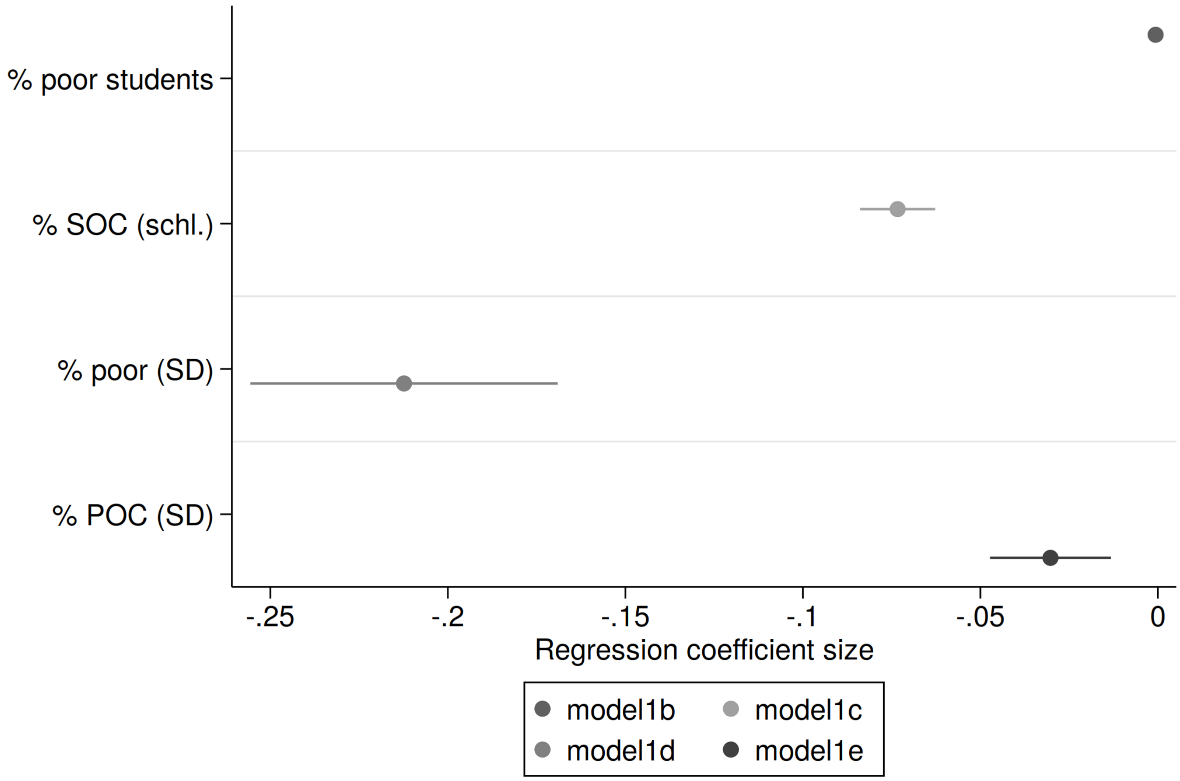

Figure 2. *Results of regressing IBL emphasis on school and social district poverty and race (corresponding to table 2)*. Each independent variable is modeled separately, and all effects are statistically significant. For clarity, not shown are controls: School level (dummies for primary, middle, high), years open (logged), number of students (logged), urban locale, and % PDF web pages.
Acronyms: IBL = inquiry-based learning; SOC = students of color; POC = people of color; SD = school district.

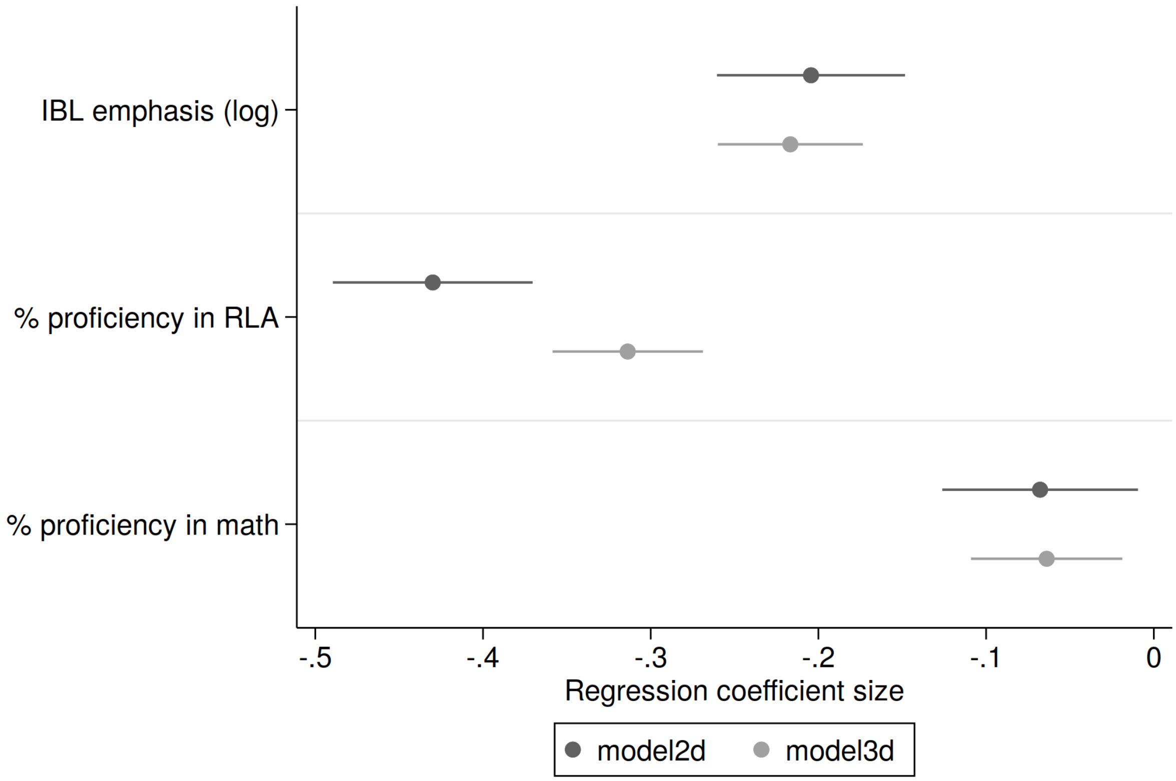

Figure 3. *Results of regressing school and social district poverty and race on IBL emphasis and academic proficiency (corresponding to table 3)*. The dependent variable of model 2d (upper dot) is % poor students; for model 3d (lower dot), it is % students of color. All effects are statistically significant. For clarity, not shown are controls: School level (dummies for primary, middle, high), years open (logged), number of students (logged), urban locale, percent PDF web pages, and blurring of RLA and math proficiency rates.
Acronyms: IBL = inquiry-based learning; RLA = reading/language arts

APPENDIX B: EXAMPLES OF CHARTER SCHOOL WEBSITES
To appear in print.

My web-crawling workflow had three steps: collect URLs, crawl web pages, and parse and filter text.

First, given that no comprehensive, reliable list of charter school URLs exists , I used the Google Places API (free for researchers: https://cloud.google.com/maps-platform/places/) to collect URLs. I specifically searched for the name and full address of each charter school, which tended best to distinguish schools and to return distinct, accurate URLs. My research team and I then cleaned this URL list by hand to identify the domain (i.e., URL) most specific to each school that also contained information on its mission, values, etc.—either on the domain itself (e.g., *www.school.com*) or in its subdomains (e.g., *www.school.com/about-us*).

I specifically used Scrapy Cluster, a scalable crawling architecture for Python that coordinates multiple web spiders (IST Research Corporation 2017). For each domain, I crawled all subdomains by following page links restricted to the root domain (e.g., *www.school.com* could link one level deeper to *www.school.com/page*, but not outside to *www.faceblast.com/school*), followed all links on those subdomains, and so on to a maximum crawling depth of 10 (most sites were not this deep). I crawled each page only once. Each PDF file (if any) was converted into text and considered a page.

Because charter school websites are diverse and lack a consistent, coherent structure, I used `BeautifulSoup` (Richardson 2007) to clean website text using simple HTML parsing: I removed inline or formatting tags (e.g., `span`, `strong`) and non-visible tags (e.g., `style`, `head`), leaving only visible text as would be viewed in a web browser. For this analysis, to represent each school I joined all its crawled, parsed pages into a single string of text. To accurately represent charter schools in their organizational contexts, I also did web searched to create a current, complete directory of CMOs, their URLs, and the schools they manage.

APPENDIX B: EXAMPLES OF CHARTER SCHOOL WEBSITES
To appear in print.

To create a current, complete directory of CMOs, their URLs, and the schools they manage, I first merged two lists of CMOs: one from a recent report (Woodworth et al. 2017), the other shared (upon request) by the National Alliance for Public Charter Schools (https://www.publiccharters.org/). I then manually cleaned and extended this list through web searches, and finally checked CMO websites to enumerate their schools.

REFERENCES (Appendix A):

IST Research Corporation. 2017. *Scrapy Cluster*. Fredericksburg, VA.

Richardson, Leonard. 2007. *Beautiful Soup Documentation*. New York, NY.

Woodworth, James L., Margaret E. Raymond, Chunping Han, Negasi Yohannes, Richardson W. Payton, and Will Snow. 2017. *Charter Management Organizations*. Stanford, CA: Center for Research on Education Outcomes, Stanford University.

APPENDIX B: EXAMPLES OF CHARTER SCHOOL WEBSITES
To appear online.

LOW in inquiry-based learning (IBL) emphasis: Phoenix College Preparatory Academy, Phoenix, AZ (https://www.phoenixcollege.edu/pc-prep-academy)

High school and college in one...

The Phoenix College Preparatory Academy (PCPA) is a charter high school accredited by the North Central Association Commission on Accreditation and School Improvement (NCA CASI), an accreditation division of AdvancED.

Students at our academy attend classes on the beautiful campus of Phoenix College, and in addition to the resources of the high school, students have access to the colleges' computer labs, libraries and other services and facilities. PCPA students have the opportunity to interact with community college students who serve as both mentors and tutors.

PCPA maintains the highest academic standards to ensure students meet all state requirements for a high school diploma. It is also possible that by graduation students can complete several community college courses.

IT'S LIKE RECEIVING A PRIVATE EDUCATION FOR FREE....

As a public charter high school, Phoenix College Preparatory Academy adheres to the open enrollment policies prescribed by the Arizona Department of Education. PCPA does not charge tuition for high school classes. Students in good academic standing may qualify for free college tuition as well.

Why attend our academy?

- Individualized instruction from state certified and/or highly qualified teachers in a small class setting
- Obtain your high school diploma
- Join in community college extra-curricular activities & organizations
- Access to college resources and facilities, including free tutoring
- May qualify to receive free tuition for college courses at any of the Maricopa Community Colleges

Phoenix College Preparatory Academy is a free public high school and does not require any citizenship or immigration status information or documentation to enroll students into high school classes. Students are not required to take college courses and are enrolled on a first come, first served, basis.

Mission Statement
Through a shared vision, Phoenix College Preparatory Academy, supported by Phoenix College, is committed to creating and sustaining a community where all learners will pursue high standards to succeed in college and career

Vision

- Every PCPA student will complete the general high school requirements in order to be admitted to a four year postsecondary institution
- Each student will have the opportunity to earn at least 30 college credits, or complete one year of college credit requirements
- 20% of students will complete their Associate's degree prior to high school graduation

----

HIGH in IBL emphasis: Anne Frank Inspire Academy, San Antonio, TX (https://www.braination.net/Anne-Frank-Inspire-Academy)

A Preeminent 21st Century School K–12

How do you make a free, public, 21st century school that is caring, creative, fun, innovative, and sustainable? How do you ensure that school retains a rigorous curriculum that pushes every student to be and do his or her best? You create an Inspire Academy!

The curriculum at our Inspire Academies adheres to state standards in all subjects, including the core areas of language arts, mathematics, science, and social studies. But in order to fully address student needs, we have designed a "three tiered curriculum" that not only meets state standards but also helps students follow their passions and become self-directed learners.

Students explore information daily in great depth via projects, field experiences, problem solving, collaborative groups, oral and written communication, entrepreneurialism, and much more, using a skill mastery approach. This approach creates a dynamic learning experience that allows independent thinking and problem solving to develop and flourish.

Our Program

Our curriculum includes:

- Core
    - A core subjects curriculum based on Texas' TEKS standards
- Choice
    - A personalized electives curriculum designed from the input of our students and parents
- Exposure
    - An experiential curriculum highlighting things and events all children should be able to experience without regard to social class or income
- AFIA Curriculum Design
    - K–5 Design
    - 6–8 Design
    - 9–12 Design

In addition to our regular curriculum, students receive unique learning opportunities through our community mentorship program that reflect our four interlocking areas of growth. At Anne Frank, we work hard to:

- Become an expert learner (with problem solving and creative thinking skills)
- Develop leadership skills
- Become a person of principle and character
- Make the world a better place through service

Our Facility

Our schools are designed to create a small, family-like environment that allows each student and teacher to know one another and ensure that no one "falls through the cracks." By partnering with award-winning, international architects at Fielding Nair, we've designed an incredible campus that includes eight outdoor learning areas (including a tree house and amphitheater); a variety of creative, indoor learning studios (like MakerSpaces and Yoga studios); and zero hallways! All of our facilities are also equipped with seamless technology via SkyDrives.

Our Mission & Vision

Mission: To increase the capacity for human greatness.
Vision: Creating 21st century learning models for use around the world through leadership, innovation, safety, technology, integration, synergy, and transformative power.

Our Core Values

Value gives meaning to what we do. But at Anne Frank, our core values of **innovation**, **embracing greatness**, **integrity**, and **joy** are the driving force behind every decision we make.
At Anne Frank, we are active members of our community—**We Belong**. We are becoming experts in every life arena—We can **Be Great**. And we know there is purpose in life—Therefore, we choose to **Find Joy**. In short, we believe students can be self-directed learners, and schools can be both fun and challenging.

APPENDIX C: COUNTS AND SIMILARITIES FOR IBL DICTIONARY TERMS

To appear in print.

Notes: The 5 **bolded** (seed) terms were used to make the full 50-term dictionary. The 15 terms in *italics* are part of the 20-term narrow IBL dictionary (plus the seed terms) used for validation.

Term	Frequency in corpus	Cosine similarity to seed terms
inquiry-based	1258	0.8504
problem-based	198	0.8673
discovery-based	29	0.8212
experiential	1916	0.8053
constructivist	251	0.8219
hands-on	48423	0.6635
problem-solving	7125	0.5836
critical thinking	6647	0.6942
real-world	6397	0.6002
project-based	4129	0.7535
real-life	2209	0.5789
interdisciplinary	1565	0.6910
student-centered	1468	0.6772
critical thinkers	1061	0.6236
expeditions	675	0.5807
child-centered	654	0.7195
experimentation	474	0.6129
student-based	355	0.7364
immersive	295	0.6753
activity-based	260	0.7891
student-driven	238	0.7030
intrinsically	225	0.7452
reality-based	196	0.7220
multi-disciplinary	173	0.7260
learner-centered	125	0.7720
interest-based	118	0.7576
minds-on	117	0.7786
metacognitive	111	0.6791
integrative	109	0.7554
experience-based	102	0.7552
multi-dimensional	99	0.7053
constructivism	97	0.7381
student-directed	74	0.7041
choice-based	58	0.7599
project-oriented	45	0.7552
intellectually stimulating	42	0.7255
inquiry/research	38	0.7239
field-based	37	0.7378
project-focused	36	0.7200
stimulate critical	36	0.7296
student-centric	36	0.7528
active inquiry	26	0.8035
inquiry-driven	25	0.8068
child-directed	24	0.7756
child-initiated	23	0.7571
activity/project	23	0.7075
experientially	20	0.7862
socratic	14	0.7994
problem-centered	12	0.7787
project-centered	10	0.8219

APPENDIX D: TECHNICAL NOTES ON WORD EMBEDDINGS
To appear online.

While a few alternative model architectures have been proposed for neural-net word embedding models, in this paper I use the "continuous skip-gram" model, which is the most accurate option for semantic comparisons (Mikolov, Chen, et al. 2013)—the focus of this paper—and is best suited for large data sets (TensorFlow 2018) like my charter schools database. Continuous skip-gram seeks to classify a word given each other word in its context, allowing each word vector to have a separate impact on its neighbors.

In my implementation of word embeddings, I take advantage of several extensions to the original continuous skip-gram model: noise reduction, undersampling of frequent words, and detection of common phrases in the corpus (Mikolov, Sutskever, et al. 2013). Common phrases, also called multi-word expressions, often possess unique meanings compared to their individual words; for instance, compare the meaning of the phrase "special education" to the individual terms "special" and "education". My model accepts phrases so long as their 'collocation score'—the ratio of a given word pair's collocations divided by the product of each word's individual appearances—exceeds some threshold. Mathematically, this score is derived as follows:

$$\mathrm{score}(w_i, w_j) = \frac{\mathrm{count}(w_i w_j) - \delta}{\mathrm{count}(w_i) \times \mathrm{count}(w_j)} \tag{3}$$

where w_i and w_j are two words in the corpus, $w_i w_j$ is their collocation, and δ is a "discounting coefficient" that makes infrequent phrases less likely (Mikolov, Sutskever, et al. 2013:6).

Computationally, word embeddings learn word vectors using a 'neural network', a machine learning architecture that tries to predict (in this case) the connections between words using a sample of observations of word connections or 'training words' (e.g., Turian, Ratinov, and Bengio 2010). In technical terms, word embeddings project the semantic connections observed in the sample (using initially random weights) through a hidden layer and into an output matrix, the error of which is used to update and optimize the weights over several iterations. The size of the hidden layer equals k, the number of vector space dimensions in the model.

Additionally, in calculating similarities, word2vec normalizes all vectors to a length of 1, essentially projecting all word vectors along the surface of a high-dimensional sphere. Unit normalization improves computational efficiency (Wilson and Schakel 2015), but it may lose some information on the consistency with which words appear in their contexts (Schakel and Wilson 2015).

For purposes of efficiency, I use the `gensim` defaults for some parameters: 10 iterations through the data, 5 noise words (for noise reduction), an initial learning rate of 0.025, and a discounting coefficient of 3. To improve the model's ability to capture semantic nuances—without inducing unnecessary computational burden—I expand the defaults for a few other parameters: word context windows of size eight (to better capture syntactic relations: Mikolov, Chen, et al. 2013; Spirling and Rodriguez 2019), 20000 words per sample, a hidden layer of size 300, and a phrase detection threshold of 8.0 (to better capture candidate multi-word expressions). As suggested in Mikolov et al. (2013:4), I use a negative sampling exponent of 0.75 to subsample frequent words.

Finally, while the number of dimensions, k, in the word embedding vector space is high (typically 100-300), this is nonetheless far smaller than the number of words in the corpus. Indeed, the dimensionality is low compared to the older and less efficient (though still common) term-document matrix: words are represented as rows, documents as columns, and each entry as the frequency or weight of a given term in a given document (Salton, Wong, and Yang 1975). For a vocabulary of size N and corpus of D documents, word embeddings represent all words and their associations using N x k dimensions, while the document-term matrix represents words using $N \times D$ dimensions. Even for a relatively small corpus like 6,300 websites, the latter matrix is enormously larger; consequently, working with the former matrix requires significantly less time and computational resources.

REFERENCES (Appendix D):

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. "Efficient Estimation of Word Representations in Vector Space." in *International Conference on Learning Representations Workshop.*

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. 2013. "Distributed Representations of Words and Phrases and Their Compositionality." Advances in Neural *Information Processing Systems 3111–3119.*

Salton, G., A. Wong, and C. S. Yang. 1975. "A Vector Space Model for Automatic Indexing." *Communications of the ACM* 18(11):613–620.

Schakel, Adriaan M. J. and Benjamin J. Wilson. 2015. "Measuring Word Significance Using Distributed Representations of Words." *ArXiv:1508.02297 [Cs].*

Spirling, Arthur and Pedro L. Rodriguez. 2019. "Word Embeddings: What Works, What Doesn't, and How to Tell the Difference for Applied Research."

TensorFlow. 2018. "Vector Representations of Words." *TensorFlow*. Retrieved September 21, 2018 (https://www.tensorflow.org/tutorials/representation/word2vec).

Turian, Joseph, Lev-Arie Ratinov, and Yoshua Bengio. 2010. "Word Representations: A Simple and General Method for Semi-Supervised Learning." Pp. 384–394 in *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. Uppsala, Sweden: Association for Computational Linguistics.

Wilson, Benjamin J. and Adriaan M. J. Schakel. 2015. "Controlled Experiments for Word Embeddings." *ArXiv:1510.02675 [Cs].*

APPENDIX E: ROBUSTNESS CHECKS AND VALIDATION
To appear in print.

The impact of objective quality signals may be greater when using further lagged predictors, for then parents and resource providers have longer than one year to observe and react to these signals. To test this possibility, I replicated the above analyses using academic proficiency rates for 2013-14 instead of 2014-15. These results were not significantly different from those reported above. Similarly, because ideologically distinct schools' may attract students outside their district catchment area, the impact of school enrollment demographics may be better revealed when operationalized relative to their surrounding district. As such, I re-ran the analyses using not within-school demographic proportions (e.g., percent students of color) but instead differentials between the school and district demographics (e.g., percent students of color at school level subtracted from proportion people of color at district level). Math proficiency grew in size (from 0.059 to 0.108 in absolute value) and statistical significance (from $p<0.05$ to $p<0.001$) when predicting the school poverty differential, but the results were otherwise unchanged.

To test whether the above results were skewed by schools with less precise academic data (perhaps due to misspecification), I also re-ran these analyses using only those 1,982 schools with proficiency scores reported in percentiles (the most precise). While math proficiency changed sign and largely became statistically insignificant, the direction, approximate effect size, and ratio between IBL emphasis and academic proficiency measures were otherwise consistent with the above.

Outliers could have also skewed the above results. To test the possibility that the above results are biased by very large websites (100 web pages or longer) or small schools (10 students or less), I also replicated the above analyses with filtered data. Thus, I replicated once for data with very large websites removed and once for data with small schools removed. Although math proficiency became marginally statistically significant when large websites were removed, the results were otherwise unchanged.

Likewise, given that IBL is more common in schools with more affluent, white students, the above results may not represent schools in more poor, racially diverse areas. To evaluate whether my results are thus limited in geographic reach, I repeated the analyses after filtering to only those schools located in school districts with above-median disadvantage by three separate metrics: proportion in poverty, proportion people of color, or population density (a continuous measure of urban status). I analyzed each filtered data set separately, amounting to three additional batches of models. Except for declining significance of those variables used for filtering—e.g., school district poverty became insignificant for the data set filtered to only those schools in districts with above-median poverty—and marginal significance for math proficiency in some cases, all these results were consistent with the above.

Moreover, to test the possibility that inclusion of full nesting would change the results, I replicated all models using random effects for CMO, state, and school district. The results were essentially identical—except that math proficiency was more significant when predicting school poverty ($p < 0.000$ instead of $p < 0.05$).

Finally, the particular words chosen for the IBL dictionary may be driving the results. In particular, the term "hands-on" occurs 48,423 times—nearly seven times as frequently as the next most common word in the dictionary, "problem-solving", at 7,125 times. Accordingly, I test the robustness of the above findings by calculating Emphasis and running mixed linear models for three additional dictionaries: the five seed terms; a 20-term, theoretically narrow IBL dictionary including only synonyms of the seed terms (a subset of the full dictionary); and the full dictionary with the "hands-on" term removed, leaving 49 terms (see Appendix C for detail on these dictionaries).

The results of these additional models strongly support the results discussed above. When predicting IBL, schools are still nested in CMOs (though less so for the seed dictionary: $\rho = 0.18$) but not states or districts, and the only key covariate (i.e., school and school district race and poverty) that loses statistical significance is school district proportion people of color ($p < 0.226$) when using seed terms. Otherwise, the only change for key covariates when predicting IBL is a decrease in effect size proportional (though not linearly) to dictionary size. Specifically, beta coefficients are 3-4 times smaller when seed terms are used (except for school district race, which is even smaller but not significant); a quarter to a third smaller when the narrow, 20-term dictionary is used; and almost identical when only "hands-on" is removed. This shrinking of betas is to be expected given that total word count decreases proportional to dictionary length—though it continues to discriminate effectively between educational ideologies, as consistent statistical significance indicates.

When predicting school poverty and race, the only significant change is an increase in betas for IBL, while the betas for academic proficiency are static. Specifically, betas roughly double for the seed dictionary, increase 1.2-1.5 times for the narrow dictionary, and are (again) almost identical for the 49-term dictionary. The inverse relationship between dictionary size and IBL betas suggests that shorter, conceptually sharper dictionaries may capture the clearest signal of organizational identity, and thus better explain demographic outcomes.