RSF: The Russell Sage Foundation Journal of the Social Sciences

Using Administrative Data for Science and Policy

VOLUME 5, NUMBER 3, MARCH 2019

copublished with The Bill & Melinda Gates Foundation

RSF: The Russell Sage Foundation Journal of the Social Sciences

Using Administrative Data for Science and Policy

VOLUME 5, NUMBER 3, MARCH 2019

RSF: The Russell Sage Foundation Journal of the Social Sciences ISSN 2377-8261

The Russell Sage Foundation

The Russell Sage Foundation, one of the oldest of America's general purpose foundations, was established in 1907 by Mrs. Margaret Olivia Sage for "the improvement of social and living conditions in the United States." The foundation seeks to fulfill this mandate by fostering the development and dissemination of knowledge about the country's political, social, and economic problems. While the foundation endeavors to assure the accuracy and objectivity of each book it publishes, the conclusions and interpretations in Russell Sage Foundation publications are those of the authors and not of the foundation, its trustees, or its staff. Publication by Russell Sage, therefore, does not imply foundation endorsement.

Mission Statement

RSF: The Russell Sage Foundation Journal of the Social Sciences is a peer-reviewed, open-access journal of original empirical research articles by both established and emerging scholars. It is designed to promote cross-disciplinary collaborations on timely issues of interest to academics, policymakers, and the public at large. Each issue is thematic in nature and focuses on a specific research question or area of interest. The introduction to each issue will include an accessible, broad, and synthetic overview of the research question under consideration and the current thinking from the various social sciences.

Opinions expressed in this journal are not necessarily those of the editors, editorial board, trustees, or the Russell Sage Foundation.

We invite scholars to submit proposals for potential issues through the *RSF* application portal: rsf.fluxx.io. Submissions should be addressed to Suzanne Nichols, Director of Publications.

To view the complete text and additional features online please go to **www.rsfjournal.org**.

Open Access Policy

RSF: The Russell Sage Foundation Journal of the Social Sciences is an open access journal. It is published under a Creative Commons Attribution-NonCommercial-No Derivs 3.0 Unported License.

Russell Sage Foundation
112 East 64th Street
New York, NY 10065

ISSN (print): **2377-8253**
ISSN (electronic): **2377-8261**
ISBN: **978-0-87154-760-6**

The Bill & Melinda Gates Foundation

Guided by the belief that every life has equal value, the Bill & Melinda Gates Foundation works to help all people lead healthy, productive lives. In developing countries, we focus on improving people's health and giving them the chance to lift themselves out of hunger and extreme poverty. In the United States, we seek to ensure that all people—especially those with the fewest resources—have access to the opportunities they need to succeed in school and life.

The foundation's U.S. Economic Mobility and Opportunity Program seeks to ensure that there are more public, private, philanthropic, and academic actors working together over the next decade to dramatically increase economic mobility. We are working to increase access to better information and tools needed to tackle barriers to opportunity and develop mobility strategies at the community, state, and national levels. We are committed to partnership, innovation, research, and programmatic investments that will help more Americans to climb the economic ladder and lead fulfilling, dignified lives.

RSF: The Russell Sage Foundation Journal of the Social Sciences

VOLUME 5, NUMBER 3, MARCH 2019

Using Administrative Data for Science and Policy

ISSUE EDITORS
Andrew M. Penner, University of California, Irvine
Kenneth A. Dodge, Duke University

CONTENTS

Using Administrative Data for Social Science and Policy

ANDREW M. PENNER AND KENNETH A. DODGE

Several years ago, the director of social services in Durham, North Carolina, came to our Durham Connects Program research team with a request for collaboration. He administered several dozen "pockets" of dollars allocated for families with young children for programs such as childcare subsidies, housing loans, and Early Head Start. His administrative data systems detailed which families received the dollars, but the files were not linked with each other and had not been used to improve accountability or impact. On our side, we had been collecting interview data from the entire population of families at birth and therefore knew what resources families needed to support their infants' healthy development. We teamed up with him, merged and analyzed the combined data files, and gained many insights (for example, some families were receiving many thousands of dollars in support but still living in poverty).

The findings were often translated into public policy. For example, the population's need for substance abuse treatment for mothers of infants far exceeded the capacity of the community's allocated resources in this domain, and so the juxtaposition of the need and the capacity was presented to the county commissioners, who then increased their allocation. In another example, a high no-show rate at the community health clinic was discovered and thought to impede delivery of health care and cost many dollars in inefficiency; interviews with mothers showed that they failed to show up because the bus did not stop in the housing project and was too expensive. So the City Council changed the bus route to be more accessible and passed an ordinance to allow anyone carrying a baby to ride the city bus for free. No-shows at the clinic declined. A bonus of this collaboration was that in return for providing useful research findings to the community, university scholars were allowed to use the data files for their research.

This case study illustrates the tremendous potential of a university-community collaboration using administrative data. In addition to the profound human service gains, administrative data infrastructure has important potential to improve cutting-edge social science research. Not only do administrative data facilitate the

Andrew M. Penner is professor of sociology at the University of California, Irvine. **Kenneth A. Dodge** is Pritzker Professor of Early Learning Policy Studies and professor of psychology and neuroscience at Duke University.

 Penner, Andrew M., and Kenneth A. Dodge. 2019. "Using Administrative Data for Social Science and Policy." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(3): 1–18. DOI: 10.7758/RSF.2019.5.3.01. We are grateful to Nicole Deterding, Thad Domina, David Grusky, Paul Hanselman, NaYoung Hwang, Ryan Lewis, Silvia Melzer, Matthew Snipp, and participants in the Russell Sage Foundation administrative data conference for useful comments and discussions. Direct correspondence to: Andrew M. Penner at penner@uci.edu, University of California, Irvine, 4181 Social Science Plaza, Irvine, CA 92697.

evaluation of policies and interventions, they also enable researchers to address novel questions. Administrative data in many countries are routinely collected and used to improve policy and contribute to science to a much greater degree than in the United States. In this double issue, we hope to improve collaborations between American communities and academic scholars.

The value of "big data" is becoming increasingly apparent. In the current data revolution, firms' profits are determined by their ability to capture and analyze data. The capacity to collect and analyze data determines not only the winners and losers in the economy, but also which societies can best educate their citizens, train their workforce, keep their population healthy, and promote the well-being and flourishing of their citizens. Schools, hospitals, and many other organizations now routinely collect data that allow them to serve their students, patients, and other stakeholders more effectively. Understanding the lessons contained in these data, and how best to extract them, plays an important role in helping practitioners, administrators, and policymakers understand the challenges that the American public faces and the effects of current practices and policies.

Recent federal initiatives such as the Murray-Ryan Evidence-Based Policymaking Commission Act of 2016 and the resulting Foundations for Evidence-Based Policymaking Act of 2018 suggest that administrative data will become increasingly central in U.S. social science and policy. However, efforts to leverage administrative data in the social sciences are uneven. In domains such as public health and demography, administrative data are used routinely, while they are less frequently used in mental health, social behavior, and other areas. But even where administrative data are commonly used, current capacity constraints hinder their utility. Data that could be used to inform decisions are often unavailable to those making the decisions, leaving them with incomplete information.

One clear example is in education. Despite their focus on preparing students who are "college- and career-ready," schools have historically struggled to obtain data on the practices that will prepare their students to be successful because widespread links between students in K–12 educational systems and higher education outcomes have become available only recently, and links between K–12 data systems and the labor market remain relatively rare. These data linkages are important to understand the efficacy of school-based vocational programs, dropout recovery interventions, college readiness programs, and advancement placement course policies. But schools, like other organizations, typically lack the capacity and expertise to build this infrastructure and analyze the resulting data.

This double issue opens with a section that includes an article by Sean Reardon showing the potential value of examining nationwide data on student achievement (2019). He amasses an unprecedented administrative data file and then breaks it down in many ways that are useful to policy decision making as well as scientific understanding of child development and the impact of education. This file, and others like it, can be used to address numerous important issues that were previously unavailable to rigorous empirical scrutiny. For example, Megan Austin, Joseph Waddington, and Mark Berends show how administrative data can inform the important policy question of the impact of school vouchers on student achievement (2019).

Administrative data research seeks to bring the data-driven approach increasingly used in fields such as micro-targeting (where commercial firms individualize advertising according to one's buying patterns) and precision medicine to help answer important societal questions, providing schools, clinics, and others serving the public good with the information they need to work more effectively. To the degree that public-sector institutions are unable to build the infrastructure needed to understand the effectiveness of their programs, it seems likely that this lack of data infrastructure will result in inefficiencies (particularly relative to private sector institutions with fewer restrictions regarding data).

In this article and the double issue that follows, we seek to highlight the promise of administrative data to answer pressing questions for both science and policy. Our goal is not to be comprehensive, but to provide an indication of the potential that is currently barely being

tapped, offer suggestions for advancing this infrastructure, and highlight some of the challenges currently facing the field. Likewise, our goal is to spark future work by highlighting the promise of these data by showcasing a few exciting administrative data initiatives and research projects.

THE CASE FOR ADMINISTRATIVE DATA INFRASTRUCTURE

Administrative data have many benefits for research that seeks to advance science and inform policy. These data typically originate not for research purposes, but instead as client-level service records or from administrators who are accountable for documentation of implementation of services. In many cases, the information available in administrative datasets and the analyses of these data will be useful to policymakers precisely because these are the variables for which policymakers are accountable. Furthermore, these datasets often attempt to reach the entire population of relevant participants and suffer less from selection bias than many original research datasets. Thus, not only does administrative data infrastructure provide important opportunities to understand the social world better, but it does so in contexts where stakeholders are ready to act based on the results of research. As a result, the analysis of administrative data has the potential to bridge basic and applied research.

Appropriate data infrastructure allows broad and equitable access, lessens the burden on the organizations providing the data, standardizes expectations around availability and privacy, and allows for best practices in data security. A robust administrative data infrastructure also lowers the marginal cost of conducting additional research, allowing researchers to address important issues using existing data that do not incur the effort or pecuniary costs of new data collection. Although the startup cost of establishing this infrastructure is not cheap, once established, it raises the quality of the research it permits and allows for more research by changing the cost curve for high-quality additional research. Furthermore, accessible administrative data files enable replication of data analyses and findings across independent research teams using the same and different files, which improves the robustness of the field. In a world where data analytics and knowledge are increasingly central to society, strong administrative data infrastructure is therefore not only important for knowledge creation, but also efficient.

Administrative data systems operate at scale. This feature allows researchers to consider questions such as whether all individuals benefit equally from a particular policy. When combined with research designs that allow for causal inference, these population-level data files afford answers about whether any individuals or groups are negatively affected and other questions of treatment-effect heterogeneity that are often prohibitively expensive to consider without the scale that administrative data provide. Having data on the entire universe of relevant individuals also allows for comparisons of small groups that are otherwise extremely difficult to isolate, for example, comparing employees with people doing the same job in the same establishment (Petersen and Morgan 1995). Operating at scale means that one can consider treatment not just at the individual level, but also at the community level, examining how spillover effects, feedback loops, and other dynamics might cause policies to work differently when implemented broadly. Finally, of course, having the population of participants available for analysis minimizes (but certainly does not eliminate) problems that survey researchers confront, such as biased attrition and missing data.

We next highlight several useful features of administrative data that are instrumental to advancing science and policy.

A Long-Term Perspective

Recent research has highlighted the utility of administrative data for understanding the long-term outcomes associated with a variety of interventions. Whereas much policy is necessarily driven by research and evaluations examining relatively short-term outcomes, administrative data can provide a longer-term perspective on the effects of past policies and thus a more accurate account of a policy's costs and benefits. Beyond following individuals who were exposed to interventions over time, administrative data can provide an important tool

for understanding multigenerational cycles of advantage and disadvantage, allowing researchers to trace the descendants of individuals from different backgrounds as well as the multigenerational effects of antipoverty policies and interventions. Administrative data also provide opportunities to connect researcher-collected data (for example, observations of student behavior) with future administrative records to help understand the long-term implications of researchers' observations (such as longer-term cost outcomes related to early school behavior problems).

Administrative data research has pushed the boundaries of what we know about longer-term temporal processes, highlighting the important implications of understanding longer-term policy effects, as well as the intergenerational transmission of advantage and disadvantage more broadly. Standard intervention research typically ends data collection at the end of the study and provides an evaluation at that point. In some large-scale studies, follow-ups track outcomes for multiple years post-intervention (see, for example, Ludwig et al. 2013). Insofar as participants in interventions continue to be captured in administrative records after the end of the study, administrative records provide opportunities for an efficient approach to understanding the benefits of interventions decades later. Given the importance of understanding the long-term impacts of policies aimed at ameliorating poverty (Bailey et al. 2017), and the goal of changing not just individuals' current circumstances but also their longer-term trajectories, one of the most exciting features of administrative data is their ability to turn any study into a de facto longitudinal study. Recent work in this vein, for example, has highlighted the adult income effects of having an experienced or effective kindergarten teacher and higher-achieving classmates (Chetty et al. 2011; Chetty, Friedman, and Rockoff 2014a, 2014b), the increases in adult civic participation arising from a psychosocial intervention in elementary school (Holbein 2017), and the benefits of cash transfers to poor families on their children's educational outcomes, mortality, and income (Aizer, Eli et al. 2016).

Administrative data research has also extended a long tradition of survey-based research highlighting the intergenerational transmission of advantage and disadvantage (Blau and Duncan 1967; Ganzeboom, Treiman, and Ultee 1991). Survey-based research has examined parent-child transmission processes (for example, Campbell et al. 2014; Poulton et al. 2002), and in some cases, grandparents (Wightman and Danziger 2014). Administrative data allow researchers to push back on the time horizon of these studies. Using Swedish registry data, for example, Martin Hällsten examines the effects of not only grandparents, but also great-grandparents (2014). Likewise, work by Hyunjoon Park uses Korean historical registry information to examine the outcomes of ethnically Korean slaves, showing substantial effects on their descendants many generations after the abolition of slavery (2014). Multigenerational data are somewhat more rare in the United States, the Utah Population Database being a notable exception (see, for example, Temby and Smith 2014). A more detailed discussion of the opportunities provided by administrative data for understanding multigenerational processes is available from Xi Song and Cameron Campbell (2017). These data provide not only exciting new opportunities to advance the science of the multigenerational transmission of advantage, but also the long-term costs of stigmatized identities, poverty, and disadvantage. As data infrastructure matures, we expect to see more studies examining how social policies aimed at one generation affect their children and grandchildren (for example, Meghir, Palme, and Schnabel 2012).

Key Sites and Populations in the Production of Inequality

Inequality in opportunities and outcomes across groups and persons is one of the most vexing problems facing contemporary society. These inequalities are often produced in spaces that are difficult to examine using surveys or experiments. In the hiring process, for example, correspondence studies can examine who receives responses from potential employers but cannot help us understand how applicant pools are created, which interviewees receive offers, or whether pay differences exist among those who receive offers. Research on inequality in other institutional spheres such as health care,

housing, and higher education faces similar challenges. Inequality in institutions is often shaped by processes that determine who is included and excluded, and how individuals are ranked within an organization. Administrative data on hiring pipelines, performance ratings, promotion decisions, and decisions about termination can provide valuable insights into the decisions made by gatekeepers at key sites regarding entry and exit from organizations, as well as important processes governing intra-organizational inequality regimes. One useful feature of workplace administrative records is that they allow researchers to compare individuals with others who have the same occupation working in the same establishment. Research making such comparisons suggests that much of the wage inequality observed across gender and race is created by sorting processes, as individuals doing the same work for the same employer receive largely similar pay (Petersen and Morgan 1995; Tomaskovic-Devey 1993).

Building on this insight, a long tradition of case studies uses administrative records from company human resource departments to understand inequality in these sorting processes. Trond Petersen and Ishak Saporta argue that in the current institutional context the opportunity for firms to discriminate is greatest at the point of hire (as opposed to discrimination in promotion or termination practices) and show that hiring is where the largest differences between men and women are observed (2004). Roberto Fernandez and his collaborators use human resource data in a series of papers that provide important insights into race and gender inequality in the hiring process; they show, for example, the importance of referrals (Fernandez and Greenberg 2013) and the importance of supply-side adjustments to perceived demand-side constraints (Fernandez and Friedrich 2011; Fernandez and Campero 2017). Recent work in this vein highlights the perhaps surprising egalitarian influence of an executive search firm (Fernandez-Mateo and Fernandez 2016), and in the current double issue Fernandez and Brian Rubineau use their extraordinary hiring data to provide novel analysis of network recruitment efforts and their impact on the gender-based glass ceiling in the biopharma industry (2019).

Beyond the labor market, administrative data from other sources also provide important insights into key sites for generating inequality. Research on NIH funding decisions, for example, uses detailed records to document the existence of gender inequality in the NIH review process (Li 2012). Research examining the criminal justice system uses administrative records from juvenile courts to estimate the effects of juvenile incarceration on later criminal justice and school outcomes (Aizer and Doyle 2015). One of the supposed keys to combating inequality and reaching life success is, of course, education. Although we know that dropping out of high school breeds failure and college graduation brings success, much less clear is the value of post–high school associate degrees, vocational diplomas, certificates, and partial college. In this double issue, ChangHwan Kim and Christopher Tamborini merge school administrative data files with earnings files to examine the long-term earnings that accrue from these post–high school accomplishments (2019).

Janelle Downing and Tim Bruckner use housing foreclosure administrative records and birth records to highlight yet another source of inequality (2019). They show that housing foreclosures (and presumably the stress they cause) contribute to premature births and increase inequality in birth outcomes across race and ethnic groups.

Administrative data also allow us to understand small, often difficult to access, populations that are theoretically important. For example, research using large administrative datasets has shown that millionaire tax flight does not occur at levels that are socially meaningful (Young et al. 2016), and that top earners are increasing isolated from the rest of the population (Godechot 2013). Insofar as many large administrative datasets include information on the whole population, these data allow researchers to examine relatively small and theoretically important groups (for example, those that are hard to capture in a probability sample without an explicit oversample) without compromising representativeness (Liebler, Bhaskar, and Porter 2016). From a local policy perspective, the ability to identify small groups of people is helpful because it allows policymak-

ers to ensure that the policies they implement are having their intended effects for all stakeholders, and to determine where adjustments to existing policies are needed (Howard et al. 2019).

Understanding Individuals in Their Social Contexts

The importance of context is a truism in social science research. From network influences to cultural factors to questions about positional goods, contextual considerations play a profound role in shaping an individual's outcomes. Despite this, interventions and policies have historically operated from a baseline that presupposes constant universal effects that operate at scale if implemented with fidelity (Dodge 2011). The density of information in administrative data is useful in providing opportunities to examine important sources of heterogeneity (particularly contextual sources, but also individual-level factors), as well as providing opportunities to investigate policies and interventions at scale. Durham Connects, for example, takes advantage of administrative data by assigning all newborns in Durham born on even days to receive a nurse home visit (Dodge et al. 2014). This design allows for children to be followed in administrative records throughout their lives in an ethical manner without the requirement of individual consent (because data can be de-identified before being analyzed but retain the essential characteristic of assignment to intervention) without necessitating additional data collection.

Further, thinking about changes to the social system more broadly (for example, moving bus stops to help mothers travel to local clinics) shifts research and policy discussions away from a methodological individualism that focuses on the effects of treatments on individuals, and toward considering how programs and policies affect social systems more broadly (see, for example, Denice and Gross 2016). These systems-level approaches are important for both science and policy, as they address an important shortcoming of much social science research. Research often asks what would happen if everything was held constant and only one consideration was changed. This can be instructive, but it fails to take into account the myriad of ways that people and their social worlds are interconnected. Analyses of how community-level policies and interventions shape not only individuals' outcomes, but also society more generally, allow researchers to capture the complicated feedback loops and spillover effects that occur when interventions and policies operate at scale, providing insight into how policies might change society more broadly (Dodge 2009; Penner et al. 2015). Although in theory such analyses are possible without using administrative data, in practice the existence of administrative records greatly facilitates them.

Research using administrative data can also provide a more complete account of certain aspects of context, including the government services context. For example, Robert Goerge and Emily Wiegand examine the overlap in families' access of government services across multiple agencies to show how some families are accessing many agencies whereas others seem to be underaccessing resources that might benefit them (2019). Goerge and Wiegand show that these differences vary across geographical locations within Illinois, suggesting that local practice might contribute to, and mitigate, any biases. Lanikque Howard and colleagues examine the relation between parents' payment of child support and children's involvement in the child welfare system (2019). Agustina Laurito and colleagues combine multiple administrative datasets to show how school climate and neighborhood crime levels affect student achievement (2019).

Administrative data can also afford studies that are simply not plausible through original data collection. For example, it is difficult to imagine survey data tracking all of the classmates that a student had or all of the co-workers over an employee's career, but educational administrative data and linked employer-employee datasets often include this information (Abowd, Haltiwanger, and Lane 2004). Administrative records can thus provide information not only about an individual research subject but also about the environment surrounding them. Christopher Candelaria exploits such data in education, where he disentangles the long-term effects of a third-grade teacher, the medium-term effects of middle-

school teachers, and the short-term effects of an eighth-grade teacher (2015).

Linking administrative data files across levels also provides innovative opportunities to understand individual behavior in broader context. Elizabeth Ananat and her colleagues, for example, link county-year-level administrative data about community-level job loss with individual student educational administrative data files in order to discover the impact of local economic downturns on student academic progress (2017). Further, as noted previously, administrative data allow for individuals to be placed in a multigenerational familial context. In providing dense coverage of populations, these data allow researchers to examine whether policies had spillover effects (either positive or negative) on those around the targeted populations, and to examine questions around how context moderates the effectiveness of treatment. Heterogeneity in treatment effectiveness is important not only for contributing to scientific understanding regarding the mechanisms through which interventions work, but also because it has important implications for generalizability and scalability (Domina et al. 2016).

An Iterative Policy Design, Implementation, and Evaluation Cycle

A common goal of scientific research is to improve societal outcomes. Social scientists often seek to do this by evaluating and informing existing policies, and it is not uncommon to hear researchers bemoan the lack of policy responsiveness to research. Although obtaining access to administrative data can be time consuming in some contexts, a potential advantage of administrative data analysis in many contexts is that the institution generating the data is also making and implementing the policies being evaluated, so that there is an audience that is positioned to make decisions about practice and policy based on researchers' findings (Howard et al. 2019). This is particularly true in researcher-practitioner partnerships, where researchers partner with organizations to help them use their administrative data in better ways to answer questions of interest to decision makers and stakeholders.

The research-policy link in social science is often conceptualized as one in which research informs or evaluates policies, but in research-practitioner partnerships, policy implementation and research can have a bidirectional synergistic relationship. These partnerships provide data researchers could not otherwise access. This unusual opportunity holds not only for companies' human resource data, but also for educational data, where laws protecting the privacy of student data allow data to be shared with organizations conducting research on behalf of educational agencies to help improve instruction. Further, researchers not only have the opportunity to study important policies in real-world settings but can often inform the implementation process. This relationship not only allows for policies that draw on researchers' expertise, but can also lead to opportunities for better research because researchers can help implement policies in ways that facilitate high-quality evaluations (such as introducing lottery-based assignments for oversubscribed programs or thresholds in assignment scores that enable regression discontinuity-based designs). In contexts that facilitate the timely incorporation of feedback, data collected can be used to inform Bayesian adaptive designs to help improve interventions in real time (Finucane, Martinez, and Cody 2017).

By bringing researchers into the policymaking process, rigorous research becomes part of the iterative process of policy implementation and adjustment and the policy adjustments made in implementation are better captured by research (Howard et al. 2019). This approach can provide both better policy and better research and serves as a model for how to accumulate and incorporate knowledge beyond simply conducting a series of single-policy evaluations. Although shortening the feedback loop among implementation, research, and redesign is likely to be positive overall, one challenge of this linkage is that long-term outcomes by definition cannot be observed quickly, and many policies—particularly those aimed at implementing organizational changes—have effects that take time to emerge or vary across different stages of the implementation cycle (Mills and Wolf 2017; Sun, Penner, and Loeb 2017).

EXPANDING ADMINISTRATIVE DATA INSIGHTS

Existing administrative data have allowed researchers to address a variety of important questions, and in many contexts, administrative data provide the best opportunities to answer important policy questions (Austin, Waddington, and Berends 2019). Current infrastructure, however, constrains research and limits the ability to answer questions that are important for science and policy. In the section that follows, we highlight important frontiers for administrative data research, highlighting noteworthy exemplars of work in these areas. We frame the points as strengths, but they could also be conceptualized as ways to address the potential weaknesses of administrative data.

Combining Datasets Across Sources

Although current administrative data research focuses on contemporary data sources, there is a long tradition of using administrative records in archival and historical research (see, for example, Kessler-Harris 1982; Wilde 2004). Research using administrative data has much in common with history and archeology, insofar as it observes the tracks that individuals leave as they move through society and draws lessons from these glimpses into their lives. A key difference is that when records outlast people, opportunities for supplementing and triangulation through interviews, surveys, or ethnography decline, leaving scientists to reconstruct meaning from the traces people have left behind. Although administrative data researchers using contemporary data draw conclusions from the traces left behind in current records in a similar manner, research using contemporary records has the potential to incorporate information directly from individuals through surveys and observations to supplement the data in administrative records.

Given their origin in a particular institutional context, administrative records are typically fragmented, and these data are often not linked to other data that would be useful for research and policy. Hospitals, for example, collect detailed information about patients' health, schools regularly collect information about student development, and employers often keep records not only about the performance of employees, but also about applicants who were ultimately not offered positions. Although various combinations of these data can provide important insights, they are typically compartmentalized. Likewise, given their origin, administrative records often lack certain kinds of information that are less likely to be collected in these records. For example, information about attitudes, affinities, and motives are not often collected in administrative records. Combining administrative data with records from other sources—either by linking administrative records across sources or by making administrative records available to be linked to data collected via other means—is thus central to building administrative data infrastructure.

Linking Administrative Data Records Across Domains

By virtue of how they come into existence, administrative data are typically focused on one facet of an individual's life, and data and insights are often siloed. Given that the potential for insight grows exponentially as data are integrated, combining administrative data across domains is of vital importance, and enables researchers to trace connections between settings like schools, criminal justice institutions, health organizations, and employers, and see how inequalities compound across these domains. Our introduction describing Durham Connects highlights the power of these insights for understanding the needs of families across diverse domains, and others likewise underscore the utility of linking administrative records across domains to understand the challenges facing families in poverty (Goerge and Wiegand 2019). Research linking data across domains documents how inequality in one domain shapes outcomes in others, highlighting, for example, the health consequences of foreclosure (Downing and Bruckner 2019) and how air pollution shapes mortality risk (Di et al. 2017). Other research in this vein allows us to understand the broad effects of policies, showing how lead abatement efforts lower children's blood lead levels and improve student achievement (Aizer, Currie et al. 2016), and how Superfund site cleanup improves children's later educational outcomes (Persico, Figlio, and Roth 2016).

Beyond helping us understand disadvantage better at any given point in time, linking data across domains can also open opportunities to follow individuals as they move through different institutional settings. Administrative records from birth, education, criminal justice, labor market, and mortality often capture different points in an individual's life; combining data across these stages allows us to understand how inequalities unfold over the arc of an individual's life. For example, research linking educational records with IRS records highlights the long-term income benefits associated with high-quality teachers (Chetty, Friedman, and Rockoff 2014a, 2014b) as well as the link between college major choices and later life income (Kim and Tamborini 2019). Similarly, research on the school-to-prison pipeline in Texas links education and justice records to trace the juvenile justice involvement of students suspended from school (Fabelo et al. 2011).

Much of the attention in administrative data infrastructure has focused on large-scale population-level data. However, as noted earlier, one of the potential advantages of using administrative data is that they provide information about social processes that are otherwise very difficult to study (such as the hiring pipeline). Research using administrative records to study otherwise inaccessible processes typically does not focus on linking across domains to the same degree as population-level administrative data research, presumably because of the underlying logic of these projects, which focuses primarily on isolating a hard-to-identify set of processes. Further, the unique relationship between data owners (often private companies) and researchers, and the difficulty in linking with public administrative sources (for example, the Census Bureau must avoid doing research that would favor one company over another) make linkage particularly challenging. That said, linking administrative records from these contexts with other administrative records could provide important insights and would appear to be an important frontier for administrative records research. For example, such data could help us understand how graduates from job search and other training programs fare at different stages in the hiring process. To date, we are aware of only one project that has linked human resource records with other individual-level data: linking human resource data on the hiring pipeline at a school district with data at the Census Bureau (Brummet and Penner 2017). Among other things, such data linkages provide opportunities for understanding the labor market implications of unsuccessful applications. We suspect that as the importance of evidence-based practices grows—both generally and in the context of securing foundation funding—opportunities for linking data from local organizations with important domain-specific information will continue to increase.

Combining Survey Data with Existing Administrative Records

The narrow specificity of some administrative data files often limits the range of scientific research questions that analyses drawing solely on that file can examine. Although this has the benefit of focusing researchers' attention on the measures salient to practitioners and policymakers, researchers often supplement administrative records with other information. For example, by linking administrative records with surveys measuring constructs of interest, researchers have examined teacher effects on motivation (Ruzek et al. 2015), shown how school climate can mitigate the academic effects of neighborhood violence (Laurito et al. 2019), and demonstrated how a manager's human resource practices moderate the relation between manager gender and gender wage inequality among workers (Abendroth et al. 2017). Future research in this vein linking implicit bias measures with hiring managers' real-world decisions from human resource data would also help us greatly expand our understanding of how organizational context and policies might moderate the effects of these biases. Likewise, researchers who have information on particular individuals often supplement that information with administrative records. A number of studies, for example, have used administrative data to examine the long-term outcomes associated with interventions, linking researchers' information about who received the treatment with administrative records (Chetty et al. 2011; Holbein 2017).

Elsewhere in this double issue, David Grusky,

Michael Hout, Timothy Smeeding, and Matthew Snipp highlight an additional benefit of combining survey and administrative data, noting that a common data infrastructure would allow surveys to be overlaid on top of administrative data and alleviate respondent burden (2019). This would enhance what is possible using either the survey or the administrative data independently.

Qualitative Research with Administrative Data

Although much of the research using administrative data uses quantitative information, administrative records also contain vast amounts of qualitative information. Archival research using administrative records provides a strong indication of the considerable value of qualitative work using administrative records. Although qualitative social science research using contemporary administrative records is also just beginning to realize its potential, several examples evince the promise of such approaches. Recent qualitative research in medicine, for example, highlights gender differences in the feedback that medical school residents receive (Mueller et al. 2017), and research on online dating profiles underscores how racial boundaries are reinforced not just by racial homogamy, but also by those looking to date across racial lines (Rafalow, Feliciano, and Robnett 2017).

In many administrative contexts, given the scale of textual data, advances in machine coding offer a promising approach to turning rich qualitative data into quantitative data. In this double issue, Emily Penner and her colleagues provide one example of this approach, showcasing how essays submitted as part of teacher applications are correlated with a variety of policy-relevant considerations (2019). The promise of such approaches in researcher-practitioner partnerships is difficult to overstate, because when these organizations begin to leverage their data in the ways that large tech firms do, there would appear to be substantial benefits for both policy and science. With text mining becoming increasingly sophisticated and common, and the growth of software to aid in the transcription, storage, coding and sorting processes in qualitative research, the distinction between quantitative and qualitative research is one that could quickly fade in administrative data research.

TECHNICAL, LEGAL, ETHICAL, AND PERCEPTUAL CHALLENGES

In this last section, we highlight a few current challenges specific to working with administrative data. Many are extensions of challenges that exist in social science more broadly around balancing the privacy of research participants with making data widely accessible to lower the barriers to conducting research. In this respect, we see parallels between current efforts to democratize access to administrative records (see, for example, Grusky et al. 2019) and the advent of the General Social Survey, which made nationally representative survey data widely available to the scientific community. Prior to the General Social Survey, social scientists collected their own surveys and typically did not provide data access to outside researchers, so that access to survey data was typically restricted to prominent scholars and their students. More recently, calls for greater transparency and reproducibility have underscored the value of open science in experimental fields (see, for example, Ioannidis 2005; Open Science Collaboration 2015). Against this broader backdrop, thinking about what open science looks like in the context of administrative data research is critical.

Aggregation of individual data into group scores provides a partial solution to the challenge of privacy in many contexts. This double issue includes two studies that use aggregated data files. Brittany Murray and colleagues report the positive relation between strong parent-teacher associations and growth in student achievement (2019). Portia Miller, Elizabeth Votruba-Drzal, and Rebekah Coley find that community-level resources explain variation in student achievement (2019).

However, aggregated data cannot answer all questions, and in many cases answering research and policy questions requires individual-level data. To facilitate sharing of individual-level data, it is likely important to establish incentives for administrative data linking efforts so that more scholars contribute to this public good. One challenge here is that, due to

the sensitivity of many administrative datasets, access is highly regulated, and it becomes prohibitively difficult and time consuming to navigate the multiple processes required to obtain access to data across different contexts. There are two broad models for addressing these challenges in international comparative research: the Comparative Organizational Inequality Network, which brings together researchers with access to the relevant data in different contexts around a set of common analyses that each researcher conducts on data from their home country; and the Luxembourg Income Study, which creates a largely harmonized set of data from across countries (currently non-administrative survey data) and allows researchers to submit code to run on datasets from different countries without accessing the original data. Given their different costs and benefits, we suspect that both models have important roles to play in comparative research.

More broadly, the challenges in working with administrative data can be broken down into technical, legal, ethical, and perceptual challenges. We review each in turn.

Technical Challenges

Important technical challenges remain to constructing administrative data infrastructure. For example, address-based matches are difficult to implement in contexts that lack a well-defined address system (see, for example, Wynn, Reyes, and Caldwell 2011). Likewise, for computationally intensive analyses (for example, some social network analyses) it is currently not practical to conduct analyses that make use of the density of information available at the population level. These and other questions notwithstanding, in our estimation the largest challenges to administrative data are not technical per se but instead technical constraints imposed in response to legal or perceptual considerations.

For example, it is not clear that there is a strong rationale for why researchers need to be in Texas to analyze data from the state of Texas (except that it may be easier to arrest a misuser within state), or that data from Georgia should be allowed to be used on projects only with a collaborator from a university located in Georgia. Nevertheless, such arrangements remain relatively common. Although they are not insurmountable, they do create nontrivial barriers to access and hinder the democratization that researchers generally support in science. That said, given the level of trust required for companies to allow researchers to analyze key intellectual property (Fernandez-Mateo and Fernandez 2016) or for countries to allow outside researchers access to tax data (King et al. 2017), some restrictions to access beyond those governing survey data are warranted. These barriers highlight the point that the most important challenge to successful administrative data scholarship is not the technical nature of data storage or security, but rather, the human and institutional relationships that must be developed and maintained. The relational nature of data access in many cases—such as in long-term researcher-practitioner partnerships—does result in important constraints that are in tension with norms around data-sharing and open science.

One important challenge surrounding administrative data is the lack of consistency regarding which data are collected and how they are collected. Although national surveys typically use standardized measures and best practices for assessing various constructs, information contained in administrative data can be highly variable in terms of coverage and quality. One advantage of working in close partnership with the organizations generating administrative data is that they typically have a deep understanding of how the data are generated and areas where information may be inaccurate or have limited coverage, and can often adjust practices to generate data that are of mutual interest. Working closely with partners on the ground can also help avoid misattributing causal relations. As with most survey data, administrative data sources require that findings be disseminated only as aggregate statistics in order to protect privacy. As far as we are aware, very few cases of researchers infringing on the privacy of individuals using administrative records have been documented. At this point, then, the technical challenges involved in building data infrastructure are largely surmountable, and the larger remaining question is whether political will is strong enough to move forward.

Legal Challenges

Currently, legal constraints affecting administrative data infrastructure focus on balancing the privacy of individuals whose data are contained in the administrative records with the ability of institutions to find answers to their pressing policy questions, which in many cases will enable them to serve better those who are represented in their data. Allowing access to outside researchers working on behalf of the organization can greatly enhance the research capacity of institutions that generate administrative data and provide expertise in areas that might be otherwise difficult to obtain. In this context, analyses of administrative data should address questions of the data owner, presumably in service of either those represented in the data or the broader public. By contrast, scientists argue that science benefits from widespread, democratic access, and that this access can yield new insights that might be broadly beneficial to society, the institutions generating the data, and their stakeholders, even if these benefits might not have been anticipated. Although making administrative data more widely available is likely generally beneficial, it is currently difficult to know how to assess and weigh the benefits from broad access.

Many forms of administrative data are legally protected in ways that limit access. Under the Family Educational Rights and Privacy Act, identifiable educational data in the United States can only be shared with researchers in a limited number of contexts, including cases where the studies will help the schools improve instruction. Similar challenges apply to health information and Health Insurance Portability and Accountability Act regulations. The lack of a well-established administrative data infrastructure means that lawmakers often do not consider the impact of legislation on administrative records. For example, out of concerns regarding administrative records being used for enforcement purposes, California lawmakers sought to enact laws prohibiting data-sharing and initially did not recognize the limitations this would create for researchers and administrative data infrastructure. Presumably a more robust and salient administrative data infrastructure will help in avoiding such issues in the future.

In many ways, legal constraints are a question of political will. On this point, the bipartisan support for administrative data represented by the Murray-Ryan commission and the Foundations for Evidence-Based Policymaking Act is encouraging. One might imagine, for example, that evidence-based policies around education and workforce training programs might benefit from administrative records from schools, even if the resulting study might not help each school improve instruction. Although individual lawmakers may differ on policy priorities, it is encouraging that they agree on the need for better data and analysis to inform them.

Ethical Challenges

Beyond strictly legal questions, there are ethical questions as well. Typically, potential research participants have the choice to opt out of a study. But this is not possible in most research that uses administrative records. Although consideration of informed consent is routine when it comes to whether a participant's data are used in traditional research designs, administrative records research is often considered to be nonhuman subjects research. To be clear, questions around individuals' rights vis-à-vis their data are a feature of administrative data more generally and not particular to research. This is apparent when one considers medical records. In approximately half of the states in the United States, physicians or hospitals own patients' records, and only in New Hampshire do these data belong to patients (in the remaining states data ownership is not clearly defined). It seems unlikely that patients would take issue with research analyzing these records for patterns that might help save their life. Likewise, it seems probable that most people would not object to their records being analyzed for research that might help save the lives of others. Nevertheless, because this research is often not considered to involve human subjects, and these data (outside New Hampshire) do not belong to the individual, it is unclear what rights patients should have to restrict the use of their data in administrative records research.

Historically, the argument has been that the primary potential harm in this research is that

of disclosure, or harm to the individual due to a breach of privacy. Some legal scholars suggest that this individualistic perspective may be problematic. In a high-profile example, the Havasupai sued Arizona State University for using existing blood samples in ways not covered by agreements. In discussing this case, Katherine Drabiak-Syed notes that our current legal system is ill equipped to consider issues beyond an individualistic framework, so that harm to a collective group may not be recognized (2010). These questions are perhaps especially salient in the context of Native Americans, where issues over the right to opt out are laden with colonial legacies of ignoring indigenous perspectives and also raise questions of tribal sovereignty. These concerns are likely heightened where blood (or other physical samples) are involved, where research focuses on historically marginalized populations, or when researchers are partnering with data-collecting organizations. The concerns are perhaps somewhat attenuated when looking at historical data (for example, the Dutch Hunger Winter), but the larger point remains relevant for administrative records research.

More broadly, the issue could be conceptualized as whether individuals should have the right to ensure that their data are not used in systems against their wishes. One might imagine, for example, critics of structural racism not wanting their data to be used by companies that might perpetuate racial differences in homeownership through credit scores. But it is difficult not to be complicit when almost everyone is part of the administrative data ecosystem that creates and reproduces these inequalities. This is a feature of our societal data infrastructure and is not specific to research using this infrastructure. Nonetheless, administrative data researchers should be cognizant of these issues, particularly in contexts like researcher-practitioner partnerships where they might influence the kinds of data collected, and where the research being conducted might be used to justify or rationalize practices that may otherwise be seen as problematic.

Perceptual Challenges

Perceptual challenges relevant to administrative data research can be divided into those within the academy and those in the public domain. Within the academy, in many social science disciplines there is a bias against work that is viewed as overly applied. The term *evaluation research* for example, is sometimes used pejoratively in contrast with *pure* science, implying that scientific work is somehow contaminated by being useful to society. We argue that whatever the origins of this bias, it is a distinction that has outlived its usefulness, and that supporting human flourishing—both through better understanding the social world in the broadest and most abstract sense, as well as through understanding the implications of the concrete choices that we as a society make—ought to be one of the aims of science. The degree to which these biases are held in any given scientific field varies, suggesting that social science disciplines can learn much from those more engaged in policy. These disciplinary biases are perhaps a space that academics are well positioned to change. Although these norms may be deeply entrenched, they are nonetheless created and maintained by academics, suggesting that we as a community can change them by changing our hiring criteria, tenure and promotion letters, award nominations, and graduate training. We suspect that these perceptual challenges within the academy are decreasing, in part because administrative data allow researchers to address questions that are not only important for real-world applications, but also make fundamental contributions to discipline-specific and transdisciplinary research goals. We believe the proliferation of administrative data research suggests that over the long term, perceptual issues within the academy are likely to become less pronounced.

Beyond the academy, public skepticism about the limits of confidentiality and data protection threaten public support for the use of administrative data. Recent hacking events and misuse of large private data files at Facebook and Cambridge Analytica have shaken public faith in keepers of supposedly private data. The threat goes beyond misuse to include possible political obstruction by groups such as ALEC (American Legislative Exchange Council), which has taken the position that all governmental action should be minimal. The possibilities of misuse by insiders, hacking by out-

siders, and opposition by politics will always be present, but we believe the marginal extra risk imposed by bringing researchers into this circle is very low. Researchers are required to be trained and credentialed in the use of sensitive data files, and universities tend to implement cutting-edge technologies in data security. Because of these threats and the public's vigilance, however, researchers would be wise to understand the treasure that they behold and to be extremely careful in their use of administrative data files.

At the same time that the public is skeptical, bipartisan support is also strong for administrative data science to improve our capacity to maximize the potential of our human resources. Data-sharing can be difficult in contexts marked by suspicion and mistrust, and larger conversations around privacy remain important. Legal protections governing administrative data use thus play central perceptual and scientific roles, as well as being important for ethical reasons (Anderson and Seltzer 2007). We believe that it is incumbent on scientists to help make the case for administrative data research by ensuring that the public benefits from the use of their data. Although in some cases this might mean working closely with policymakers and practitioners generating and using the data, press coverage of novel findings using engaging data visualizations that reach the public more broadly also play an important role in highlighting the utility of these data to the broader public. Wide dissemination of research findings not only helps inform public discourse around important social questions, but also plays an educational role by engaging people's curiosity and helping them understand how the social world works.

CONCLUSION

As a society, we have the data and expertise to address questions that are vital to our communal life, but we currently do not have the infrastructure to bring data from disparate sources together and provide access to researchers with high-impact projects. U.S. administrative data infrastructure has lagged behind that of its peers, leading to policies that are not as well tuned as they might be, and in many cases leading American social scientists to work with better data from other countries. Important policy-relevant scientific questions go unanswered, and scholars and policymakers are left to infer how things might work in the United States based on evidence from elsewhere. The lack of data infrastructure has human costs for our students, patients, and their families; has pecuniary costs for taxpayers; and puts American science at a disadvantage. Recent efforts to create administrative data infrastructure have great promise to rectify the situation, making it an exciting time to be an administrative data researcher.

One final word of caution is perhaps in order: in America, the logic of competition drives many of our collective efforts. When building infrastructure, however, coordination is important. To use a metaphor from physical infrastructure, having five sets of highway systems that do not connect with each other is considerably less useful than having a single, well-planned system. We have an opportunity to create world-class data infrastructure that will enable policymakers to make better policies, scientists to understand society better, teachers to instruct students better, and physicians to treat patients better. In moving forward, coordinating efforts to ensure that we build the best data infrastructure possible, and that our data can benefit the public as much as possible, is paramount.

REFERENCES

Abendroth, Anja-Kristin, Silvia Melzer, Alexandra Kalev, and Donald Tomaskovic-Devey. 2017. "Women at Work: Women's Access to Power and the Gender Earnings Gap." *ILR Review* 70(1): 190–222.

Abowd, John M., John Haltiwanger, and Julia Lane. 2004. "Integrated Longitudinal Employer-Employee Data for the United States." *American Economic Review* 94(2): 224–29.

Aizer, Anna, Janet Currie, Peter Simon, and Patrick Vivier. 2016. Do Low Levels of Blood Lead Reduce Children's Future Test Scores? No. w22558. National Bureau of Economic Research.

Aizer, Anna, and Joseph J. Doyle Jr. 2015. "Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly Assigned Judges." *Quarterly Journal of Economics* 130(2): 759–803.

Aizer, Anna, Shari Eli, Joseph Ferrie, and Adriana Lleras-Muney. 2016. "The Long-Run Impact of Cash Transfers to Poor Families." *American Economic Review* 106(4): 935–71.

Ananat, Elizabeth O., Anna Gassman-Pines, Dania V. Francis, and Christina M. Gibson-Davis. 2017. "Linking Job Loss, Inequality, Mental Health, and Education." *Science* 356(6343): 1127–28.

Anderson, Margo, and William Seltzer. 2007. "Challenges to the Confidentiality of U.S. Federal Statistics, 1910–1965." *Journal of Official Statistics* 23(1): 1–34.

Austin, Megan, R. Joseph Waddington, and Mark Berends. 2019. "Voucher Pathways and Student Achievement in Indiana's Choice Scholarship Program." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(3): 20–40. DOI: 10.7758/RSF.2019.5.3.02.

Bailey, Drew, Greg J. Duncan, Candice L. Odgers, and Winnie Yu. 2017. "Persistence and Fadeout in the Impacts of Child and Adolescent Interventions." *Journal of Research on Educational Effectiveness* 10(1): 7–39.

Blau, Peter, and Otis Dudley Duncan. 1967. *The American Occupational Structure*. New York: John Wiley & Sons.

Brummet, Quentin, and Emily Penner. 2017. "After School: An Examination of the Career Paths and Earnings of Former Teachers." Paper presented at APPAM 2017. Chicago (November 2–4, 2017).

Campbell, Frances, Gabriella Conti, James J. Heckman, Seong Hyeok Moon, Rodrigo Pinto, Elizabeth Pungello, and Yi Pan. 2014. "Early Childhood Investments Substantially Boost Adult Health." *Science* 343(6178): 1478–85.

Candelaria, Christopher A. 2015. "Rethinking Teacher Effects on Student Achievement." Paper presented at the Association for Education Finance and Policy. Washington, D.C. (February 26–28, 2015).

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2011. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR." *Quarterly Journal of Economics* 126(4): 1593–660.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014a. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review* 104(9): 2593–632.

———. 2014b. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review* 104(9): 2633–79.

Denice, Patrick, and Betheny Gross. 2016. "Choice, Preferences, and Constraints: Evidence from Public School Applications in Denver." *Sociology of Education* 89(4): 300–20.

Di, Qian, Yan Wang, Antonella Zanobetti, Yun Wang, Petros Koutrakis, Christine Choirat, Francesca Dominici, and Joel D. Schwartz. 2017. "Air Pollution and Mortality in the Medicare Population." *New England Journal of Medicine* 376(26): 2513–22.

Dodge, Kenneth A. 2009. "Community Intervention and Public Policy in the Prevention of Antisocial Behavior." *Journal of Child Psychology and Psychiatry* 50(1–2): 194–200.

———. 2011. "Context Matters in Child and Family Policy." *Child Development* 82(1): 433–42.

Dodge, Kenneth A., W. Benjamin Goodman, Robert A. Murphy, Karen O'Donnell, Jeannine Sato, and Susan Guptill. 2014. "Implementation and Randomized Controlled Trial Evaluation of Universal Postnatal Nurse Home Visiting." *American Journal of Public Health* 104(S1): S136–43.

Domina, Thurston, Andrew McEachin, Paul Hanselman, Priyanka Agarwal, NaYoung Hwang, and Ryan Lewis. 2016. "Beyond Tracking and Detracking: The Dimensions of Organizational Differentiation in Schools." RAND Labor & Population working paper no. WR1155. Santa Monica, Calif.: RAND Corporation.

Downing, Janelle, and Tim Bruckner. 2019. "Subprime Babies: The Foreclosure Crisis and Initial Health Endowments." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(2): 123–40. DOI: 10.7758/RSF.2019.5.2.07.

Drabiak-Syed, Katherine. 2010. "Lessons from *Havasupai Tribe v. Arizona State University Board of Regents*: Recognizing Group, Cultural, and Dignity Harms as Legitimate Risks Warranting Integration into Research Practice." *Journal of Health and Biomedical Law* 6(2): 175–226.

Fabelo, Tony, Michael D. Thompson, Martha Plotkin, Dottie Carmichael, Miner P. Marchbanks, and Eric A. Booth. 2011. "Breaking Schools' Rules: A Statewide Study of How School Discipline Relates to Students' Success and Juvenile Justice Involvement." New York: Council of State Governments Justice Center.

Fernandez, Roberto M., and Santiago Campero. 2017. "Gender Sorting and the Glass Ceiling in High-Tech Firms." *ILR Review* 70(1): 73–104.

Fernandez, Roberto M., and Colette Friedrich. 2011. "Gender Sorting at the Application Interface." *Industrial Relations: A Journal of Economy and Society* 50(4): 591–609.

Fernandez, Roberto M., and Jason Greenberg. 2013. "Race, Network Hiring, and Statistical Discrimination." In *Networks, Work and Inequality*, edited by Steve McDonald. Bingley, UK: Emerald Group Publishing.

Fernandez, Roberto M., and Brian Rubineau. 2019. "Network Recruitment and the Glass Ceiling: Evidence from Two Firms." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(3): 88–102. DOI: 10.7758/RSF.2019.5.3.05.

Fernandez-Mateo, Isabel, and Roberto M. Fernandez. 2016. "Bending the Pipeline? Executive Search and Gender Inequality in Hiring for Top Management Jobs." *Management Science* 62(12): 3636–55.

Finucane, Mariel McKenzie, Ignacio Martinez, and Scott Cody. 2017. "What Works for Whom? A Bayesian Approach to Channeling Big Data Streams for Public Program Evaluation." *American Journal of Evaluation* 39(1): 109–22.

Ganzeboom, Harry B. G., Donald J. Treiman, and Wout C. Ultee. 1991. "Comparative Intergenerational Stratification Research: Three Generations and Beyond." *Annual Review of Sociology* 17(1): 277–302.

Godechot, Olivier. 2013. "Financialization and Sociospatial Divides." *L'Année Sociologique* 63(1): 17–50.

Goerge, Robert M., and Emily R. Wiegand. 2019. "Understanding Vulnerable Families in Multiple Service Systems." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(2): 86–104. DOI: 10.7758/RSF.2019.5.2.05.

Grusky, David B., Michael Hout, Timothy M. Smeeding, and C. Matthew Snipp. 2019. "The American Opportunity Study: A New Infrastructure for Monitoring Outcomes, Evaluating Policy, and Advancing Basic Science." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(2): 20–39. DOI: 10.7758/RSF.2019.5.2.02.

Hällsten, Martin. 2014. "Inequality Across Three and Four Generations in Egalitarian Sweden: 1st and 2nd Cousin Correlations in Socio-Economic Outcomes." *Research in Social Stratification and Mobility* 35(1): 19–33.

Holbein, John B. 2017. "Childhood Skill Development and Adult Political Participation." *American Political Science Review* 111(3): 572–83.

Howard, Lanikque, Lisa Klein Vogel, Maria Cancian, and Jennifer L. Noyes. 2019. "Building Connections: Using Integrated Administrative Data to Identify Issues and Solutions Spanning the Child Welfare and Child Support Systems." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(2): 70–85. DOI: 10.7758/RSF.2019.5.2.04.

Ioannidis, John P. A. 2015. "Why Most Published Research Findings Are False." *PLoS Medicine* 2(8): e124.

Kessler-Harris, Alice. 1982. *Out to Work: A History of Wage-Earning Women in the United States*. Oxford: Oxford University Press.

Kim, ChangHwan, and Christopher R. Tamborini. 2019. "Are They Still Worth It? The Long-Run Earnings Benefits of an Associate Degree, Vocational Diploma or Certificate, and Some College." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(3): 64–85. DOI: 10.7758/RSF.2019.5.3.04.

King, Joseph, Andrew M. Penner, Nina Bandelj, and Aleksandra Kanjuo-Mrčela. 2017. "Market Transformation and the Opportunity Structure for Gender Inequality: A Cohort Analysis Using Linked Employer-Employee Data from Slovenia." *Social Science Research* 67 (September): 14–33.

Laurito, Agustina, Johanna Lacoe, Amy Ellen Schwartz, Patrick Sharkey, and Ingrid Gould Ellen. 2019. "School Climate and the Impact of Neighborhood Crime on Test Scores." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(2): 141–66. DOI: 10.7758/RSF.2019.5.2.08.

Li, Danielle. 2012. "Essays on the Organization of Science and Education." Ph.D. diss., Massachusetts Institute of Technology.

Liebler, Carolyn A., Renuka Bhaskar, and Sonya R. Porter. 2016. "Joining, Leaving, and Staying in the American Indian/Alaska Native Race Category Between 2000 and 2010." *Demography* 53(2): 507–40.

Ludwig, Jens, Greg J. Duncan, Lisa A. Gennetian, Lawrence F. Katz, Ronald C. Kessler, Jeffrey R. Kling, and Lisa Sanbonmatsu. 2013. "Long-Term

Neighborhood Effects on Low-Income Families: Evidence from Moving to Opportunity." *American Economic Review* 103(3): 226–31.

Meghir, Costas, Mårten Palme, and Marieke Schnabel. 2012. "The Effect of Education Policy on Crime: An Intergenerational Perspective." NBER working paper no. w18145. Cambridge, Mass.: National Bureau of Economic Research.

Miller, Portia, Elizabeth Votruba-Drzal, and Rebekah Levine Coley. 2019. "Poverty and Academic Achievement Across the Urban to Rural Landscape: Associations with Community Resources and Stressors." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(2): 106–22. DOI: 10.7758/RSF.2019.5.2.06.

Mills, Jonathan N., and Patrick J. Wolf. 2017. "Vouchers in the Bayou: The Effects of the Louisiana Scholarship Program on Student Achievement After 2 Years." *Educational Evaluation and Policy Analysis* 39(3): 464–84.

Mueller, Anna S., Tania M. Jenkins, Melissa Osborne, Arjun Dayal, Daniel M. O'Connor, and Vineet M. Arora. 2017. "Gender Differences in Attending Physicians' Feedback to Residents: A Qualitative Analysis." *Journal of Graduate Medical Education* 9(5): 577–85.

Murray, Brittany, Thurston Domina, Linda Renzulli, and Rebecca Boylan. 2019. "Civil Society Goes to School: Parent-Teacher Associations and the Equality of Educational Opportunity." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(3): 41–63. DOI: 10.7758/RSF.2019.5.3.03.

Open Science Collaboration. 2015. "Estimating the Reproducibility of Psychological Science." *Science* 349(6251): aac4716.

Park, Hyunjoon. 2014. "The Legacy of Disadvantaged Origins: Blocked Social Mobility of Descendants of Nobis in Late Joseon (Korea)." Ann Arbor: Nam Center for Korean Studies, University of Michigan.

Penner, Andrew M., Thurston Domina, Emily K. Penner, and AnneMarie Conley. 2015. "Curricular Policy as a Collective Effects Problem: A Distributional Approach." *Social Science Research* 52 (July): 627–41.

Penner, Emily K., Jane Rochmes, Jing Liu, Sabrina Solanki, and Susanna Loeb. 2019. "Differing Views of Equity: How Prospective Educators Perceive Their Role in Closing Achievement Gaps." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(3): 103–27. DOI: 10.7758/RSF.2019.5.3.06.

Persico, Claudia, David Figlio, and Jeffrey Roth. 2016. "Inequality Before Birth: The Developmental Consequences of Environmental Toxicants." NBER working paper no. w22263. Cambridge, Mass.: National Bureau of Economic Research.

Petersen, Trond, and Laurie A. Morgan. 1995. "Separate and Unequal: Occupation-Establishment Sex Segregation and the Gender Wage Gap." *American Journal of Sociology* 101(2): 329–65.

Petersen, Trond, and Ishak Saporta. 2004. "The Opportunity Structure for Discrimination." *American Journal of Sociology* 109(4): 852–901.

Poulton, Richie, Avshalom Caspi, Barry J. Milne, W. Murray Thomson, Alan Taylor, Malcolm R. Sears, and Terrie E. Moffitt. 2002. "Association Between Children's Experience of Socioeconomic Disadvantage and Adult Health: A Life-Course Study." *The Lancet* 360(9346): 1640–45.

Rafalow, Matthew H., Cynthia Feliciano, and Belinda Robnett. 2017. "Racialized Femininity and Masculinity in the Preferences of Online Same-Sex Daters." *Social Currents* 4(4): 306–21.

Reardon, Sean F. 2019. "Educational Opportunity in Early and Middle Childhood: Using Full Population Administrative Data to Study Variation by Place and Age." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(2): 40–68. DOI: 10.7758/RSF.2019.5.2.03.

Ruzek, Erik A., Thurston Domina, AnneMarie M. Conley, Greg J. Duncan, and Stuart A. Karabenick. 2015. "Using Value-Added Models to Measure Teacher Effects on Students' Motivation and Achievement." *Journal of Early Adolescence* 35(5–6): 852–82.

Song, Xi, and Cameron D. Campbell. 2017. "Genealogical Microdata and Their Significance for Social Science." *Annual Review of Sociology* 43(1): 75–99.

Sun, Min, Emily K. Penner, and Susanna Loeb. 2017. "Resource- and Approach-Driven Multidimensional Change: Three-Year Effects of School Improvement Grants." *American Educational Research Journal* 54(4): 607–43.

Temby, Owen F., and Ken R. Smith. 2014. "The Association Between Adult Mortality Risk and Family History of Longevity: The Moderating Effects of Socioeconomic Status." *Journal of Biosocial Science* 46(6): 703–16.

Tomaskovic-Devey, Donald. 1993. *Gender & Racial*

Inequality at Work: The Sources and Consequences of Job Segregation. Ithaca, N.Y.: Cornell University Press.

U.S. Census Bureau. 1996. "Puerto Rico, Virgin Islands, and the Pacific Island Territories." In *1990 Census of Population and Housing, History*, chapter 13. Washington: Government Printing Office.

Wightman, Patrick, and Sheldon Danziger. 2014. "Multi-Generational Income Disadvantage and the Educational Attainment of Young Adults." *Research in Social Stratification and Mobility* 35(1): 53–69.

Wilde, Melissa J. 2004. "How Culture Mattered at Vatican II: Collegiality Trumps Authority in the Council's Social Movement Organizations." *American Sociological Review* 69(4): 576–602.

Wynn, John D., Daniel A. Reyes, and Willard E. Caldwell. 2011. "United States Census 2010: A Description of the 2010 Census Operations and Data Products of the Island Areas, and How They Compare to the 50 States and the District of Columbia." Washington: U.S. Census Bureau.

Young, Cristobal, Charles Varner, Ithai Z. Lurie, and Richard Prisinzano. 2016. "Millionaire Migration and Taxation of the Elite: Evidence from Administrative Data." *American Sociological Review* 81(3): 421–46.

PART IV

Education

Voucher Pathways and Student Achievement in Indiana's Choice Scholarship Program

MEGAN AUSTIN, R. JOSEPH WADDINGTON, AND MARK BERENDS

This article examines the pathways that students can follow within the Indiana Choice Scholarship Program and the associations with their math and English language arts achievement in upper elementary and middle school. We analyze student-level longitudinal data by matching voucher and nonvoucher students to estimate the role of participating in the voucher program, taking advantage of the uniqueness of Indiana public and private schools taking the same standardized assessment over time. The different student pathways for using vouchers are related to student achievement with significant achievement losses for students who switch from a public to a private school with a voucher. Students who have always attended a private school, both before and after receiving a voucher, experience no significant changes in achievement.

Keywords: parental choice scholarships, vouchers, student mobility, school effects, fixed effects

Megan Austin is a researcher at American Institutes for Research. **R. Joseph Waddington** is assistant professor in the Department of Educational Policy Studies and Evaluation at the University of Kentucky, where he also holds a secondary appointment in the Martin School of Public Policy and Administration. **Mark Berends** is professor of sociology at the University of Notre Dame, where he directs the Center for Research on Educational Opportunity.

 Austin, Megan, R. Joseph Waddington, and Mark Berends. 2019. "Voucher Pathways and Student Achievement in Indiana's Choice Scholarship Program." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(3): 20–40. DOI: 10.7758/RSF.2019.5.3.02. Author order was determined randomly; each contributed equally to this paper. This paper was supported by Notre Dame's Center for Research on Educational Opportunity and Institute of Educational Initiatives, the Spencer Foundation, and the Walton Family Foundation. We are grateful to the Indiana Department of Education for providing access to the state administrative records and for supporting independent analyses. We are also grateful for the substantial feedback we have received from numerous colleagues, including David Agrawal, Bob Bifulco, Jo Blacketor, Kenneth Couch, Josh Cowen, Kenneth Dodge, John Elcesser, Joe Ferrare, Phil Gleason, Joshua Goodman, Doug Harris, Brian Jacob, Helen "Sunny" Ladd, Hank Levin, Maddy Mavrogordato, Andrew Penner, Ann Primus, Macke Raymond, Amy Ellen Schwartz, James Shuls, Erdal Tekin, Genia Toma, Patrick Wolf, Ron Zimmer, three anonymous reviewers, and participants at the Using Administrative Data for Science and Policy conference held in January 2018 at the Russell Sage Foundation headquarters in New York. All opinions expressed in this paper represent those of the authors and not necessarily the institutions with which they are affiliated. All errors in this paper are solely the responsibility of the authors. Direct correspondence to: Megan Austin at maustin@air.org, 10 S. Riverside Plaza, Suite 600, Chicago, IL 60606.

One school choice option for families is providing them with scholarships, or vouchers, to give parents the opportunity to send their children to the school of their choice. Using public funds otherwise spent on a student's public school education by the district or state, vouchers are allocated to families who use them to make partial or full tuition payments at private schools. The number of voucher bills that states have passed and the number of states introducing voucher bills increased dramatically in the last few years (Berends 2018). The number of students participating in voucher programs across the nation also increased significantly in the last decade, although the approximately 182,700 students receiving vouchers remains a small fraction of the total number of U.S. students (for a description of ongoing voucher programs, see EdChoice 2018). The Indiana Choice Scholarship Program (ICSP), authorized in 2011, provides vouchers to more than thirty-five thousand of these students—serving more than 20 percent of all voucher students nationwide, but only about 3 percent of the 1,139,822 K–12 students in Indiana (Indiana Department of Education 2018).

As vouchers have become more prominent as a form of school choice, they have been highly contested and debated in educational policy, the media, and research communities. Dating back to studies of the first voucher program in Milwaukee in the 1990s, researchers have used administrative data to evaluate the effects of vouchers on student outcomes, and although these analyses been central to our understanding of voucher effects, findings have been mixed (see Berends, Cannata, and Goldring 2011). Although policymakers expected that evaluation of the Milwaukee program and subsequent voucher programs in other locales would settle disputes about the effects of vouchers on student outcomes, the research findings instead have been nuanced and mixed with some positive, some negative, and some null effects on student academic achievement outcomes (for reviews, see Austin and Berends 2018; Epple, Romano, and Urquiola 2017; Shakeel, Anderson, and Wolf 2016; Zimmer and Bettinger 2015).

The use of Indiana Department of Education (IDOE) administrative data—in particular, deidentified longitudinal student records linked to the schools students attended—has been central to the evaluation of voucher program effectiveness. Because the administrative data are comprehensive in that they include annual scores on the same accountability test for the population of students in both public and private schools in Indiana, these data allow us to use rigorous nonexperimental methods to assess the impact of the ICSP on student outcomes. Through a data-sharing agreement with IDOE established within a researcher-practitioner partnership, we have access to longitudinal information on students' test scores before and after moving to a new school using a voucher. Recently, we found that students who use an Indiana Choice Scholarship to move from a public to a private school do worse, on average, on math tests in their new private school than they did in their prior public school and about the same on English language arts (ELA) tests (Waddington and Berends 2018).

Administrative data, which include records for the full population of Indiana students, also enable us to evaluate one potential explanation for this negative outcome—the disruption of transitioning to a new school (Grigg 2012; Langenkamp 2010, 2011; Schwartz, Stiefel, and Cordes 2017; Xu, Hannaway, and D'Souza 2009). In addition, because we also collected qualitative interview data from principals, teachers, parents, and students from thirteen private schools in Indiana, we were able to identify the intended and unintended pathways that families took to enter the ICSP program (Austin 2019). Understanding the ICSP pathways other than the initial one where students switched from a public to a private school with a voucher is critical for assessing the ICSP because of the substantial numbers of students who take these additional pathways and the potential for heterogeneous effects of vouchers by pathway.

The administrative data are analyzed in the context of a researcher-practitioner partnership between the University of Notre Dame's Center for Research on Educational Opportunity and IDOE. Spanning several administrations, the partnership's mission is to conduct independent, rigorous research to inform educational

policy and decision making in Indiana (Berends and Austin 2017). Thus, analyses of IDOE administrative records are done in such a way to inform decision making at the state, district, and school levels.

GROWTH AND EVOLUTION OF THE ICSP

Compared with other statewide programs in Ohio (Figlio and Karbownik 2016) and Louisiana (Abdulkadiroglu, Pathak, and Walters 2018; Mills and Wolf 2017), the Indiana voucher program is unique. First, important changes occurred over time in Indiana's voucher program, opening access to more families. In 2011, when the program began, students needed to attend a public school before using a voucher to attend a private school.[1] In 2013, the criteria for eligibility expanded to include kindergarten students, siblings of voucher students, special education students, and those located in the attendance zones of failing public schools; in addition, no cap was set on the number of eligible Indiana students who can receive a Choice Scholarship. With these additional pathways, in the 2017–2018 school year, 57 percent of students receiving a voucher have never attended a public school (Indiana Department of Education 2018). Moreover, compared with the 3,911 students participating when the voucher program began in 2011–2012, 35,458 students participated in 2017–2018, attending 318 private schools (Indiana Department of Education 2018).

Second, the administrative data on students, teachers, and schools in Indiana allow us to longitudinally track students over time with the same outcomes whether they are attending a public or a private school. Unlike other state administrative data used to analyze the effects of statewide voucher programs that only administer tests to students receiving a voucher (Louisiana, Ohio, and Washington, D.C.), the Indiana data include scores on the same tests for not only voucher students but also other students in public and private schools. Such comprehensiveness enables us to produce more precise estimates of students' achievement as they move into and out of the ICSP and between public and private schools than would be possible without comprehensive achievement data on the population of Indiana test takers. Our data also differ from the data used in many articles in this issue: because all students are required by law to take the state standardized test, our data are comprehensive. The data also come from a single state agency. As a result, we do not have to merge data from multiple sources or address concerns about selection into the dataset itself.

Third, recent quasi-experimental findings of negative effects on students across voucher programs in Indiana, Ohio, Washington, D.C., and New Orleans have received a great deal of attention. For example, in Indiana, Joseph Waddington and Mark Berends find no change in ELA performance but significant losses in math performance in students' first two years after switching to a private school, which persisted for students through four years of participation in the program (2018). In Ohio, David Figlio and Krzysztof Karbownik find significant negative effects on students' math and reading scores following a move into a voucher school (2016). Atila Abdulkadiroglu, Parag Pathak, and Christopher Walters find significant negative effects on Louisiana students' math and reading performance in their first year after transitioning from a public to a private school in the Louisiana Scholarship Program (2018). Jonathan Mills and Patrick Wolf find that by their second year in a private school, Louisiana voucher recipients no longer had worse outcomes than their public school peers in reading, but that negative effects in math performance persisted (2017).

These studies provided only black-box estimates of the effects of the voucher programs across all students; to better understand these findings, additional research is needed that seeks to identify potential mechanisms and heterogeneous effects for different groups of students. This article takes that next step by

1. An additional pathway by which students could qualify for the voucher program during the 2011–2012 school year was by having previously received a corporate tax-credit scholarship through a scholarship granting organization (SGO). Most students receiving vouchers through the SGO scholarship pathway were previously enrolled in private schools and may not have spent any time in a public school.

disaggregating students according to the pathway by which they came to receive a voucher.

LITERATURE REVIEW

Most studies of student transitions from public to private voucher schools evaluate overall voucher program effectiveness but do not contextualize students' transitions into a voucher school as a special case of school transitions more generally (Abdulkadiroglu, Pathak, and Walters 2018; Figlio and Karbownik 2016; Mills and Wolf 2017; Waddington and Berends 2018). School choice expands the number and type of schools in which students can choose to enroll. An expanded choice set of schools, along with the financial support provided by school voucher programs, may create incentives to move between schools and potentially increase the prevalence of school transfers as students move into schools that previously were not financially available to them (Lareau and Goyette 2014; Schwartz, Stiefel, and Cordes 2017).

Research on the effects of school transitions on student outcomes often finds that transitioning to a new school is associated with a decrease in student test performance, especially in the short term (Grigg 2012; Langenkamp 2010, 2011; Hanushek, Kain, and Rivkin 2004; Mehana and Reynolds 2004; Schwartz, Stiefel, and Cordes 2017; Xu, Hannaway, and D'Souza 2009). To better isolate the true effect of voucher programs on student outcomes, it is important to consider the school transitions that occur as part of a voucher program within the larger context of transitions between schools. Considering the impact on student achievement of making any school transition provides valuable insight into the potential consequences of transitions that take place in the context of a school voucher program. This study examines the extent to which the changes in student performance attributed to voucher programs in the studies cited may be intertwined with changes related to transitioning between any two schools, independent of voucher receipt.

Transitions. Students make transitions between schools for both structural reasons, such as transitioning from elementary to middle school, and nonstructural ones, such as residential moves or academic or social struggles (Benner 2011). Nationally, limited data on student transitions show that transitions are most common among elementary school students and least common among high school students: in 2000, NAEP data showed that 35 percent of fourth-grade students, 21 percent of eighth-grade students, and 10 percent of twelfth-grade students had changed schools at least once in the past two years (Anderson 2017; Rumberger 2015). Among the kindergarten cohort of 1998, only 34 percent of students remained in the same school from kindergarten to fifth grade; 42 percent made one change; and 24 percent made two or more changes (Rumberger 2015).

Generally, research has suggested that transitions have a negative effect on student outcomes, especially in the first year following a move (Schwartz, Stiefel, and Cordes 2017). However, "strategic moves" between schools are associated with better outcomes than "reactive moves" (Rumberger 2015; Xu, Hannaway, and D'Souza 2009). It is difficult to disentangle the effects of any transition between schools from the effects of voucher school attendance, given the central role that the transition from public to private school plays in the estimation of voucher program effects (Waddington and Berends 2018). This article provides an initial step in disentangling the influence of these different types of transitions on students' academic achievement.

Structural and nonstructural transitions at different grade levels. Research on transitions does not always distinguish between structural and nonstructural transitions; when it does, it tends to focus on structural transitions—from middle school to high school or, to a lesser extent, from elementary school to middle school (Benner 2011; Langenkamp 2009, 2010, 2011; Langenkamp and Carbonaro 2018; Schiller 1999). However, recent causal research using New York City school administrative data that does distinguish between structural and nonstructural transitions finds different patterns by transition type: structural transitions have a significant negative short-term effect on ELA performance and significant negative medium-term effects on both ELA and math performance. In contrast, students who make nonstructural transitions show a significant positive increase in ELA performance (Schwartz, Stiefel, and Cordes 2017).

Some researchers have hypothesized that age at transition is an important factor in how successfully students navigate a transition (Eccles, Lord, and Midgley 1991). For example, for elementary school students, school transitions are associated with lower achievement and a higher likelihood of grade retention (Alexander, Entwisle, and Dauber 1996; Gruman et al. 2008). Some of these associations disappeared once family background characteristics were included, consistent with research showing that students who experience instability in other areas of their lives—family instability, residential moves—both are more likely to make school transitions and may be particularly vulnerable to negative effects of school transitions (Langenkamp 2010, 2011). Other research suggests that transitions during middle school are particularly hard on students (Herbers, Reynolds, and Chen 2013) and are associated with declines in math achievement and grade point average (Anderson 2017). However, other research finds consistent declines in student achievement associated with transitions across grades five through eight (Alspaugh and Harting 1995). The declines across all grade levels were concentrated in the first year following the transition, with student achievement recovering to pretransition levels by the second year following the transition.

School culture. School transitions may lead to decreased performance for interpersonal or academic reasons. Students who transition to a new school must leave relationships with friends and teachers or other school staff and form new friendships and connections. Especially in the case of nonstructural moves, classes in students' new schools often do not align seamlessly with their prior coursework, leaving them ahead or behind their new peers academically (Rumberger 2015). In addition, voucher programs increase the incidence of transitions between school sectors—for example, the ICSP as originally written required that students move from a public school into a private school; other voucher programs similarly target students seeking to leave their public schools to attend a private school. A transition to a school in a different sector requires students to adapt not only to a new school but also to a new type of school.

Students' ability to successfully adapt to their new schools is shaped by the school culture of both their prior school and the school they enter (Langenkamp 2009; Langenkamp and Carbonaro 2018). The school culture of both schools influences the effect of a transition on academic outcomes (Eccles, Lord, and Midgley 1991). A positive school climate, strong relationships with other students, and higher feelings of bonding with teachers in students' prior school all provide protection against a drop in student outcomes following a school transition—and in some cases are associated with higher performance following a transition.

School culture in the receiving school also influences students' outcomes following a transition (Langenkamp 2009; Langenkamp and Carbonaro 2018). Private schools, of which many in Indiana are Catholic, have a unique school culture that facilitates social bonding (Coleman and Hoffer 1987; Bryk, Lee, and Holland 1993; Hallinan 2006; Hoffer 2009). As a result, transitioning into a private school may influence students' transition experiences differently than transitioning between schools within a sector.

DATA AND MEASURES

For this study, we used seven years (school years 2009–2010 through 2015–2016) of students' longitudinal demographic and test score records. We obtained these data through a data-sharing agreement with the Indiana Department of Education. Similar to those of Sean Reardon (2019), these records include the full population of students in grades three through eight who attended public (traditional, charter, and magnet) and private schools (including voucher and nonvoucher students) and who participated in the Indiana Statewide Testing for Educational Progress Plus (ISTEP+) program. Testing took place each spring in math and ELA.[2]

One unique feature of these data and of

2. The ISTEP+ is aligned to the Indiana Academic Standards and serves as the main accountability-linked assessment for Indiana students in grades three through eight. It is vertically equated across grades and consists of multiple choice, constructed response, and extended response items scored using item response theory

studying Indiana's voucher program is the sizeable number of private schools that participated in the ISTEP+ testing program and other state reporting (318 schools statewide as of 2018). Participation in statewide testing is required for private schools to participate in the voucher program.[3] However, many private schools, including nearly all K–8 Catholic schools, participated in statewide testing well before the start of the voucher program as part of their accreditation process. All students in voucher-participating private schools took the test, whether they received a voucher or not.

This robust participation in annual statewide testing by private schools results in a comprehensive set of administrative records that provide multiple advantages in analysis of voucher student outcomes. First, the robust data allow us to make apples-to-apples achievement comparisons over time, both between voucher private and nonvoucher public school students and between voucher and nonvoucher private school students. Second, the robust number of participating schools allows for more generalizability of our findings across a broader spectrum of Indiana's private schools. Third, because each student's testing records are longitudinally linked, we can observe changes in an individual student's achievement over time, both before and after receiving a voucher and regardless of school sector. Fourth, the longitudinal linking of student records between public and private schools also allows us to estimate the impacts of transitioning between schools and sectors. Fifth, we can examine multiple pathways of entry into Indiana's Choice Scholarship Program, including students previously enrolled in private schools.

Measures. Students' ISTEP+ test scores in math and ELA are our outcomes of interest. We standardized each student's annual scaled test scores relative to the mean and standard deviation (SD) of students statewide within each subject, grade, and year of testing.[4] The units of our outcomes are SDs from the state average of all test takers within each tested grade.

We also created indicators for the following student-level demographic and background covariates: each student's sex, race-ethnicity, free or reduced-price lunch status, English-language learner (ELL) status, special education status, and grade level. We created an indicator for grade retention from the previous year. We also observed whether a student receives a voucher in each year, enabling us to construct an annual indicator of voucher receipt as well as indicators for the number of years since first receiving a voucher.

Along with voucher recipient status, we also observed the student's school of record within each year. We merged additional school-level data from the Common Core of Data and the Private School Universe Survey to augment and enhance the available school-level data from IDOE.[5] We then created binary indicators of the school type (public, charter, magnet, Catholic,

methods. Reliability coefficients range from 0.88 to 0.94 in ELA and 0.88 to 0.95 in math (Indiana Department of Education 2011). Annual testing in grades three through eight occurs in math and ELA and less frequently in other subjects such as social studies or science.

3. The Indiana Department of Education holds private schools participating in voucher programs accountable through their performance on the ISTEP+ assessment by restricting their ability to enroll students receiving vouchers should the school have two consecutive years of poor testing performance.

4. Although the ISTEP+ is vertically equated, we do not use untransformed raw scaled scores for our outcome because the variation in scales differs between grade levels. This introduces additional measurement error; however, we adjust for differences between years and across tests by controlling for grade-by-year fixed effects in all models. We standardized scores across all test takers, whether public or private, voucher or nonvoucher.

5. The Common Core of Data (CCD) contains annual demographic and background information for the universe of public schools. Similarly, the biennial Private School Universe Survey (PSS) contains similar information for private schools. We applied CCD data to all public schools for each corresponding year, except 2015–2016, for which we use data from the 2014–2015 CCD. Similarly, we applied PSS data from the most recent prior year to all private schools. Private schools in the IDOE data from 2011–2012 through 2015–2016 contain PSS information from 2011–2012, the last year of publicly available data.

or other private) and manually entered any missing school type information from other IDOE public records and internet searches.

Using annual school records, we created a set of indicators for the types of move a student made into or out of the voucher program across several pathways. These pathways include first receiving a voucher and using it to transition from a public to a private school, first receiving a voucher while enrolled in a private school, first receiving a voucher in a private school and later transitioning to a public school, and first transitioning from a private to a public school (to become eligible to receive a voucher) and transitioning back to a private school with a voucher the subsequent year. These variables serve as the treatment indicators in all analyses.

We also created two student-level mobility indicators to identify any school changes students made. The structural move indicator is for students who change schools between years in the course of normal grade progression. The nonstructural move indicator is for students who switch schools for any other reason (we do not observe the underlying reasons). Both variables indicate a switch only in the year (t) immediately after the switch, even though the switch takes place between years $t–1$ and t.[6] We then created a set of interactions between each of the mobility indicators and the student's grade level at time ($t–1$), the baseline year before a transition. For voucher students moving from public to private schools, this is the year prior to receiving a voucher. For voucher students moving from private to public schools, this is the final year of receiving a voucher before transitioning to a public school.

SAMPLE AND ESTIMATION STRATEGY

Our main research questions involve unpacking the academic achievement gains that voucher students make and how the timing of receiving a voucher (which grade level) and the type of pathway alters that relationship. Ideally, we would randomly assign vouchers for students to attend private schools and then draw comparisons between the treatment and control groups. However, in Indiana, vouchers were not randomly assigned as part of the ICSP. In addition, the numerous pathways through which a student can qualify for a voucher further complicate any research design to understand voucher outcomes in Indiana.

Without random assignment of vouchers or a natural experiment such as a lottery, any assessment of the effects of Indiana's voucher program is subject to selection bias due to the many unobservable factors that may influence whether (and when) a family may choose to apply for a voucher.[7] Thus, we cannot simply compare the achievement of voucher and nonvoucher students. However, we can use the IDOE's robust longitudinal administrative dataset, which maximizes sample sizes in each pathway, in conjunction with our estimation approach and the creation of comparison groups to mitigate selection bias. For both the sample construction and estimation strategies, we drew upon important lessons from within-study comparison research that uses nonexperimental approaches to replicate experimental estimates (Bifulco 2012; Cook, Shadish, and Wong 2008; Fortson et al. 2014) and the implementation of those lessons in the nonexperimental evaluation of charter schools (Angrist, Pathak, and Walters 2013; Dobbie and Fryer 2013, 2017) and voucher programs (Waddington and Berends 2018).

We enforced several data restrictions prior to sample construction. These restrictions included requiring each student to have at least three years of test scores, including two years before receiving a voucher (a pre-baseline and baseline year), and thus exclude any voucher students we observed receiving a voucher beginning with their first observation in our data. After these restrictions, we had 9,955 voucher students available for our sample construction

6. The association between mobility and student achievement is negative (Schwartz, Stiefel, and Cordes 2017). These indicators help us parse out the impact of switching schools from any voucher impacts. This is particularly important for the groups of voucher students we investigate who change schools after receiving a voucher.

7. More than 97 percent of families who apply for a voucher for their child receive a voucher. Based on this high take-up rate along with qualitative interviews with parents and schools, we conclude that families apply for a voucher only if they know they are eligible.

approach, which we describe in the following section.

Voucher Student Sample and Eligibility Pathways

A student must meet several eligibility criteria to qualify for a voucher in Indiana. One universal criterion is based on family income. The voucher income thresholds based on household size directly correspond to the thresholds for free or reduced-price lunch eligibility (Indiana Department of Education 2018).[8] Students in families at or below the income threshold for reduced-lunch eligibility can receive a "full" voucher for up to 90 percent of tuition at a private school. Students in families at or below 150 percent of the income threshold for reduced-lunch eligibility can receive a "half" voucher for up to 50 percent of tuition.

We focus on the results for students who receive full vouchers because we can construct a counterfactual group of nonvoucher recipients consisting of individuals who are voucher eligible by way of receiving free or reduced-price lunch. We refer to this group of voucher students as low income and include students who either receive a full voucher or had received free or reduced-price lunch in the two years prior to receiving a voucher.[9] By focusing on this low-income group of voucher and nonvoucher students, we can reduce issues with selection bias in our empirical models and avoid introducing additional unobserved variation based on family income that pertains to voucher eligibility.[10]

A second criterion from the initial implementation of the policy required students to have attended a public school (either traditional public, charter, or magnet) for at least one year immediately prior to receiving a voucher. In our cleaned data, 5,219 students moved from a public to a private school for the first time after receiving a voucher.[11] In addition, we observed 199 students who were once enrolled in a private school without a voucher, then left for one year to attend a public school to qualify for a voucher, and then returned to a private school the following year after receiving a voucher. This is a unique group of students who, by our hypothesis, "game the system" to qualify for a voucher. However, this gaming process requires students to make at least two moves between different schools.

After the voucher eligibility criteria expanded beginning in the 2013–2014 school year, additional students became eligible to receive a voucher, including students who had never attended a public school.[12] We observe an additional 4,537 students who were enrolled in a private school without a voucher before receiving a voucher for the first time.[13] For all voucher students, whether previously enrolled in a public school or not, we can establish a baseline

8. Although free or reduced-price lunch status is a rough indicator of students' family income levels, it is the criterion used in Indiana's voucher policy, so we also use it in this study.

9. Because income fluctuates, we wanted to account for indications that a family is low income in either the year before receiving a voucher (baseline year) or the year after. We use a similar procedure for public school students.

10. Half voucher students have higher achievement before receiving a voucher and are less diverse demographically and academically. Because we do not have information about family income for nonvoucher recipients, a key voucher qualification criterion, we are unable to produce trustworthy estimates to compare all voucher and nonvoucher students. Further descriptive details for recipients of half voucher recipients are available on request.

11. Some of these students eventually exit a private school after receiving a voucher and return to a public school. We include both students who remain in a private school and those who return to a public school in our analysis.

12. From the start of the voucher program in the 2011–2012 school year, students who had previously received a scholarship from an SGO were also eligible to receive a voucher without having previously attended a public school. For our analysis, we do not distinguish this group of students from non-SGO students previously enrolled in private schools as the schooling context is no different (all SGO and non-SGO voucher qualifying students were previously enrolled in private schools).

13. As with the group of public school students receiving a voucher and attending a private school, some of the previously enrolled voucher students eventually exit a private school after receiving a voucher and attend a public school. We include both groups of students in our analysis.

year based on the year immediately prior to receiving a voucher for the first time.

Comparison Student Sample

We constructed unique comparison groups for each of the pathways of voucher recipients to estimate the influence of receiving a voucher on student achievement. One important takeaway from the within-study comparison literature in education research and the broader quasi-experimental literature is that treatment and comparison groups should be drawn from the same geographic location and time frame—that is, the same school (Cook, Shadish, and Wong 2008). Therefore, we construct comparison groups for each group of voucher recipients that match students based on several geographic, time, and demographic criteria.

To match voucher and nonvoucher students, we develop matching cells that exactly match students based on being of the same race and sex and in the same grade, year, and school as a voucher student at baseline. For students using a voucher to transition from a public to a private school, we constrain the comparison sample to the group of public school students who never receive a voucher and are in the same matching cell as the voucher student's baseline year.[14] Thus, after baseline, our estimates compare voucher students who are attending a private school with a counterfactual group of peers who remained in a public school.

We use a similar approach for matching voucher students who were initially enrolled in a private school before receiving a voucher. This includes the group of students who initially left a private school without a voucher, always remained enrolled in a private school, or attended a public school for one year and then returned to a private school after qualifying for and receiving a voucher. Here, we constrain the comparison sample to the group of private school students who never receive a voucher and are in the same matching cell as the voucher student's baseline year.[15] After baseline, we compare voucher students attending a private school with a counterfactual peer group who remained in a private school without a voucher.

Because voucher eligibility is based on family income, we include only low-income public and private comparison students. As described, these students are most likely to be eligible to receive a full voucher. We constrain our comparison sample to nonvoucher students who received free or reduced-price lunch in the baseline or first post-baseline year. This limits the number of nonvoucher comparison students in private schools who would be eligible for a full voucher, given that most low-income students in private schools apply for and receive a voucher once they met all the eligibility criteria.[16] Our estimates comparing students starting in private schools before receiving a voucher with their nonvoucher private school peers therefore have large standard errors.

The exact matching of students based on several criteria helps mitigate selection bias in terms of who does and does not receive a voucher when conducting our estimates. By matching students by school, grade, and year, the voucher and comparison students are compared, beginning at baseline, in the same schooling contexts. Also, matching exactly based on a student's race and sex further accounts for variation in the selection process. For example, if students of a certain race-ethnicity were more likely to participate in the voucher program, we are now comparing them

14. Some public school students have peers who leave to attend a private school with a voucher across several grades and years. To avoid replicating individual students in our sample, we randomly choose which of a given public school student's years serves as the baseline year.

15. Some private school students have peers who later receive a voucher in a private school across several grades and years. To avoid replicating individual students in our sample, we randomly choose which of a given private school student's years serves as the baseline year.

16. Not all students previously enrolled in a private school that are eligible for a voucher based on income meet the other eligibility criteria, which allows us to create a counterfactual group of private school students. The other eligibility criteria, without having spent the previous year in a public school, include being a sibling of a student who is receiving a voucher, receiving a voucher since kindergarten, being classified as a special education student, residing in a public school district with an F state rating, or having received an SGO scholarship.

with their same-race peers who should share the same likelihood of selection into the program based on the observable characteristic of race-ethnicity.

This approach shares characteristics with propensity score matching (Rosenbaum and Rubin 1983). Both approaches rely on the matching of students based on a finite set of observable criteria associated with the selection process. The exact matching process is more precise than propensity score matching; however, the number of matching criteria must be limited when using exact matching to achieve a desirable match rate. We believe that race-ethnicity, sex, and sharing a baseline year, grade, and school are a reasonable set of criteria to mitigate selection bias based on prior studies (see also Angrist, Pathak, and Walters 2013; Dobbie and Fryer 2017). Yet, as with propensity score matching, we are adjusting for key observable differences between voucher and nonvoucher students. Despite having a comparable sample, our estimates of the voucher program's effects remain subject to bias based on any unobservable characteristics that may drive selection into the program through any pathway. We further detail these concerns and our approach to estimation in the following section.

Estimation Strategy

Our basic empirical models rely on the inclusion of matched cell fixed effects as an approach to mitigate unobserved, between-group confounding influences. We estimated these models for each individual year after treatment, resulting in a total of three individual models to estimate the voucher program associations with student achievement in the first, second, and third year after receiving a voucher. We also estimated separate models for each outcome subject (math and ELA), though the structure of the equation remained the same.

We begin with an ordinary least squares regression model with several covariates as shown in equation (1) for students starting in public schools.

$$Y_{icgt} = \alpha + \beta_1 PubtoPriv_{icgt} + \beta_2 PubtoPrivtoPub_{icgt} + \pi Y_{icg\,(t=0)} + \omega Y_{icg\,(t=-1)} + \delta \mathbf{X}_{icg\,(t=0)} + \theta_g + \tau_c + \upsilon_{icgt}. \quad (1)$$

Here, the achievement level (Y) for each student (i) in matching cell (c) in grade (g) and year (t) is a function falling into one of three groups of students who received a voucher and attended a private school. This includes students who switch from public to private schools and remain there after receiving a voucher ($PubtoPriv_{icgt}$), students who switch from public to private schools after receiving a voucher but later return to a public school ($PubtoPrivtoPub_{icgt}$), and the matched public school comparison peers. We included in our preferred models two measures of a student's prior achievement in the same subject as the outcome, one at baseline ($Y_{icg(t=0)}$) and one pre-baseline ($Y_{icg(t=-1)}$). Because lagged achievement scores are endogenous in the post-baseline years, these controls remain as the baseline and pre-baseline achievement measures for our estimates in the second, and third years post-baseline. By controlling for prior achievement, we are netting out the differences between voucher and nonvoucher student achievement within any one matching cell that may drive selection into receiving a voucher. And, by accounting for baseline and pre-baseline achievement, we are accounting for pre-treatment trends in achievement that may drive selection into receiving a voucher. Thus, the voucher coefficients in this model (β_1 and β_2) represent the within-cell difference between voucher and nonvoucher students in the achievement gain (or loss) from baseline in a given post-baseline year.

If voucher students experienced a substantial drop in student performance, this could be a signal to parents to apply for a voucher, and in the case of students beginning in a public school, to change schools. This phenomenon is known in the job-training literature as Ashenfelter's Dip, and without adjusting for pre-treatment differences in achievement, our results may be biased (Ashenfelter 1978). We can test for differing pre-treatment trends between voucher and nonvoucher students by incorporating a series of interaction terms between the voucher student indicators and the prior achievement variables. If these estimates are not statistically significant, we should not be concerned about pre-treatment trends driving our estimated outcomes.

We also controlled for baseline student char-

acteristics ($X_{icg(t=0)}$) that include ELL and special education status, structural and nonstructural school changes in year (t) interacted with grade level (g), public school type (magnet or charter versus traditional public), and school locale (suburban or town-rural versus urban). The matched cell fixed effect (τ_c) in the model accounts for unobserved differences between voucher and nonvoucher students within each race-sex-grade-year-public school cell at baseline. These also inherently account for systematic differences in exams across years as students within each cell take exams always within the same calendar year posttreatment. Grade fixed effects (θ_g) account for differences across exams and over time. The term υ_{icgt} represents cluster-robust standard errors to account for serial correlation among students within the same baseline cohort (same grade, year, and school) and is used in all models.

Similarly, we estimate equation (2) for students starting in private schools.

$$Y_{icgt} = \alpha + \gamma_1 PrivtoPub_{icgt} + \gamma_2 AlwaysPriv_{icgt} + \pi Y_{icg(t=0)} + \delta \mathbf{X}_{icg(t=0)} + \theta_g + \tau_c + \upsilon_{icgt}. \quad (2)$$

This equation is nearly identical to equation (1), but now only includes students who received a voucher while in a private school and later switched to a public school ($PrivtoPub_{icgt}$), students who received a voucher while in a private school and remained there ($AlwaysPriv_{icgt}$), and the matched private school comparison peers. The matched cell fixed effect (τ_c) in this model accounts for unobserved differences between voucher and nonvoucher students within each race-sex-grade-year-private school cell at baseline. Thus, in equation (2), the voucher coefficients (γ_1 and γ_2) represent the within-cell difference in the achievement gain (or loss) from baseline in a given post-baseline year between voucher and nonvoucher student achievement.[17]

For the sample of voucher students starting in private schools, we only control for baseline achievement ($Y_{icg(t=0)}$) and do not include a control for pre-baseline achievement. We do not have a large enough sample of students in private schools with two years of pre-baseline data without a voucher. To ensure the validity of these results, we conducted a robustness check with our public to private voucher sample and found no differences in our results with our preferred model (1) when we remove the control for pre-baseline achievement.

Across all voucher types, most of our estimates are constrained to students in grades five to eight.[18] We estimate varying voucher impacts by type of move (structural-nonstructural) and grade level by including interactions with the voucher pathway indicators in equations (1) and (2).

RESULTS

Table 1 includes a descriptive comparison of all types of voucher students with their matched nonvoucher public and private school peers at baseline. The numbers of students reported in table 1 represent the matched students. Compared with traditional public school students, a higher proportion of students who switch from a public to a private school with a voucher are Latino/a or ELLs. These students have higher average math and ELA scores than their public school peers the year before receiving a voucher (baseline); they have lower math scores but higher ELA scores in the first year after receiving a voucher.

17. We used a modified version of equation (2) to separately estimate impacts for students who switched from private to public schools and then returned to a private school after qualifying for and receiving a voucher. This allowed us to obtain estimates for the year a student spent in a public school and the years after they returned to a private school with a voucher.

18. This could limit the generalizability of our findings, but we cannot yet estimate the relationship between achievement and students always observed receiving a voucher in our data. Since the Indiana voucher law change in 2013 to open pathways and not require students to previously attend a public school before receiving a voucher, "students always observed receiving a voucher" will appear in the data for the first time in the 2017–2018 school year. For example, first graders receiving a voucher and always attending a private school was only possible starting in the 2013–2014 school year; these students are now fifth graders in the 2017–2018 school year. Likewise, kindergartners in 2013–2014 will be fifth graders in the 2018–2019 school year.

Table 1. Descriptive Statistics for Voucher, Public Comparison, and Private Comparison Students

	Public Schools at Baseline			Private Schools at Baseline			
	Public to Private with Voucher	Public to Private with Voucher to Public	Public Comparison	Private with Voucher to Public	Always Private with Voucher	Private to Public to Private with Voucher	Private Comparison
Students	3,288	686	42,835	126	616	151	326
Female	0.528	0.522	0.521	0.492	0.502	0.596	0.506
Black	0.211	0.311	0.244	0.341	0.170	0.080	0.114
Latino/a	0.264	0.150	0.166	0.119	0.183	0.160	0.199
Other race-ethnicity	0.074	0.080	0.029	0.032	0.037	0.086	0.012
Free or reduced-price lunch	0.854	0.864	0.916	0.770	0.784	0.616	0.831
English language learner	0.169	0.073	0.107	0.103	0.106	0.027	0.086
Special education	0.081	0.089	0.119	0.087	0.080	0.060	0.040
Nonstructural move to receive or stop receiving voucher	0.657	0.644	—	0.944	—	1.000	—
Baseline math score	−0.208	−0.378	−0.322	−0.399	−0.155	−0.168	−0.154
	[0.880]	[0.825]	[0.921]	[0.959]	[0.888]	[0.815]	[0.802]
Baseline ELA score	−0.201	−0.325	−0.323	−0.354	−0.009	0.144	0.005
	[0.876]	[0.876]	[0.905]	[0.922]	[0.950]	[0.854]	[0.787]
Year 1 post-baseline math	−0.353	−0.538	−0.323	−0.662	−0.135	−0.221	−0.104
	[0.863]	[0.793]	[0.928]	[0.965]	[0.860]	[0.756]	[0.765]
Year 1 post-baseline ELA	−0.201	−0.325	−0.311	−0.436	−0.013	0.097	0.041
	[0.826]	[0.849]	[0.902]	[0.936]	[0.898]	[0.797]	[0.761]

Source: Authors' compilation based on administrative data provided by the Indiana Department of Education.

Note: Table displays matched students with at least three years of test scores. Proportions reported for demographic characteristics; means and standard deviations [in brackets] reported for test scores in baseline year. For voucher students, this only includes students with at least two years of test scores prior to receiving a voucher for the first time. For public and private school comparison students, only students of the same sex and race/ethnicity in the same grade, school, and baseline year of voucher students are included.

Students who were previously enrolled in private schools without a voucher, received a voucher, and later switched to public schools are somewhat more likely to be African American and less likely to be Latino/a than are students who always attended private schools with or without a voucher. They have lower achievement at baseline and in the first post-baseline year.

Students whose families "game" the system—moving them from a private to a public school (to become eligible for a voucher) and then back to a private school after receiving a voucher—are more likely to be female, less likely to be ELL, and less likely to be black than students who have always attended a private school with a voucher. These private to public to private movers have lower average math but higher ELA achievement than students who always attended a private school at baseline and in the first year after receiving a voucher.

Regression Model Results

We next describe the change from baseline in a student's math (table 2) and ELA (table 3) achievement gains after receiving a voucher across the four voucher pathways. Students who made an initial switch into a private school after receiving a voucher experienced an average math achievement loss from baseline of −0.146 SD ($p \leq .001$) the first year after receiving a voucher. This decline in math achievement increases to −0.163 SD ($p \leq .001$) in the second year from baseline and is similar in year three (−0.156 SD, $p \leq .001$).

Meanwhile, students who initially were enrolled in a public school, later received a voucher to attend a private school, and then eventually left to attend a public school experienced greater average losses in math in year one from baseline (−0.192 SD, $p \leq .001$), year two (−0.197 SD, $p \leq .001$), and year three (−0.333 SD, $p \leq .01$). These students' losses accumulate to one-third of a SD loss in math in year three from baseline before this group of students returned to a public school. We found that students who switched from private (with a voucher) to public schools and students who remained always enrolled in a private school after receiving a voucher experienced no significant gains or losses from baseline in math.

We observed a different pattern between the groups of voucher students when looking at students' average ELA achievement gains from baseline. Most of the estimates in table 3 are statistically insignificant and small in magnitude, indicating no average ELA achievement differences from baseline for most of the student groups. Students who initially switched from a public into a private school after receiving a voucher, students who received a voucher while in a private school then switched to a public school, and students who remained always enrolled in a private school after receiving a voucher did not differ in the ELA achievement compared with the comparison groups. Students who switched from a public to a private school with a voucher and then returned to a public school experienced a statistically significant average achievement loss from baseline to year two in ELA (−0.130 SD, $p \leq .01$), which increased in magnitude to −0.161 SD ($p \leq .05$) in year three from baseline.

Beyond the main estimates of receiving a voucher in tables 2 and 3, we disaggregated the findings based on both whether students made structural (versus nonstructural) switches between schools and the year from baseline, as well as the grade level at baseline immediately before students first received a voucher. This set of results helps us better understand the relationship between the type and timing of student mobility and voucher-related achievement changes. Due to small cell sizes for structural movers, these results cannot be estimated for students who made private (with voucher) to public switches.[19]

In math, for students who switched from public to private schools with a voucher, the average achievement loss from baseline was similar one and two years from baseline for students who made nonstructural moves (−.152 SD, $p \leq .001$ in year 1; −0.160, $p \leq .001$ in year 2) compared to students who made structural moves (−0.133 SD, $p \leq .001$ in year 1; −0.168, $p \leq .001$ in year 2). However, students who made structural

19. We also report disaggregated results by baseline grade level only for voucher students always enrolled in private schools, as nearly all students do not change schools (that is, make a structural or nonstructural move).

Table 2. Changes from Baseline in Student Math Achievement for Voucher Recipients

	Public to Private with Voucher	Public to Private with Voucher to Public	Private with Voucher to Public	Always Private with Voucher
Pre-baseline achievement level	0.101***	−0.017	—	—
	(0.017)	(0.035)		
Baseline achievement level	0.121***	−0.036	−0.025	0.178**
	(0.016)	(0.032)	(0.232)	(0.062)
Voucher year 1	−0.146***	−0.192***	0.025	0.015
	(0.012)	(0.026)	(0.098)	(0.038)
Voucher year 2	−0.163***	−0.197***	−0.112	−0.018
	(0.015)	(0.044)	(0.232)	(0.057)
Voucher year 3	−0.156***	−0.333***	—	−0.086
	(0.022)	(0.104)		(0.102)
Voucher year 1*structural	−0.133***	−0.163***	—	—
	(0.020)	(0.043)		
Voucher year 2*structural	−0.168***	−0.306***	—	—
	(0.023)	(0.072)		
Voucher Year 3*structural	−0.226***[(s)]	−0.467**	—	—
	(0.034)	(0.182)		
Voucher year 1*nonstructural	−0.152***	−0.208***	—	—
	(0.014)	(0.032)		
Voucher year 2*nonstructural	−0.160***	−0.151***	—	—
	(0.020)	(0.054)		
Voucher year 3*nonstructural	−0.108***[(s)]	−0.272*	—	—
	(0.027)	(0.122)		
Voucher year 1*grade four baseline	−0.193***[(s)]	−0.248***	0.115	0.036
	(0.023)	(0.040)	(0.226)	(0.074)
Voucher year 1*grade five baseline	−0.132***[(s)]	−0.157***	0.005	−0.171[(s)]
	(0.021)	(0.040)	(0.140)	(0.091)
Voucher year 1*grade six baseline	−0.109***[(s)]	−0.130*	−0.133	0.095[(s)]
	(0.024)	(0.060)	(0.155)	(0.085)
Voucher year 1*grade seven baseline	−0.129***	—	—	0.059[(s)]
	(0.025)			(0.063)

Source: Authors' compilation based on administrative data provided by the Indiana Department of Education.

Note: Overall results reported from preferred models 1 and 2. To parse out subgroup effects, we include interactions between mobility indicators or grade indicators, and the voucher indicator. We computed separate models by number of years after receiving a voucher and between students starting in public schools versus private schools. ISTEP+ Math scores measured in standard deviation units, relative to the Indiana statewide mean and standard deviation within each grade and year (across all public and private school test takers). Robust standard errors, adjusted for the clustering of students within schools, are in parentheses. [(s)]Indicates significant differences between subgroups from F-test of estimates ($p \leq .050$). Missing cells indicate fewer than fifteen students or inability to estimate subgroup differences (no structural or nonstructural school transitions for always private voucher students).

*$p \leq .05$; **$p \leq .01$; ***$p \leq .001$

Table 3. Changes from Baseline in Student ELA Achievement for Voucher Recipients

	Public to Private with Voucher	Public to Private with Voucher to Public	Private with Voucher to Public	Always Private with Voucher
Pre-baseline achievement level	0.095***	0.009	—	—
	(0.016)	(0.034)		
Baseline achievement level	0.109***	0.010	−0.086	0.060
	(0.016)	(0.035)	(0.257)	(0.062)
Voucher year 1	0.005	−0.026	0.178	0.084
	(0.011)	(0.024)	(0.135)	(0.047)
Voucher year 2	−0.011	−0.130**	−0.105	0.115
	(0.015)	(0.043)	(0.265)	(0.064)
Voucher year 3	0.021	−0.161*	—	0.118
	(0.020)	(0.081)		(0.102)
Voucher year 1*structural	0.032	−0.018	—	—
	(0.020)	(0.037)		
Voucher year 2*structural	0.006	−0.038	—	—
	(0.023)	(0.064)		
Voucher year 3*structural	−0.015	−0.200	—	—
	(0.032)	(0.137)		
Voucher year 1*nonstructural	−0.008	−0.030	—	—
	(0.014)	(0.031)		
Voucher year 2*nonstructural	−0.023	−0.170**	—	—
	(0.020)	(0.054)		
Voucher year 3*nonstructural	0.046	−0.143	—	—
	(0.026)	(0.100)		
Voucher year 1*grade 4 baseline	−0.057**(s)	−0.036	0.245	−0.075(s)
	(0.021)	(0.036)	(0.244)	(0.101)
Voucher year 1*grade 5 baseline	−0.030(s)	−0.037	0.419	−0.061(s)
	(0.022)	(0.038)	(0.253)	(0.105)
Voucher year 1*grade 6 baseline	0.087***(s)	0.012	−0.124	0.228**(s)
	(0.023)	(0.057)	(0.275)	(0.085)
Voucher year 1*grade 7 baseline	0.057*(s)	—	—	0.211**(s)
	(0.024)			(0.079)

Source: Authors' compilation based on administrative data provided by the Indiana Department of Education.

Note: Overall results reported from preferred models 1 and 2. To parse out subgroup effects, we include interactions between mobility indicators or grade indicators, and the voucher indicator. We computed separate models by number of years after receiving a voucher and between students starting in public schools versus private schools. ISTEP+ ELA scores measured in standard deviation units, relative to the Indiana statewide mean and standard deviation within each grade and year (across all public and private school test takers). Robust standard errors, adjusted for the clustering of students within schools, are in parentheses. (s)Indicates significant differences between subgroups from F-test of estimates ($p \leq .050$). Missing cells indicate l fewer than fifteen students or inability to estimate subgroup differences (for example, no structural/non-structural school transitions for always private voucher students).

*$p \leq .05$; **$p \leq .01$; ***$p \leq .001$

moves experienced even greater math losses three years from baseline (−0.226 SD, $p \leq .001$) than students who made nonstructural moves (−0.108 SD, $p \leq .001$). We tested the statistical significance of the difference between these groups at the 5 percent level using an F-test and denoted any significant differences with an (s) in table 2.

When examining the grade level at which students received a voucher and switched from a public to a private school, we generally found greater math losses in the first post-baseline year for students transitioning at lower than higher grade levels. For example, table 2 shows that students who first received a voucher and switched from a public to a private school immediately following fourth grade experienced an average math loss of −0.193 SD ($p \leq .001$), which was a statistically greater loss than for students who made a similar move immediately following fifth (−0.132 SD, $p \leq .001$) or sixth grade (−0.109 SD, $p \leq .001$). We also see a similar pattern in ELA, whereby students who transition from public to private schools after fourth (−0.057 SD, $p \leq .01$) and fifth grade (−0.030 SD, not significant) have statistically lower ELA achievement in the first year post-baseline than their older peers. Students experienced gains in ELA achievement when transitioning after sixth (0.087 SD, $p \leq .001$) and seventh grade (0.057 SD, $p \leq .05$).

For students who transition from public to private schools and then return to a public school, we did not find statistically different impacts between students who initially make structural moves and their peers who make nonstructural moves. Similarly, we did not observe differences in first-year impacts by the grade level of transition. We also did not observe any grade-level heterogeneity in the first-year impacts of receiving a voucher for students who were initially enrolled in a private school with a voucher before exiting to attend a public school.

We observed meaningful grade-level differences in first-year impacts of receiving a voucher for students who were always enrolled in a private school. Students who received a voucher for the first time in sixth (0.095 SD, not significant) and seventh grade (0.059 SD, not significant) have statistically higher achievement in math than their peers who received a voucher for the first time after fifth grade (−0.171 SD, not significant). In ELA, we see a similar pattern, whereby students who first received a voucher in later grades experienced a statistically significant achievement gain (after sixth grade, 0.228 SD, $p \leq .01$; after seventh grade, 0.211 SD, $p \leq .01$) relative to their peers in lower grades who experienced null gains (after fourth grade, −0.075 SD, not significant; after fifth grade, −0.061 SD, not significant).

In a separate set of analyses, displayed in table 4, we examined the relationship between achievement and students who moved from private to public schools and then returned to a private school after receiving a voucher. In their interim year in a public school, this group of students experienced an average gain from their baseline score, measured when previously enrolled in a private school, in ELA (0.299 SD, $p \leq .001$) and no change in math achievement from baseline. After returning to a private school (controlling for their performance in the public school year), this group of voucher students experienced no change in achievement in either subject. The results look similar to the group of voucher students consistently enrolled in private schools. Variation by the grade level in which these switches occurred is minimal, except for students who completed their transition back to private schools between seventh and eighth grade, who experienced a statistically significant loss in math (−0.210 SD, $p \leq .01$) This loss was also statistically different from their peers who transitioned back to private schools between fifth and sixth grade (0.018 SD, not significant).

DISCUSSION

In this article, we analyze the pathways students can take in the Indiana voucher program, including voucher students who switched from public to private schools, from private to public schools, from private to public to private schools, or always attended a private school. Students who always attend private schools experience no differences in achievement gains before or after receiving a voucher. This is unsurprising because these students do not change school context and therefore have consistent experience in a private school environ-

Table 4. Impacts of Receiving Voucher for Students Moving from Private to Public to Private Schools with a Voucher

	Math	ELA
Baseline achievement level	−0.087	−0.112
	(0.071)	(0.074)
Year in public school	0.104	0.299***
	(0.072)	(0.075)
Voucher year 1	−0.072	−0.002
	(0.045)	(0.049)
Voucher year 2	−0.015	−0.066
	(0.063)	(0.069)
Voucher year 3	−0.047	0.113
	(0.087)	(0.124)
Voucher year 1*grade 4 baseline	0.018[(s)]	−0.120
	(0.079)	(0.086)
Voucher year 1*grade 5 baseline	−0.029	0.065
	(0.073)	(0.095)
Voucher year 1*grade 6 baseline	−0.210**[(s)]	0.054
	(0.079)	(0.071)

Source: Authors' compilation based on administrative data provided by the Indiana Department of Education.

Note: Overall results reported from preferred model 2, exclusive to voucher students moving from private to public and back to private schools. We computed separate models for the year a student spent in a public school and for each of the years after receiving a voucher. Baseline grade reported for year enrolled in private school before transition to public school. ISTEP+ math and ELA scores measured in standard deviation units, relative to the Indiana statewide mean and standard deviation within each grade and year (across all public and private school test takers). Robust standard errors, adjusted for the clustering of students within schools, are in parentheses. [(s)]Indicates significant differences between subgroups from F-test of estimates ($p \leq .050$).

*$p \leq .05$; **$p \leq .01$; ***$p \leq .001$

ment. However, it also indicates that students may not experience a stigma associated with beginning to receive a voucher, or at least that any potential stigma experienced does not influence their academic performance.

The negative findings in math for students who switch from a public to a private school with a voucher mirror what we have found using other estimation strategies (Berends and Waddington 2018; Waddington and Berends 2018). Here, we are able to compare the losses of these students with those of other groups of voucher students. It appears that students who have no experience in private schools experience a larger decline in their math scores than their counterparts enrolled in private schools for at least part of their schooling. These losses are largest for students who transition in earlier grades. Similarly, transitions from public to private schools with a voucher after fourth grade are associated with achievement loss in ELA, but transitions in the middle school years are associated with gains. The pattern of more negative outcomes following earlier transitions is one that additional research should examine (Langenkamp and Carbonaro 2018). In our future research, we also will examine further how these achievement losses vary by school context, such as racial-ethnic and socioeconomic composition and school climate.

Students who start in a public school, transition to a private school with a voucher, and then subsequently transition back to public school generally experienced larger achievement losses than students who start in a public school, transition to a private school with a

voucher, and remain in a private school with a voucher. These students are a group that our future research will examine to test two hypotheses: perhaps the low-performing students are being pushed out of private schools, or perhaps their families are motivated to pull them out because their child may not be performing well in private schools (as evidenced by some of the subgroup outcomes for nonstructural movers). These hypotheses are important to examine because some opponents of school choice argue that if private schools retain the ability to expel students for disciplinary or academic reasons, they should not receive public funds (Ravitch 2010).

Families who gamed the system early in the voucher program and moved their children from a private to a public and back to a private school to become eligible for a voucher may be a more motivated group. Their potential motivation may help explain why these students' achievement losses are more mitigated compared with other groups of voucher students.

Most voucher programs are designed to provide new learning opportunities, particularly for low-income students. Because of changes in the ICSP over time, students have taken different pathways to receiving a voucher. Our results here reveal variation in the association between these pathways and academic achievement scores. Moving from public to private schools with a voucher results in decreasing test scores, at least in math and especially for students who transition in earlier grades. Students who always attended private schools have higher math achievement at baseline experience no changes in their achievement before and after receiving a voucher. The achievement losses for public to private movers may be cause for concern for policymakers advocating for voucher programs. However, how students come to receive a voucher is as important to consider as the impact of receiving a voucher per se.

Although the results presented here are not causal estimates, they point to some potential considerations for parents considering using a voucher to move their child from a public to a private school and for policymakers considering a new or revised voucher program. Parents should be aware that research on transitions in general suggests that students often experience a decline in achievement in the first year following a transition to a new school of any type. Our findings suggest that attending a private voucher school generally does not compensate for that achievement loss. In addition, parents considering the timing of such a transition should note that transitions at earlier grades are associated with larger achievement losses, but that both structural and nonstructural transitions are associated with achievement loss in math for students who start in a public school and transition to a private school with a voucher, regardless of whether they ultimately transition back to a public school or remain in a private school.

The association between voucher transitions and achievement changes in ELA is weaker; achievement losses in ELA are concentrated in year two post-baseline following both structural and nonstructural changes and following a move after fourth grade. Students who transition to a private school with a voucher following sixth or seventh grade in a public school experienced average achievement gains.

Policymakers considering voucher policies should consider how voucher eligibility criteria shape the timing of students' school transitions and the pathways through which they make those transitions. Qualitative research in Indiana suggests that elementary and middle school students have more trouble adjusting to school expectations when they transition from a public to a private school in older grades (Austin 2019). However, recent quantitative analyses of the Washington, D.C., voucher program are consistent with our findings (Dynarski et al. 2017). Analyses of administrative data like those presented here will continue to inform policymakers as they weigh these and other important considerations for voucher policy design.

In addition, Indiana's K–12 administrative records can now be linked with data from the Indiana Commission on Higher Education and the National Student Clearinghouse through a statewide longitudinal data system. Future research will be able to explore the relationships between receiving a voucher and high school graduation, college readiness, and educational attainment. To date, forty-seven states have received at least one federal grant to support

building their statewide longitudinal data system, the information from which will provide social scientists with valuable data to assess program impacts, various student and adult outcomes, and educational inequality more generally. As the articles in this issue highlight, access to and use of such data can be burdensome, frustrating, and time consuming. Establishing researcher-practitioner partnerships not only promotes the use of research findings to inform practices, programs, and policies but also builds strong relationships with agency staff who know the data well and may help address some of the challenges of using administrative data. Addressing such challenges is key as data systems expand by adding data from state agencies, such as workforce data, child services, criminal justice and corrections. Such developments will enable researchers to connect individuals' schooling experiences to a variety of adult outcomes that will inform research and policy—making societal contributions far into the future.

REFERENCES

Abdulkadiroglu, Atila, Parag A. Pathak, and Christopher R. Walters. 2018. "Free to Choose: Can School Choice Reduce Student Achievement?" *American Economic Journal: Applied Economics* 10(1): 175–206.

Alexander, Karl L., Doris R. Entwisle, and Susan L. Dauber. 1996. "Children in Motion: School Transfers and Elementary School Performance." *Journal of Education Research* 90(1): 3–12.

Alspaugh, John W., and Roger D. Harting. 1995. "Transition Effects of School Grade-Level Organization on Student Achievement." *Journal of Research and Development in Education* 28(3): 145–49.

Anderson, Sara. 2017. "School Mobility Among Middle School Students: When and for Whom Does It Matter?" *Psychology in the Schools* 54(5): 487–503.

Angrist, Joshua D., Parag A. Pathak, and Christopher R. Walters. 2013. "Explaining Charter School Effectiveness." *American Economic Journal: Applied Economics* 5(1): 1–27.

Ashenfelter, Orley. 1978. "Estimating the Effect of Training Programs on Earnings." *Review of Economics and Statistics* 60(1): 47–57.

Austin, Megan J. 2019. "Organizational and Social Costs of Participating in a Voucher Program." In *School Choice at the Crossroads,* edited by Mark Berends, R. Joseph Waddington, and J. Schoenig. New York: Routledge.

Austin, Megan J., and Mark Berends. 2018. "School Choice and Learning Opportunities." In *Handbook of the Sociology of Education in the 21st Century*, edited by Barbara Schneider and Guan K. Saw. New York: Springer.

Benner, Aprile D. 2011. "The Transition to High School: Current Knowledge, Future Directions." *Educational Psychology Review* 23(3): 299–328.

Berends, Mark. 2018. "The Evolving Choice Landscape in the U.S." In *Handbook of Education Policy*, edited by Rosemary Papa and Shadow Armfield. Hoboken, N.J.: Wiley-Blackwell.

Berends, Mark, and Megan Austin. 2017. "The Promises and Pitfalls of Research-Practice Partnerships." In *Toward a More Perfect Psychology: Improving Trust, Accuracy, and Transparency in Research*, edited by Matthew C. Makel and Jonathan Plucker. Washington, D.C.: American Psychological Association.

Berends, Mark, Marissa Cannata, and Ellen B. Goldring, eds. 2011. *School Choice and School Improvement*. Cambridge, Mass.: Harvard Education Press.

Berends, Mark, and R. Joseph Waddington. 2018. "School Choice in Indianapolis: Effects of Charter, Magnet, Private, and Traditional Public Schools." *Education Finance and Policy* 13(2): 227–55.

Bifulco, Robert. 2012. "Can Nonexperimental Estimates Replicate Estimates Based on Random Assignment in Evaluations of School Choice? A Within-Study Comparison." *Journal of Policy Analysis and Management* 31(3): 729–51.

Bryk, Anthony S., Valerie E. Lee, and Peter B. Holland. 1993. *Catholic Schools and the Common Good.* Cambridge, Mass.: Harvard University Press.

Coleman, James S., and Thomas Hoffer. 1987. *Public and Private High Schools: The Impact of Communities*. New York: Basic Books.

Cook, Thomas D., William R. Shadish, and Vivian C. Wong. 2008. "Three Conditions Under Which Experiments and Observational Studies Produce Comparable Causal Estimates: New Findings from Within-Study Comparisons." *Journal of Policy Analysis and Management* 27(4): 724–50.

Dobbie, Will, and Roland G. Fryer. 2013. "Getting Be-

neath the Veil of Effective Schools: Evidence from New York City." *American Economic Journal: Applied Economics* 5(1): 28–60.

———. 2017. "Charter Schools and Labor Market Outcomes." NBER working paper no. 22502. Cambridge, Mass.: National Bureau of Economic Research.

Dynarski, Mark, Ning Rui, Ann Webber, Babette Gutmann, and Meredith Bachman. 2017. *Evaluation of the DC Opportunity Scholarship Program: Impacts After One Year*. NCEE 2017–4022. Washington: U.S. Department of Education. Eccles, Jacquelynne S., Sarah Lord, and Carol Midgley. 1991. "What Are We Doing to Early Adolescents? The Impact of Educational Contexts on Early Adolescents." *American Journal of Education* 99(4): 521–42.

EdChoice. 2018. "School Choice in America." Last modified June 26, 2018. Accessed September 18, 2018. https://www.edchoice.org/school-choice/school-choice-in-america.

Epple, Dennis, Richard E. Romano, and Miguel Urquiola. 2017. "School Vouchers: A Survey of the Economics Literature." *Journal of Economic Literature* 55(2): 441–92.

Figlio, David F., and Krzysztof Karbownik. 2016. *Evaluation of Ohio's EdChoice Scholarship Program: Selection, Competition, and Performance Effects.* Columbus, Ohio: Fordham Institute.

Fortson, Kenneth, Philip Gleason, Emma Kopa, and Natalya Verbitsky-Savitz. 2014. "Horseshoes, Hand Grenades, and Treatment Effects? Reassessing Whether Nonexperimental Estimators Are Biased." *Economics of Education Review* 44 (February): 100–13.

Grigg, Jeffrey. 2012. "School Enrollment Changes and Student Achievement Growth: A Case Study in Educational Disruption and Continuity." *Sociology of Education* 85(4): 388–404.

Gruman, Diana H., Tracy W. Harachi, Robert D. Abbott, Richard F. Catalano, and Charles B. Fleming. 2008. "Longitudinal Effects of Student Mobility on Three Dimensions of Elementary School Engagement." *Child Development* 79(6): 1833–52.

Hallinan, Maureen T., ed. 2006. *School Sector and Student Outcomes.* Notre Dame, Ind.: University of Notre Dame Press.

Hanushek, Eric A., John F. Kain, and Steven G. Rivkin. 2004. "Disruption versus Tiebout Improvement: The Costs and Benefits of Switching Schools." *Journal of Public Economics* 88(9–10): 1721–46.

Herbers, Janette E., Arthur J. Reynolds, and Chin-Chih Chen. 2013. "School Mobility and Development Outcomes in Young Adulthood." *Development and Psychopathology* 25(2): 501–15.

Hoffer, Thomas. 2009. "Perspectives on Private Schools." In *Handbook of Research on School Choice*, edited by Mark Berends, Matthew G. Springer, Dale Ballou, and Herbert J. Walberg. New York: Routledge.

Indiana Department of Education. 2011. *2011–2012 ISTEP+ Program Manual: Policies and Procedures for Indiana's Assessment System*. Indianapolis: Indiana Department of Education.

———. 2018. *Choice Scholarship Program Annual Report: Participation and Payment Data*. Indianapolis: Office of School Finance.

Langenkamp, Amy. 2009. "Following Different Pathways: Social Integration, Achievement, and the Transition to High School." *American Journal of Education* 116(1): 69–97.

———. 2010. "Academic Vulnerability and Resilience During the Transition to High School: The Role of Social Relationships and District Context." *Sociology of Education* 83(1): 1–19.

———. 2011. "Effects of Educational Transitions on Students' Academic Trajectory: A Life Course Perspective." *Sociological Perspectives* 54(4): 497–520.

Langenkamp, Amy, and William Carbonaro. 2018. "How School Socioeconomic Status Affects Achievement Growth Across School Transitions in Early Educational Careers." *Sociology of Education* 91(4): 358–78.

Lareau, Annette, and Kimberly Goyette. 2014. *Choosing Homes, Choosing Schools.* New York: Russell Sage Foundation.

Mehana, Majida, and Arthur J. Reynolds. 2004. "School Mobility and Achievement: A Meta-Analysis." *Children and Youth Services Review* 26(1): 93–119.

Mills, Jonathan N., and Patrick J. Wolf. 2017. "Vouchers in the Bayou: The Effects of the Louisiana Scholarship Program on Student Achievement After 2 Years." *Educational Evaluation and Policy Analysis* 39(3): 464–84.

Ravitch, Diane. 2010. *The Death and Life of the Great American School System: How Testing and Choice Are Undermining Education*. New York: Basic Books.

Reardon, Sean F. 2019. "Educational Opportunity in Early and Middle Childhood: Using Full Population Administrative Data to Study Variation by Place and Age." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(2): 40–68. DOI: 10.7758/RSF.2019.5.2.03.

Rosenbaum, Paul R., and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70(1): 41–55.

Rumberger, Russell W. 2015. *Student Mobility: Causes, Consequences, and Solutions*. Boulder, Colo.: National Education Policy Center.

Schiller, Kathryn S. 1999. "Effects of Feeder Patterns on Students' Transition to High School." *Sociology of Education* 72(4): 216–33.

Schwartz, Amy E., Leanna Stiefel, and Sarah A. Cordes. 2017. "Moving Matters: The Causal Effect of Moving Schools on Student Performance." *Education Finance and Policy* 12(4): 419–46.

Shakeel, M. Danish, Kaitlin P. Anderson, and Patrick J. Wolf. 2016. "The Participant Effects of Private School Vouchers Across the Globe: A Meta-Analytic and Systematic Review." EDRE working paper no. 2016–07. Fayetteville: University of Arkansas.

Waddington, R. Joseph, and Mark Berends. 2018. "Impact of Indiana Choice Scholarship Program: Achievement Effects for Students in Upper Elementary and Middle School." *Journal of Policy Analysis and Management* 37(4): 783–808.

Xu, Zeyu, Jane Hannaway, and Stephanie D'Souza. 2009. *Student Transience in North Carolina: The Effect of School Mobility on Student Outcomes Using Longitudinal Data.* Washington, D.C.: National Center for Analysis of Longitudinal Data in Education Research.

Zimmer, Ron W., and Eric P. Bettinger. 2015. "Beyond the Rhetoric: Surveying the Evidence on Vouchers and Tax Credits." In *Handbook of Research in Education Finance and Policy,* edited by Helen F. Ladd and Margaret E. Goertz. New York: Routledge.

Civil Society Goes to School: Parent-Teacher Associations and the Equality of Educational Opportunity

BRITTANY MURRAY, THURSTON DOMINA, LINDA RENZULLI, AND REBECCA BOYLAN

Where do parent-teacher associations (PTAs) and other similar school-linked nonprofits form? What role do PTAs play in distributing educational opportunities between and within public schools? In this paper, we link IRS data describing nonprofit organizations associated with North Carolina public schools to school- and student-level administrative data in order to answer these questions. Our analyses suggest PTAs form in a wide variety of school contexts, but high-revenue PTAs form primarily in affluent, predominantly white schools. Students in schools with active PTAs enjoy relatively strong achievement growth compared to their peers in schools without active PTAs. However, our analyses suggest that in reading, the benefits associated with PTAs flow disproportionately to nonpoor students.

Keywords: administrative data, civil society, PTAs, social capital, opportunity hoarding

As economic inequality increases (compare Piketty and Saez 2014; Saez and Zucman 2016), intergenerational mobility declines (Chetty et al. 2017), and the link between family income and student academic achievement strengthens (Duncan and Murnane 2011; Reardon 2011), scholars and public intellectuals are increasingly interested in the ways in which affluent parents secure advantage for their children (Calarco 2014; Chua 2011; Currid-Halkett 2017;

Brittany Murray is a doctoral student in the School of Education at the University of North Carolina at Chapel Hill. **Thurston Domina** is associate professor of educational policy and sociology in the School of Education at the University of North Carolina at Chapel Hill. **Linda Renzulli** is professor of sociology and department head in the College of Liberal Arts Department of Sociology at Purdue University. **Rebecca Boylan** is a postdoctoral researcher in the College of Liberal Arts Department of Sociology at Purdue University.

 Murray, Brittany, Thurston Domina, Linda Renzulli, and Rebecca Boylan. 2019. "Civil Society Goes to School: Parent-Teacher Associations and the Equality of Educational Opportunity." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(3): 41–63. DOI: 10.7758/RSF.2019.5.3.03. We thank Cole D. Smith for his support in matching nonprofit and school administrative data. We also thank William L. Pryor IV for generating maps of PTAs across North Carolina's elementary schools. This material is based in part on work supported by the National Science Foundation under grant numbers SES-1626891. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Direct correspondence to: Brittany Murray at bcmurray@live.unc.edu, 122 Peabody Hall, CB# 3500, Chapel Hill, NC 27599; Thurston Domina at tdomina@email.unc.edu, 119 Peabody Hall, CB# 3500, Chapel Hill, NC 27599; Linda Renzulli at lrenzull@purdue.edu, 700 W. State St., West Lafayette, IN 47907; and Rebecca Boylan at rboylan@purdue.edu, 700 W. State St., West Lafayette, IN 47907.

Horvat, Weininger, and Lareau 2003; Lareau 2000; Lewis and Diamond 2015; Putnam 2016; Reeves 2017). Several scholars have identified parent-teacher associations (PTAs) as one such mechanism (Cucchiara 2013; McKenna 2016; Posey-Maddox 2014). PTAs are parent-led organizations that attempt to create structures for parents to communicate collectively with school leaders and avenues for parents to contribute time, money, and energy into their children's schools. They are thus strategic sites for investigating the links between families and schools and their implications for educational inequality.

From one perspective, PTAs are *dream hoarders*, institutions that help facilitate the unequal distribution of educational opportunities both within and across schools (Reeves 2017). PTAs are increasingly effective fundraisers, and many PTAs in affluent communities raise hundreds of thousands of dollars annually (Brown, Sargrad, and Benner 2017; Nelson and Gazley 2014). These discretionary dollars support school programs, supplementary curricular materials, equipment and facilities, and even instructional staff. PTAs may thus create meaningful advantages for the children they serve. At the same time, they may potentially undermine efforts to equalize educational opportunities by exacerbating the disadvantages that poor students and resource-strapped schools face. Furthermore, PTAs may influence school instructional patterns in ways that reinforce existing inequalities within schools (Lewis and Diamond 2015; Posey-Maddox 2014).

But a different view of PTAs is available. In his 2001 book, *Bowling Alone*, the political scientist Robert Putnam argues that PTAs and similar voluntary associations serve as essential building blocks of democratic governance. Putnam and other scholars in this Tocquevillian tradition see PTAs as sources of social capital, a resource that is available and beneficial to rich and poor alike. From this perspective, PTAs create social spaces in which diverse school constituents can share views and work together to solve self-identified communal challenges (Noguera 2001). By building trust among parents, students, and educators, PTAs can help schools coordinate their efforts to create effective learning opportunities for all students (Bryk 2010; Lee, Bryk, and Smith 1993). Further, by facilitating communication among parents and educators, PTAs may draw attention to struggling students and create resources to help these students thrive (Coleman 1988).

In this article, we provide new evidence about the characteristics of schools in which PTAs do and do not operate and the relationship between PTA operation and student achievement. We acknowledge that PTAs and other voluntary organizations can simultaneously be forces for the maintenance of social inequality as well as important sources of social capital. Nonetheless, the social reproductionist and social capital views of the PTA differ in their expectations about the various school communities in which PTAs operate and the effects of PTAs on the distribution of educational opportunities within these schools. The reproductionist view suggests that PTAs operate primarily in affluent communities, and the social capital view expects them to operate in a wide range of communities. Likewise, the reproductionist view expects the effects of PTAs to accrue primarily to the children of the affluent, and the social capital view expects PTAs to benefit students broadly.

The tension between these two views of the PTA and parental collective action in school thus implies the following questions:

> How do school demographics, organizational characteristics, and neighborhood characteristics associate with PTA operation? To what extent do changes in these organizational and spatial factors predict PTA formation or dissolution?
>
> Do students in schools with active PTAs experience faster achievement growth than demographically similar students in schools with no active PTA? To what extent do poor and nonpoor students benefit equally from attending schools with active PTAs?

We assemble a unique collection of data from multiple administrative data sources. We use nonprofits' Internal Revenue Service (IRS) filings to identify the universe of parent-teacher associations, parent-teacher organizations, parent-student-teacher associations,

school booster clubs, and other nonprofits directly affiliated with North Carolina public elementary schools between 1999 and 2015. These records, collected not for research but for tax oversight purposes, provide an unprecedented view of parent-led organizational activity in contemporary schools. We link these records to school-level data from the National Center for Education Statistics' (NCES) Common Core of Data (CCD) geocoded demographic estimates from the American Community Survey and the decennial census. Pooling these data creates a comprehensive panel of North Carolina elementary schools during the 1999 to 2015 period, including time-varying measures of PTA activity. We further link these data to student-level administrative data from the North Carolina Education Research Data Center (NCERDC) to document PTA activity in the schools students attend as they age through elementary school.

PARENT-TEACHER ASSOCIATIONS IN AMERICAN PUBLIC ELEMENTARY SCHOOLS

Although not all PTAs are associated with the National PTA, that umbrella organization continues to create a model for parent collective action in contemporary public schools. Founded in 1897 as the National Congress of Mothers, the National Parent Teacher Association provides an important avenue to political activity, particularly for white, middle-class mothers (Woyshner 2009). The National Parent Teacher Association was central to the implementation of school lunch and child immunization programs and continues to organize letter-writing and other lobbying efforts around public school funding and related issues (Putnam 2001; Woyshner 2003). National PTA membership has declined since the 1960s and in many schools, unaffiliated parent-teacher organizations or parent-student-and-teacher organizations now operate instead (Crawford and Levitt 1999; Putnam 2001). These non-PTA affiliated organizations emerge less frequently and may provide fewer opportunities to engage parents in issues that transcend their children's schools, given their preference for organizational independence. However, because the internal operations of PTA and non-PTA groups do not appear to vary systematically, we consider these organizations as functional equivalents. Therefore, we consider all such school-based nonprofit organizations, regardless of national affiliation, as PTAs.

Modest in terms of organizational structure and human resources, PTAs recruit members from across the school. Nonetheless, our field research indicates that a relatively small group of parents typically play a disproportionately large role in PTA operations. Much PTA labor occurs outside of open school-wide meetings: on email chains and social media, in closed meetings of elected PTA board members, or in smaller committees on which parents volunteer to organize around particular tasks or interests, such as school grounds and facilities, fundraising, or newsletter production. Fundraising is also a central activity for many PTAs. These organizations collect membership dues and parent and community donations and run bake sales, raffles, and school festivals (Murray et al. 2018). Despite these efforts, PTAs account for less than 1 percent of school budgets, and even in the most extreme cases of highly funded PTAs in poorly financed schools, PTAs rarely account for more than 5 percent of an overall operating budget (Brown, Sargrad, and Benner 2017; Hanushek 1997; Hanushek, Rivkin, and Taylor 1996; Nelson and Gazley 2014). Our observations of parental collective action in several North Carolina elementary schools indicate that, despite this limited financial impact, by bringing parents into schools, PTAs provide a setting for parents to informally track their children's educational progress, get to know one another, and learn about the institution's day-to-day operations (Murray et al. 2018).

The literature provides little evidence about PTAs and their operations. From an equity perspective, our first research question asks which school communities benefit from PTA resources. Our second question focuses within schools, asking which students benefit from the educational opportunities that PTAs help create. Research and theory in the social capital and social reproductionist traditions suggest two sets of answers to these two questions. In the following discussion, we elucidate and test a series of hypotheses emerging from the two

perspectives about PTAs and their consequences for educational inequality.

Social Capital: Trust, Diversity, and the PTA

From a social capital perspective, PTAs are part of a virtuous cycle, institutionalizing the trust present in schools to create new social resources widely available to all members of a school community. Several school reform models thus emphasize PTAs as a mechanism to improve communication among parents, teachers, and administrators (Bryk 2010; Comer 1995; Epstein 2001). Although this perspective suggests that the formation of a PTA has a positive effect on school operation (Bryk and Schneider 2002; Bryk et al. 2010; Orr and Rogers 2011; Putnam 2001), it also implies that PTAs are unlikely to emerge in schools characterized by a sense of distrust or devoid of social norms (Coleman 1988). This perspective thus regards the PTA—and other voluntary civic organizations—as both a reflection of the degree of social capital present in a community and as a broker of additional social ties.

It also suggests that school and community characteristics associated with the development of communal trust should increase PTA activity in a school. Although the relationship between diversity and social capital is far from resolved (Portes and Vickstrom 2011), organizational size (Andrews 2017), racial and ethnic diversity, as well as economic inequality are in many settings associated with lower levels of trust (Alesina and La Ferrara 2002) and civic engagement (Costa and Kahn 2003; van der Meer and Tolsma 2014). We hypothesize that PTAs operate disproportionately in relatively small, well-established, and demographically homogeneous schools. Similarly, charter schools and other schools of choice should have relatively high levels of PTA operation, given that families who have chosen to enroll in such institutions likely subscribe to similar educational philosophies.

PTAs institutionalize the social ties among parents, creating opportunities for parents to get to know and work with other parents from across a school community (Small 2009). These social networks create intergenerational closure in the school environment, making it possible for parents to share information about issues within the school and look out for one another's children (Muller and Kerbow 1993; McNeal 1999; Noguera 2001). Further, given the loosely coupled organizational structure of most American public schools, involved and well-informed parents may provide an important channel for communication among teachers and administrators, building trust and organizational capacity in schools (on structure, Hallett 2010; Meyer 1977; Paino 2018; on communication, Lee, Bryk, and Smith 1993). Social capital theory suggests that these resources are distinctive—and distinctively powerful—because they operate as a public good (Coleman 1988). That is, in contrast to physical capital, where the benefits of ownership typically accrue primarily to the owner, the benefits of social capital accrue broadly throughout the social organizations in which social capital operates (Coleman 1988). We thus hypothesize that students who attend a school with an active PTA will enjoy stronger academic gains, on average, than children who attend a school with no PTA. Further, based on the assumption that the network in which PTA social capital operates is roughly coterminous with the school, we hypothesize that the achievement gains associated with attending a school with a PTA will be distributed broadly across a school's population.

Social Reproduction: Resource Hoarding via the PTA

For PTAs to have the broadly egalitarian consequences that school reform advocates hope for, they must be at least as likely to operate in relatively disadvantaged schools as in advantaged schools and must create opportunities that are equitably distributed within schools. Research in the social reproductionist mold suggests that the opposite may be true.

Affluent and middle-class citizens often have time, money, and professional skills to dedicate to political activity. It is not surprising, therefore, that they tend to be more engaged in the political process than their socioeconomically disadvantaged fellow citizens (Verba, Schlozman, and Brady 1995). Consistent with this evidence, Ashlyn Nelson and Beth Gazley demonstrate that the financial resources associated with PTAs and other school-

supporting institutions tend to cluster in large and relatively affluent school districts (2014). This study sheds important light on the distribution of PTA-generated funds but provides no evidence about the schools in which lower-revenue PTAs operate. Consistent with a view of PTAs as complicit in the reproduction of inequality, we hypothesize that high-revenue PTAs will be disproportionately concentrated in schools with small proportions of poor and minority students.

Although school reform models rely on PTAs to generate broadly accessible social capital, the assumption does not necessarily follow that the social networks in which PTAs generate and deploy social capital encompass entire schools. Case study evidence suggests that within a given school district, PTAs may play a particularly important role in schools that educate socioeconomically diverse student populations. In these settings, relatively advantaged parents use the PTA to attract other relatively advantaged families to the school and ensure that their students receive preferential treatment within the school (Cucchiara and Horvat 2009; Lareau and Muñoz 2012; Posey-Maddox 2014; Lewis and Diamond 2015). Based on this work, we hypothesize from a social reproduction perspective that PTAs operate disproportionately in schools that enroll relatively affluent students as well as in schools that are socioeconomically and racially diverse.

Schools, like any other social context, can be cliquish or otherwise socially fragmented (McFarland et al. 2014). In socially fragmented schools, one group of parents may act collectively to form and operate a PTA and use the organization and the social capital it creates to disproportionately benefit their own children. When they do, PTAs can be a tool for the hoarding of educational opportunities rather than creators of broad-based public goods. The sociologist L'Heureux Lewis-McCoy provides a striking example of the ways in which the PTA can facilitate opportunity hoarding in his study of racial inequality in one diverse suburban school district (2014). White middle-class parents in this district organized via the PTA to counter what they viewed as an undue focus on racial inequality, arguing instead for an achievement-for-all focus in district policy and creating networks for parents to lobby for preferred teachers and educational opportunities. In this case, the PTA seems to have redirected educator priorities away from efforts that might have disproportionately benefited students of color and other educationally disadvantaged students. In other settings, PTAs may create socioeconomically exclusive social networks through which advantaged parents share information about, and advocate for, enhanced educational opportunities for their children, a process described as "negotiated advantage" (Lareau and Calarco 2012; Calarco 2018). In either case, based on the social reproduction perspective, we hypothesize that the benefits associated with PTAs accrue to the relatively advantaged children with close network ties to the PTA and its leaders, at the expense of students in other social networks.

DATA AND METHODS

Our analyses provide a new window into PTA activity over time and space by exploring the schools in which IRS-reporting PTAs operate and the students that PTA operation benefit. Although these data cannot speak to the rate at which individual parents participate in the life of their children's schools via the PTA, they document the existence of PTAs and provide evidence of when and where relatively large PTAs (as measured by PTA revenues) operate. By pairing these administrative data with student-level achievement scores, we shed light on the relationship between PTAs and the distribution of student opportunities between and within schools.

Drawing on the Urban Institute's National Center for Charitable Statistics' (NCCS) databases, we construct a panel dataset that attempts to describe all PTAs, or similar organizations connected to a North Carolina public elementary school, that registered with the IRS between 1999 and 2015. We link these nonprofit data with school-level data on all North Carolina public elementary schools from the CCD, as well as student-level data from NCERDC. The result is a sixteen-year panel on school and civic life that allows us to investigate the distribution of PTAs across North Carolina elementary schools and their consequences for educational opportunities for children in the state.

Documenting Active PTAs

We draw on the NCCS databases that regularly document and compile financial data on organizations that register with the IRS. Under federal law, all 501(c)(3) public charities, with the exception of religious congregations, that raise $5,000 or more in annual revenue must register with the IRS. We identify educational nonprofits by extracting PTAs, PTOs, and single-school supporting nonprofits from the database and subsequently match them to North Carolina elementary schools using common addresses between the NCCS organizational and NCES school administrative files. Once matched, we create school-year level indicators for whether the school has an organization active in a given year based on whether NCCS reported an active organization in the same year or within a window of three years. The appendix provides a detailed description of the techniques we used to identify school-supporting organizations and match these organizations with the schools they serve.

Although these data may only provide a record of school-linked associations that file with the IRS, our analyses suggest nearly all active school-linked organizations regularly do so. The National PTA broadly circulates the necessary forms for associations to file with the IRS and provides considerable encouragement and assistance in doing so. Furthermore, our cross-validation efforts, including telephone interviews with a sample of schools and web searches, indicate that fewer than 5 percent of schools for whom no matched IRS data are available have links to a parent-teacher association with scheduled meetings.

Dependent Variables

We use these data to construct a time-varying indicator identifying each of the years in which a PTA or similar organization is active in each North Carolina public elementary school. In supplementary models, we investigate the school-level predictors of high-revenue PTAs by replacing the outcome in previous models with an indicator flagging schools during the years in which their PTAs reported revenues of $50,000 or more. Subsequent analyses of student-level outcomes in schools with and without PTAs rely upon data from the NCERDC for all youth enrolled in North Carolina public elementary schools during the 2007–2008, 2008–2009, 2009–2010, 2010–2011, and 2011–2012 school years. We linked these measures to school-level PTA data described for the corresponding school years.

Independent Variables

Our first set of analyses examines the characteristics of schools in which PTAs operate. To facilitate these analyses, we merge time-varying measures of PTA activity with panel data describing the demographics, organizational characteristics, and neighborhood characteristics of schools to explore the relationships between several measures of school trust and resources and PTA operation. Our analyses shed light on the sorts of resources necessary to build associational life in schools. Our second set of analyses measures variation in student achievement explained by PTA presence, net of students' race-ethnicity, gender, special education status, English-language learner status, and prior achievement. We specifically examine the role of socioeconomic status (SES) in moderating the relationship between attending a school with a PTA and subsequent academic benefits.

Racial-Ethnic Predictors of PTA Activity

We use data from the 1999–2015 Public Elementary and Secondary School Universe surveys in the CCD to capture racial-ethnic homogeneity as a measure of trust across school years. We constructed the Blau index (sometimes known as the Gini-Simpson index) to measure racial diversity in a school using total membership counts and the proportion of black, white, and Hispanic students by school year (Blau 1977). Blau index measures were generated using the following equation:

$$\lambda = 1 - \left(Pr^2_{white} + Pr^2_{black} + Pr^2_{hispanic} + Pr^2_{other}\right).$$

The resulting measure ranges from 0 to 1, where 0 represents a perfectly homogenous group and increases as the group becomes more diverse. The resulting proportion represents the probability that an individual in the school is from a certain racial-ethnic group.

Organizational Predictors of PTA Activity

We use CCD data to identify new schools as they appear in our panel, determine their size based on annual membership counts, and flag charter schools distinctively from traditional public schools. New schools in our panel are often public charters; North Carolina removed the cap on the number of charter schools allowed in the state in 2011, prompting an increase in the number of charters and proportion of students attending charters.

Socioeconomic and Spatial Predictors of PTA Activity

To measure school-level resources, we use counts of students receiving free and reduced-price lunch available in the CCD. We use the Education Demographic and Geographic Estimates (EDGE)—also prepared by NCES—to capture the socioeconomic context surrounding schools at the district level. EDGE provides data from the American Community Survey and the decennial census that measure local civilian unemployment rates and proportions of single-parent households among families with school-age children living within the school district boundaries. We use mapping software to estimate proportions of students receiving free and reduced-price lunch at the school attendance boundary level. Available for most traditional public schools, the School Attendance Boundary Survey (SABS), administered by NCES, provides shapefiles for use with measures in the CCD.

Student-level analyses rely on a measure from North Carolina's Department of Public Instruction termed economically disadvantaged. This classification captures students who receive, or are in a family that receives free or reduced-price lunch services or other federal assistance programs such as Temporary Assistance for Needy Families, or Supplemental Security Income.

Analytic Strategy

Our unbalanced panel includes 23,209 school-year observations from 1999 through 2015. Our analyses consider 1,631 unique schools that enrolled students in any configuration of kindergarten through fifth grades during at least one year of the study period. We then investigate the relationships between PTA exposure and 1.3 million reading and math achievement scores for fourth- and fifth-grade students from school year 2007–2008 to school year 2011–2012. The results from the analyses that follow provide insight into the prominence of PTAs in North Carolina, the characteristics of schools that predict PTA formation, and the consequences of PTA presence on student outcomes.

Research Question 1: Demographic and Organizational Predictors of PTA Activity

We use these school-level panel data to conduct a series of bivariate and multivariate analyses to investigate the school and community-level characteristics associated with PTA operation. We use random-effects models to examine differences between schools across the study period. This model takes the following functional form:

$$Ln\left(\frac{P_{it}}{1-P_{it}}\right) = \alpha + \beta_1 X_{it} + \sum \beta_2 Year_{it} + \mu_i + \varepsilon_{it},$$

where the outcome is the logit transformation of a school i's time-varying odds of having an active PTA in year t, X_{it} is vector of the school-level controls (demographics, size, and so on) described, $Year_{it}$ is a series of dummy variables, years 1999 through 2015 (1999 is reference), and μ_{it} and ε_{it} are the between and within school error terms. We also report the results of identically specified linear probability models, in which the dichotomous outcome is not log transformed. Such a model is necessarily heteroskedastic and therefore likely returns biased standard error estimates. However, this model is useful for interpretation when its results are consistent with the logistic model's results.

We then replace the school random effect in the model specified with a school fixed effect to investigate whether changes in school characteristics predict the formation or dissolution of PTAs. Although findings for these school fixed-effects models generalize only to the subset of schools in which a PTA formed or dissolved over the study period, they make it possible to separate the effects of school characteristics on PTA operations from the potential

confounding influences of a wide range of time-invariant aspects of school culture and location.

Research Question 2: PTA Activity and Student Achievement Growth

Our second set of analyses attempts to provide preliminary information about the link between PTA operation and the distribution of educational opportunities. We use the student-level data presented earlier to estimate models of the following general form:

$$Y_{is} = \alpha + \beta_1 PTA_s + \beta_2 EDS_{is} + \beta_3 X_{is} + \beta_4 Y_{is(Year\,n-1)} + \sum \beta_4 Year_{is} + \sum \beta_5 Grade_{is} + \mu_s + \varepsilon_{is}.$$

The dependent variables in these models, Y_{is}, are grade-by-year standardized scores on the standards-based end-of-grade tests that all North Carolina students take in the spring of the fourth and fifth grades. Because students take tests in both reading and mathematics each spring, we estimate all models separately for reading and mathematics skills. The key predictor in these models, PTA_s is an indicator flagging students enrolled in a school with an active PTA based on IRS filings during the school year under study.

This set of analyses includes a vector of time-varying student-level controls, including an indicator flagging students identified as economically disadvantaged; indicators identifying student gender, race-ethnicity, and English-language learner status; lagged values on students' math and reading achievement scores using third-grade scores for students observed in the fourth-grade and fourth-grade scores for students observed in the fifth grade; and a set of year and grade fixed effects to account for achievement time trends. We also include a school-by-year random effect to adjust standard errors for school-level clustering. In a second set of models, we add an interaction between the school-level PTA indicator and the student-level economically disadvantaged indicator to examine the extent to which poor and nonpoor students share in educational opportunities associated with PTA operation.

To interpret the results of these models as unbiased estimates of the independent effects of PTA operation on student achievement, one must assume that the measured controls capture all school and student characteristics associated both with PTA operation and student achievement. This assumption is highly restrictive, particularly given that PTAs likely both reflect and generate a wide range of relationships, resources, and practice within the schools where they operate. We thus consider these models to be exploratory. Nonetheless, this evidence regarding achievement growth trajectories in schools with and without PTAs—and particularly evidence regarding the development of socioeconomic achievement gaps in PTA and non-PTA schools—provides at least a partial test of social capital and opportunity hoarding hypotheses about the distribution of educational opportunities. Social capital implies a common and widely distributed positive association between PTAs and student achievement; opportunity hoarding implies that this association will be more pronounced for advantaged students than for poor students.

RESULTS

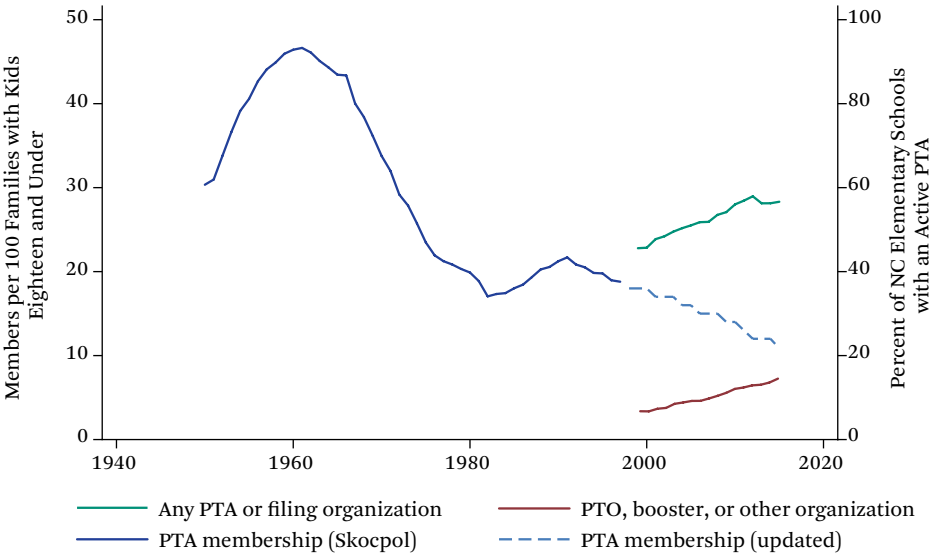
In what follows, we describe the frequency and permanence of parent associational life among North Carolina elementary schools. We then adjudicate between social capital and social reproductionist views of PTAs by providing evidence for the organizational, demographic, and spatial predictors of PTA operation and the academic consequences for students who attend schools with PTAs.

An Institution in Decline?

Robert Putnam, Theda Skocpol, and others view the PTA as emblematic of a broad decline in associational life in the United States (Putnam 2001; Skopol and Fiorina 2004). The solid line in figure 1, which reproduces Putnam's data on the proportion of parents of school-age children who were dues-paying members of the National PTA between 1960 and 2000, illustrates that the proportion of such U.S. parents peaked at 47 percent in 1960, before declining to 17 percent in 1980 (Putnam 2001). The dotted line updates Putnam's time series to 2016. In that year, the National PTA reported a membership of four million, or about 20 percent of U.S. parents of school-age children.

Our analyses suggest that, at least in North

Figure 1. Time Trend of PTA Membership and Organizational Presence



Source: Authors' calculations updating Theda Skocpol's data on PTA membership to 2017 and including organizational activity from the National Center for Charitable Statistics, Core and Business Master files.

Carolina, declining PTA membership in the recent past does not reflect a decline in the proportion of elementary school-age children who attend a school with an active PTA. Indeed, as the gray line in this figure illustrates, that proportion has risen steadily across the last fifteen years. Sixty percent of North Carolina public elementary school students currently attend a school with an active PTA or similar organization. Putnam speculates that some of the decline in PTA membership may be due to the rise of parent-teacher organizations, parent-teacher student organizations, and other school-level groups that are independent from the National PTA. Putnam worries that these organizations reflect increased attention on the interests of one's own children rather than a broader common good (2001). Our analysis indicates that under 20 percent of North Carolina school-affiliated parent organizations are independent from the National PTA, and that proportion has remained largely unchanged since 1999.

Another potential source for the decline in National PTA membership is a decline in the proportion of parents in schools with active PTAs who join the national organization. Our data cannot speak directly to this possibility, but survey data collected by the National Center for Education Statistics in 1995, 2003, 2007, and 2012 suggest that the vast majority of parents report active engagement in their children's schools and that this rate has not changed in the recent past. Nearly 90 percent of parents of school-age children reported having attended a PTA or related school-wide meeting in the past year throughout the 1995 to 2012 period, and nearly half of parents reported volunteering at their children's school in each of these study years (Herrold, O'Donnell, and Mulligan 2008; Noel et al. 2016; Wirt et al. 2001; Vaden-Kiernan, McManus, and Chapman 2005). On their face, these survey responses are difficult to reconcile with the evidence that Putnam has assembled regarding the decline in PTA membership. Nonetheless, it is possible that PTAs, although widely distributed across schools, represent an increasingly narrow membership within schools.

Table 1 provides a descriptive time trend for the proportion of schools with any PTA, and with big PTAs raising at least $50,000 annually relative to school and school district demographic characteristics (see table A1 for annually reported data). PTAs remain prominent

Table 1. Characteristics of North Carolina Elementary Schools

	1999–2002	2003–2006	2007–2010	2011–2014	2015
Dependent variables					
Any PTA?	0.47	0.51	0.54	0.57	0.57
More than 50K PTA?	0.04	0.06	0.07	0.04	—
Demographic predictors					
Free or reduced-price lunch	0.47	0.49	0.46	0.59	0.66
Black	0.33	0.33	0.29	0.25	0.26
White	0.59	0.55	0.53	0.51	0.50
Hispanic	0.05	0.08	0.11	0.15	0.16
Other	0.03	0.03	0.04	0.08	0.08
Racial-ethnic diversity	0.38	0.41	0.45	0.48	0.49
Organizational predictors					
Size	501	493	509	510	503
New school	0.02	0.02	0.02	0.01	0.02
New school (lagged)	0.01	0.02	0.02	0.01	0.00
Charter school	0.04	0.04	0.04	0.05	0.06
Neighborhood predictors					
Unemployment rate	0.12	0.12	0.11	0.11	0.11
Single-parent household rate	0.52	0.52	0.52	0.52	0.52

Source: Authors' calculations based on National Center for Charitable Statistics, Core and Business Master files, merged with school administrative data from National Center for Education Statistics, Common Core of Data. Neighborhood predictors are generated using district-level decennial census and ACS estimates from National Center for Education Statistics, Education Demographic and Geographic Estimates data.
Note: PTA revenues only available until year 2012. Racial-ethnic diversity calculated using the Blau index.

over the sixteen-year period, appearing at just under 50 percent of all elementary schools in the state in 1999 (N = 750) and almost 60 percent by 2015 (N = 979). Although PTAs grew by about 20 percent over the study period, large PTAs tripled their presence in the years in which we have revenue data (1999–2012), from about fifty schools raising $50,000 or more to almost 150.

Socioeconomic and Spatial Correlates of PTA Activity

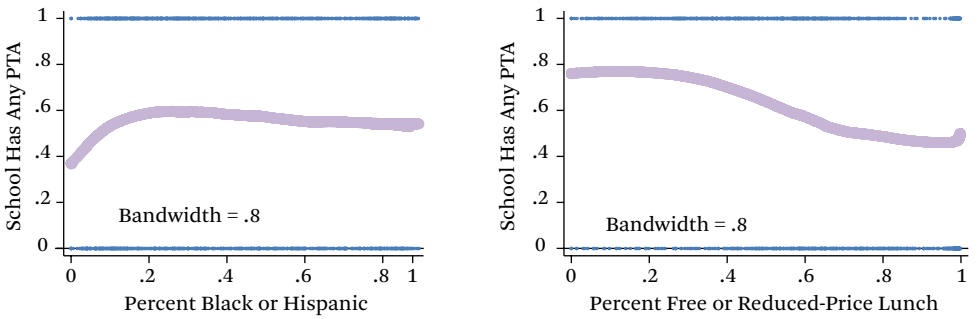
Table 2 gives an overview of school characteristics summarized by PTA history: whether it formed or dissolved and whether it always or never existed across the panel. PTAs are prominent and stable organizations across our panel: most North Carolina elementary schools had a PTA in all years of the study period (N = 667). There are relatively few cases of PTA births and deaths, indicating that once a PTA forms, it is likely to remain active. The relationship between school size, racial diversity, and the longevity of a PTA appears to be positive. Table 2 also suggests schools that host PTAs in all years are remarkably similar, on average, to those in which a PTA never existed. Racial and socioeconomic gaps between schools with and without PTAs are strikingly small. Although PTAs are evenly distributed across racial and socioeconomic indicators, the disparity among charter schools based on PTA history is noticeable. Charter schools make up 8 percent of schools that never had a PTA, but they are not represented among schools that had an organization in all years, in part because most North Carolina charters formed during the study period.

Table 2. School Characteristics by PTA History, 1999–2015

	Never Had a PTA	PTA Died	PTA Born	PTA All Years
Demographic predictors				
Free or reduced-price lunch	0.56	0.61	0.41	0.47
Black	0.28	0.38	0.15	0.32
White	0.55	0.45	0.72	0.51
Hispanic	0.10	0.11	0.08	0.11
Other	0.05	0.05	0.04	0.05
Racial-ethnic diversity	0.40	0.45	0.35	0.48
Organizational predictors				
Size	438	460	525	571
Charter school	0.08	0.02	0.08	0.00
Neighborhood predictors				
Unemployment rate	0.12	0.12	0.11	0.11
Single-parent household rate	0.50	0.53	0.49	0.54

Source: Authors' calculations based on National Center for Charitable Statistics, Core and Business Master files, merged with school administrative data from National Center for Education Statistics, Common Core of Data. Neighborhood predictors are generated using district-level decennial census and ACS estimates from National Center for Education Statistics, Education Demographic and Geographic Estimates data.
Note: Racial-ethnic diversity calculated using the Blau index.

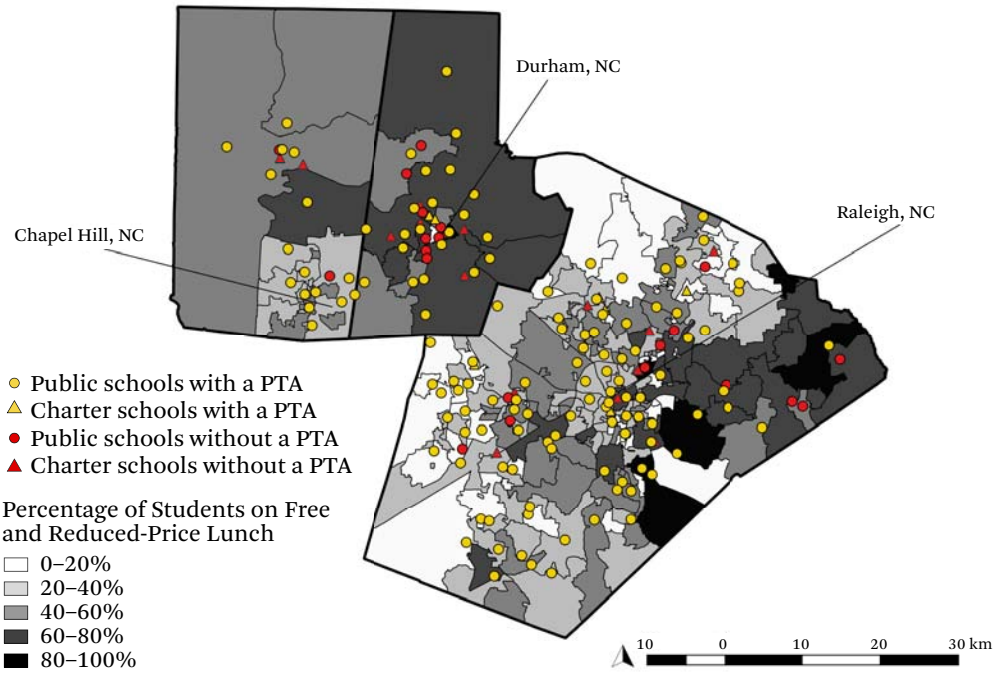
Figure 2. Bivariate Loess Plots of School Demographic Composition and PTA Presence



Source: Authors' calculations using the National Center for Charitable Statistics, Core and Business Master files, and the National Center for Education Statistics, Common Core of Data.

The bivariate loess plots in figure 2 provide another representation of the relationship between school demographic composition and PTA presence. The figure displays the proportion of students who qualify for free or reduced-price lunch and the proportion of students who are African American or Hispanic plotted with the likelihood that a school has a PTA. Although approximately 80 percent of the state's most homogeneously affluent schools have an active PTA, PTAs are also present in almost 50 percent of schools where nearly all students qualify for free or reduced-price lunch. That said, there is almost no bivariate relationship between school racial composition and PTA presence except among the most highly segregated black schools. More than 40 percent of schools with predominantly white enrollments and more than 50 percent of relatively integrated schools have operating PTAs.

Figure 3. Elementary School PTAs by Local Free Lunch Participation Rates, North Carolina Research Triangle, 2015



Source: Author-generated map using the National Center for Charitable Statistics, Core and Business Master files, with administrative data and shapefiles provided by the National Center for Education Statistics, Common Core of Data and School Attendance Boundary Survey.

However, as the proportion of black students approaches 1, the likelihood of having a PTA declines.

Figure 3 provides a geographic representation of the evenness by which PTAs are dispersed across poor and nonpoor communities in North Carolina's Triangle area. The referenced map displays traditional public and public charter elementary schools, both with and without PTAs, alongside the concentration of poor students in enrollment zones that make up Wake County Public School System, Durham Public Schools, Orange County Public Schools, and Chapel Hill-Carrboro City Schools. PTAs appear to be most prominent in school zones with the most economic diversity, although they still emerge in homogeneously poor and affluent communities.

The analyses reported in table 3 provide a multivariate examination of the school-level characteristics predicting PTA operation. In models 1 and 2, we use time-varying measures of school characteristics to predict PTA operation in the same year, adding school random effects to account for repeated observations. In models 3 and 4, we replicate these models using school fixed effects to estimate the extent to which changes in school characteristics affect changes in PTA operation, net of time-invariant school characteristics. We report linear random- and fixed-effects analyses in models 1 and 3 for ease of interpretation. Because the key predictors in these models are z-score standardized, each of the model's coefficients—with the exception of coefficients for indicator variables flagging new schools—can be interpreted as the difference in expected PTA activity rates associated with a 1 standard deviation change in the predictor, net of all other controls. We report logistic analyses in models 2 and 4.

Models 1 and 2 draw attention to several

Table 3. Predictors of PTA Presence Using Linear and Logistic Regression, 2015

	Model 1 Linear RE	Model 2 Logistic RE	Model 3 Linear FE	Model 4 Logistic FE
Demographic predictors				
Free or reduced-price lunch (std)	−0.01*	−0.11	0	−0.04
Black (std)	0.01	0.41***	0	0.03
Hispanic (std)	−0.03***	−0.63***	−0.03***	−0.75***
Racial-ethnic diversity (std)	0.06***	1.58***	0.05***	1.29***
Organizational predictors				
Size (std)	0.02***	0.43***	0.01	0.12
New school	−0.08***	−2.81***	−0.08***	−2.77***
New school (lagged)	−0.03*	−1.21***	−0.03**	−1.31***
Year fixed effects				
1999 (reference)	0	0	0	0
2000	0	0.07	0	0.08
2001	0.02**	0.65**	0.02**	0.62**
2002	0.03***	0.74***	0.03**	0.70***
2003	0.04***	0.97***	0.04***	0.92***
2004	0.04***	1.14***	0.04***	1.07***
2005	0.05***	1.22***	0.05***	1.17***
2006	0.05***	1.32***	0.05***	1.29***
2007	0.04***	0.93***	0.04***	0.91***
2008	0.05***	1.16***	0.05***	1.16***
2009	0.06***	1.38***	0.06***	1.38***
2010	0.07***	1.74***	0.07***	1.70***
2011	0.08***	1.82***	0.07***	1.75***
2012	0.08***	2.05***	0.08***	1.99***
2013	0.08***	1.78***	0.07***	1.72***
2014	0.07***	1.57***	0.07***	1.54***
2015	0.08***	1.77***	0.07***	1.72***
Constant	0.46***	−3.35***	0.48***	
N	23,018	23,018	23,018	5,248

Source: Authors' calculations based on National Center for Charitable Statistics, Core and Business Master files, merged with school administrative data from National Center for Education Statistics, Common Core of Data. Neighborhood predictors are generated using district-level decennial census and ACS estimates from National Center for Education Statistics, Education Demographic and Geographic Estimates data.

Note: These models use time-varying measures of school characteristics to predict PTA operation in the same year. Models 1 and 2 use school random effects and models 3 and 4 use school fixed effects. Models 1 and 3 are linear probability models and Models 2 and 4 are logistic models.

*$p < .05$; **$p < .01$; ***$p < .001$

school characteristics that associate with PTA operation. Large schools are significantly more likely to have active PTAs. Conversely, the negative conditional association between the proportion of students enrolled in free or reduced-price lunch in a school and the odds of active PTA operation is statistically significant. In both cases, these relationships are weak: a standard deviation increase in enrollment and decrease in percent free or reduced-price lunch is

associated with a 1 or 2 percentage point difference in PTA operation rates. Nonetheless, both associations point to the role that two basic organizational resources—the numbers of parents who are available to participate in PTA activities and these parents' finances—play in PTA operation. The significant negative coefficients for new schools in models 1 and 2 also likely speak to the role that organizational resources play in PTA operation.

The relationship between schools' racial composition and PTA operation is somewhat more challenging to explain. Although social capital theory implies PTAs should form in relatively homogeneous schools, models 1 and 2 in table 3 indicate that school racial diversity is significantly and positively associated with PTA operation. Indeed, this association is relatively strong: a 1 standard deviation change in school diversity is associated with a 6 percentage point difference in a school's likelihood of having an active PTA, net of controls. An increase in school size, a smaller concentration of economically disadvantaged students, and an increase in school racial diversity all independently relate to a school's likelihood of having an operating PTA. However, none of these associations provide warrant for strong statements regarding the school demographic factors that increase or decrease the chances of PTA operation in a given school.

The fixed-effects analyses reported in models 3 and 4 aim to provide a more stringent test of the role school characteristics play in determining PTA formation or dissolution. These analyses, which pool data for all available schools across our 1999–2015 panel and add a school-level fixed effect, estimate the relationship between year-over-year changes in each of the key independent variables and the chances of school PTA operation. These models thus focus particular attention on schools in which PTAs dissolve or form over the study period to control for all time-invariant school characteristics that may confound cross-sectional estimates of the effects of a given measured school factor on the chances of PTA operation. The results reported in models 3 and 4 of table 3 indicate that after controlling for unmeasured time-invariant school characteristics, the socioeconomic circumstances of a school's student body do not significantly influence the likelihood of having a PTA operate in a school. However, this analysis indicates when school enrollments become more racially diverse, the odds of PTA organization increase significantly. Although the magnitude of this effect is not large, it suggests that PTAs may come into being to serve as social bridges in relatively diverse school communities, or enable advantaged parents to control the flow of school opportunities and resources when in racially competitive school contexts. Finally, this model indicates that new schools are significantly less likely to have PTAs, particularly in their first year of operation.

Table 4 reports the results of a series of parallel models designed to isolate the characteristics of schools that host high-revenue PTAs. The first two random-effects models, which explore the characteristics of schools in which an active PTA raised at least $50,000, suggest unsurprisingly that school size is positively associated with high-revenue PTA operation. Further, this model suggests that schools that educate large proportions of students who qualify for free or reduced-price lunch, schools that educate large proportions of black and Hispanic students, and racially diverse schools are less likely to house high-revenue PTAs.

The analyses reported in models 3 and 4 suggest that some part of the association between school demographics and PTA revenue may be systematic. These fixed-effects models use year-to-year changes in school characteristics to predict formation and dissolution of PTAs raising $50,000 or more annually. These models reveal growing school enrollments may expand fundraising possibilities facilitating the creation of high-revenue PTAs. Although the model suggests that changes in proportions of free or reduced-price lunch enrollment have no significant effect on the likelihood of having a high-revenue PTA, we find that these resource-rich organizations are less likely to form in schools that educate minority youth. As the significant negative coefficients for percent black and percent Hispanic indicate, increases in black and Hispanic enrollments decrease the odds of high-revenue PTA operation.

Table 4. Predictors of PTA Presence (Over $50,000) Using Linear and Logistic Regression, 2015

	Model 1 Linear RE	Model 2 Logistic RE	Model 3 Linear FE	Model 4 Logistic FE
Demographic predictors				
Percent free or reduced-price lunch (std)	−0.01**	−0.62***	0	0.07
Percent black (std)	−0.04***	−1.65***	−0.06***	−1.77***
Percent Hispanic (std)	−0.03***	−0.96***	−0.03***	−0.98***
Racial-ethnic diversity (std)	−0.01**	0.29	−0.02***	−0.27
Organizational predictors				
Size (std)	0.06***	1.43***	0.06***	1.15***
New school	0.01	0.86*	0	0.39
New school (lagged)	0.03**	1.11**	0.02	0.81*
Neighborhood predictors				
Unemployment rate	−0.03***	−1.69***	0	0
Single-parent household rate	0.03***	0.64***	0	0
Year fixed effects				
1999 (reference)	0	0	0	0
2000	0.01	0.58	0.01	0.54
2001	0.02*	0.63	0.02**	0.63*
2002	0.02***	0.76*	0.03***	0.80*
2003	0.04***	1.19***	0.04***	1.33***
2004	0.04***	1.40***	0.04***	1.39***
2005	0.05***	1.66***	0.05***	1.65***
2006	0.05***	1.86***	0.06***	1.88***
2007	0.05***	1.75***	0.06***	1.90***
2008	0.05***	1.43***	0.05***	1.89***
2009	0.06***	1.79***	0.06***	2.17***
2010	0.06***	2.22***	0.07***	2.15***
2011	0.07***	2.47***	0.07***	2.41***
2012	0.08***	2.78***	0.08***	2.70***
Constant	0.02*	−9.67***	0.01	
N	23,018	23,018	23,018	5,248

Source: Authors' calculations based on National Center for Charitable Statistics, Core and Business Master files, merged with school administrative data from National Center for Education Statistics, Common Core of Data. Neighborhood predictors are generated using district-level decennial census and ACS estimates from National Center for Education Statistics, Education Demographic and Geographic Estimates data. Individual math and reading achievement scores from North Carolina Education Research Data Center.

Note: These models use time-varying measures of school characteristics to predict PTA operation in the same year. Models 1 and 2 use school random effects and models 3 and 4 use school fixed effects. Models 1 and 3 are linear probability models and Models 2 and 4 are logistic models.

*$p < .05$; **$p < .01$; ***$p < .001$

PTAs and the Distribution of Educational Opportunities

Having investigated the school-level correlates of PTA operation, our analysis now turns to a consideration of the relationship between PTAs and student learning opportunities. The analyses reported in table 5 investigate the extent to which achievement varies between students in schools with active PTAs and schools without one. We report results for mathematics achievement in panel A and reading achievement in panel B.

North Carolina students who attend schools with PTAs score an average of 0.16 standard deviations higher on fourth and fifth grade mathematics achievement tests than students in schools with no PTA. Observable student characteristics explain the vast majority of this achievement gap between PTA and non-PTA schools. Nonetheless, after controlling for prior achievement, race, ethnicity, economic disadvantage, and other potential confounders in model 2, we find that students in PTA schools score statistically significantly higher on end of grade mathematics tests than their peers in schools without PTAs. Net of a robust set of controls, including prior math and reading achievement, students in schools with active PTAs score 0.02 standard deviations higher on mathematics achievement than their peers in schools without active PTAs. This difference is substantively quite small, suggesting that active PTAs are associated with few independent educational advantages.

The advantages associated with attending a school with a PTA appear to accrue evenly for poor and nonpoor students alike. Model 2 reveals a 0.08 standard deviation gap in mathematics achievement growth between economically disadvantaged students and their peers who receive neither free or reduced-price lunch nor other targeted government assistance. Model 3 indicates that this class-based mathematics achievement gap does not vary significantly between schools with PTAs and schools without PTAs.

The results for reading tell a different story. As model 1 in panel B of table 5 indicates, reading achievement is 0.10 standard deviations higher in schools with active PTAs than in schools without. After controlling for student prior achievement and other student characteristics, this gap is statistically indistinguishable from zero. Interestingly, however, this null main effect obscures important variation between poor and nonpoor students in the relationship between PTA activity and reading achievement. After adding a PTA*economically disadvantaged student interaction in model 3, the significant positive main effect for PTA indicates that nonpoor students experience significantly larger year-to-year reading achievement gains when they attend schools with PTAs than poor students do. Economically disadvantaged students, by contrast, appear to experience no benefit in reading when they attend schools with active PTAs. Although this model provides no evidence to suggest that attending a school with a PTA is associated with lower reading scores for disadvantaged students, these findings are consistent with a view of PTAs as opportunity hoarding organizations, suggesting that the additional educational opportunities provided by PTAs, although small, flow exclusively to nonpoor students.

DISCUSSION

Our analyses represent a novel use of administrative data to understand the distribution and consequences of parental collective action in contemporary elementary schools.

Drawing on nonprofit tax records, we construct a comprehensive panel describing PTAs and other nonprofit organizations affiliated with North Carolina public schools between 1999 and 2015. By linking data on these charitable organizations with school and district-level administrative data as well as student-level achievement data, we track the predictors of PTA activity in schools and document the implications of PTA activity for the distribution of student learning opportunities.

The evidence displayed about where PTAs form provides mixed support for both the social capital and social reproductionist views of PTAs. Although membership in the National PTA has declined over time, we find that PTAs remain a prominent institution across North Carolina's elementary schools, broadly accessible to a large share of schoolchildren, regardless of socioeconomic backgrounds. PTAs are unlikely to form in new schools, implying a

Table 5. Predictors of Student-Level Math and Reading Achievement, 2015

	Panel A: Math			Panel B: Reading		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Any PTA?	0.157***	0.016***	0.017***	0.096***	0.002	0.006**
EDS		−0.081***			−0.029***	
Any PTA* EDS			−0.002			−0.008***
Grade fixed effects						
4	−0.122***	0.019***	−0.018***	−0.090***	−0.018***	0.135***
5	−0.116***	0.038***	0	−0.088***	0	0.153***
Year fixed effects						
2007	0			0		
2008	0	0	0	−0.001	0	0
2009	−0.005	−0.031***	−0.031***	0.005	−0.078***	−0.078***
2010	−0.006	−0.026***	−0.026***	0.001	−0.076***	−0.076***
2011	−0.011	−0.034***	−0.034***	0.004	−0.082***	−0.082***
2012	−0.008	−0.025***	−0.025***	0.003	−0.076***	−0.077***
2013	−0.002	0.124***	0.124***	−0.001	−0.082***	−0.082***
Student controls						
Prior read (standard)		0.196***			0.602***	
Prior math (standard)		0.624***			0.102***	
American Indian		−0.022**			−0.001	
Asian		0.107***			0.018***	
Black		−0.030***			0.007***	
Hispanic		0.053***			0.022***	
Multirace		−0.008*			0.009**	
Male		0.013***			−0.035***	
Moved school		−0.012***			−0.004*	
Limited English proficient		−0.005			−0.052***	
Constant	0	0	0.037***	0	0.155***	0
N	1,341,353	1,055,592	1,055,592	1,369,613	1,062,344	1,062,344

Source: Authors' calculations based on National Center for Charitable Statistics, Core and Business Master files, merged with school administrative data from National Center for Education Statistics, Common Core of Data. Neighborhood predictors are generated using district-level decennial census and ACS estimates from National Center for Education Statistics, Education Demographic and Geographic Estimates data. Individual math and reading achievement scores from North Carolina Education Research Data Center.

Note: These models use time-varying measures of student characteristics and parent-teacher association (PTA) operation to predict student math achievement in panel A and reading achievement in panel B. Models 2 and 5 add student controls. Models 3 and 6 add an interaction effect between a student's economically disadvantaged status (EDS) and whether the student attends a school with a PTA.

*p < .05; **p < .01; ***p < .001

required level of trust and organizational resources among school constituents before parents decide to build venues for civic engagement. However, consistent with the social reproductionist view, most trust variables are negatively correlated with PTA formation: PTAs tend to form in schools with large, racially diverse populations and are least likely to form in charter schools. Our findings regarding the location of high-revenue PTAs further align with a social reproductionist view. Although PTAs on average are distributed relatively evenly across various demographic contexts, PTAs raising at least $50,000 annually operate almost

exclusively in homogeneously white and affluent schools. This disparity in access to high-revenue PTAs supports a social reproductionist view given that poor and minority students do not have access to additional resources well-established PTAs provide.

Similarly, PTAs distribute benefits to students in a pattern consistent with both social capital and social reproductionist views. The social capital view suggests that PTA presence would benefit the entire student body. Consistent with this view, students who attend schools with active PTAs experience small but significantly higher year-to-year gains in mathematics than their demographically similar peers who attend schools that do not have an active PTA. We find no evidence to suggest that the mathematics gains associated with PTAs are restricted to any particular part of the school population. In reading, however, we find that the academic benefits associated with attending a PTA school accrue exclusively to nonpoor students. To be clear, our analyses indicate that economically disadvantaged students learn no less in reading when they attend schools with PTAs than they might in schools without PTAs. Nonetheless, our results highlight the possibility that PTAs work for an advantaged segment of the school population. Our analyses demonstrate that PTAs might be both a common good and a mechanism for opportunity hoarding among the advantaged.

The weak association between PTAs and student outcomes that we document may be unsurprising in light of literature documenting a weak relationship between school resources and student academic achievement (compare Hanushek 1997; Hanushek, Rivkin, and Taylor 1996). Further, these findings are also consistent with the results of a broad array of well-evaluated educational interventions that indicate that even the most promising interventions produce highly uneven effects when implemented in disparate educational settings. The growing literature on the uneven effects of federally funded school improvement grants is instructive (see, for example, Dee 2012; Heissel and Ladd 2018; Schueler, Goodman, and Deming 2017; Strunk et al. 2016).

As is the case in the research on school turnaround, further research is needed to investigate the role that PTAs and other organizations play in different school contexts, among various populations of students and in producing alternative outcomes. The relationship between attending a school with a PTA and student achievement scores might be moderated by school characteristics. For example, the relationship may be stronger in larger schools, urban school districts, in schools with fewer state and district expenditures, or in schools with PTAs grossing $50,000 or more annually. Further, this article investigates differences only among poor and nonpoor students. Given the propensity for PTAs to form in racially competitive environments, future analyses should investigate differential PTA effects among racial-ethnic groups. Furthermore, given the ability for parent involvement in collective organizations to affect student socialization into the fabric of school culture (Domina 2005; McNeal 1999), we might find stronger PTA effects for student behavioral and attendance outcomes.

We see this work as a first step toward a broader research agenda on parental collective action and social inequality. When, where, and why do PTAs benefit students broadly? When, where, and why do they work to reinforce social inequality? How do PTAs determine their (formal and informal) organizational agendas, and what voices are represented in that deliberative process? What consequences do PTAs have for the social organization of schools? Do they influence course assignments, gifted and special education placements? In light of the positive correlation between racial diversity and PTA operation, how is race implicated in PTA organizations and their effects? Furthermore, given our finding that high-resource PTAs are much more highly clustered in affluent schools than other PTAs, to what extent do the effects of PTAs vary with the size of their operational budgets?

Administrative data provide powerful tools for answering many of these questions. Although we document only the formation, dissolution, and revenues of elementary school-supporting associations in a single state, these data are available for a diverse array of nonprofit organizations across the country. Our approach to matching data and verifying match quality is labor intensive, and we are currently working to develop a more automated approach

that uses probabilistic matching techniques to link organizations to schools on a much larger scale using directory variables such as name, address, and geocodes (see Downing and Bruckner 2019; Goerge and Wiegand 2019). Ultimately, our goal is to provide data on PTAs and other school-linked nonprofit organizations for all U.S. public schools via a CCD link. These data will provide important indicators of school social capital and document resources that are currently unobservable in educational administrative data. Additionally, NCCS documents tax filings across a range of public sectors—including but not limited to health, education, human services, and the arts—providing a novel opportunity to investigate organizational resources in school-community contexts.

Our future analyses will take advantage of linked educational administrative data to develop rich measures of classroom assignment patterns across schools and over time (Domina et al. 2017) and explore how PTAs influence these aspects of students' educational experiences. We also plan to match these organizational-level data to the Stanford Education Data Archive to investigate the role of PTAs in producing academic achievement gaps nationwide. Other analyses will take advantage of policy shifts to explore the effects of desegregation efforts and school choice on PTA formation and effects.

APPENDIX: IDENTIFYING PTAS

Our panel uses data compiled in two NCCS databases: the IRS Business Master File (BMF), a monthly updated census of IRS-registered nonprofit associations in the United States based on the IRS Forms 1023 and 1024, and the NCCS Core Financial Files (Core files), an annually updated data source that includes more detailed financial information for larger nonprofit associations that file IRS Form 990, 990-EZ, or 990-PF.

The BMF contains descriptive information for all organizations that have filed IRS Forms 1023 or 1024 within the last three years, successfully registering for tax-exempt status. Although it contains information including the name and address of each organization, its employer identification number, and its ruling date, or the month and year it first filed for tax-exempt status, the BMF provides limited financial information for the organizations. Instead, the primary purpose of the data source is to record whether an organization is active at a given time.

Under the Pension Protection Act of 2006, all registered nonprofit organizations must file a tax return each year to maintain tax-exempt status. NCCS updates the BMF data source several times a year to capture new organizations that have recently filed for tax-exempt status and to exclude those whose tax-exempt status has expired. NCCS creates the Core files using IRS Return Transaction Files paired with the IRS BMF. These data sources provide a more complete picture of the financial viability for all tax-exempt organizations, except religiously affiliated ones, raising at least $25,000 annually. We use this information to learn about North Carolina PTAs that raise substantial amounts of money annually. Between 2005 and 2010, all organizations with revenues of greater than $25,000 file Forms 990 or 990N. In 2011–2012, this rule changed, and now only organizations with revenues greater than $50,000 are required to file those forms.

We used an extensive process to whittle the full array of nonprofit organizations down to traditional parent-teacher associations, parent-teacher organizations, arts- and sports-supporting boosters, and other single-school supporting organizations. Restricting the sample to educational nonprofits using only the National Taxonomy of Exempt Entities classification system codes still returned many results for school district-wide charities, private foundations, special interest groups, student- and family-serving nonprofits, in addition to organizations that support higher and alternative education. From here, we use a two-step process to further identify parent-teacher associations and similar school-linked associations. First, we use a series of search terms, including PTA, P.T.A., PTO, parent-teacher, school booster, and related variants, to identify nonprofit organizations that are likely closely connected to schools. Second, we match these associations to the schools around which they are organized using association and school names, addresses, and zip codes.

In approximately 90 percent of cases, these

matches are unproblematic—for example, the E. K. Powe PTA shares a name, a Durham (North Carolina) address, and zip code with E. K. Powe Elementary School in Durham. In the remaining cases, we identify matches by hand, making it possible to link associations and schools that share a name but report from neighboring zip codes; associations in which school names are shortened, abbreviated, or misspelled; or other similar cases. Using this process, we match parent-teacher associations or similar organizations to approximately 50 percent of North Carolina elementary schools open between 1999 and 2015.

We telephoned approximately one hundred elementary schools for which there appeared to be no connected parent-teacher association in the NCCS data. Ninety-one percent of these schools acknowledged having no PTA. These conversations add to our confidence that our data fairly accurately capture the prevalence of reasonably formal associations that are connected to—but organizationally independent of—North Carolina public elementary schools.

The dependent variables of interest are active status of PTAs overall, and at various levels of revenue. The term PTAs in this analysis includes both parent-teacher organizations, arts- and sports-booster clubs, as well as other single-school supporting parent organizations. This measure was created based on whether an organization was deemed to be alive in a given year, considering its year of formation and its consistency of showing up in the BMF. Organizations were given a three-year window as a grace period according to NCCS guidelines, but otherwise were deemed to be inactive after three years of absence from the BMF files. Similarly, researchers assessed the activity of larger PTAs ranked by revenue. Using the Core files, we ranked PTAs based on their revenue as any PTA active, active PTA raising at least $25,000 dollars annually, active PTA raising at least $50,000 dollars annually, and active PTA raising at least $100,000 dollars annually.

REFERENCES

Alesina, Alberto, and Eliana La Ferrara. 2002. "Who Trusts Others?" *Journal of Public Economics* 85(2): 207–34.

Andrews, Rhys. 2017. "Organizational Size and Social Capital in the Public Sector: Does Decentralization Matter?" *Review of Public Personnel Administration* 37(1): 40–58.

Blau, Peter M. 1977. *Inequality and Heterogeneity: A Primitive Theory of Social Structure*, vol. 7. New York: Free Press.

Brown, Catherine, Scott Sargrad, and Meg Benner. 2017. "Hidden Money: The Outsized Role of Parent Contributions in School Finance." Washington, D.C.: Center for American Progress. Accessed March 5, 2018. https://www.americanprogress.org/issues/education/reports/2017/04/08/428484/hidden-money.

Bryk, Anthony S. 2010. "Organizing Schools for Improvement." *Phi Delta Kappan* 91(7): 23–30.

Bryk, Anthony S., and Barbara Schneider. 2002. *Trust in Schools: A Core Resource for Improvement*. New York: Russell Sage Foundation.

Bryk, Anthony S., Penny Bender Sebring, Elaine Allensworth, John Q. Easton, and Stuart Luppescu. 2010. *Organizing Schools for Improvement: Lessons from Chicago*. Chicago: University of Chicago Press.

Calarco, Jessica M. 2014. "Coached for the Classroom: Parents' Cultural Transmission and Children's Reproduction of Educational Inequalities." *American Sociological Review* 79(5): 1015–37.

———. 2018. *Negotiating Opportunities: How the Middle Class Secures Advantages in School*. New York: Oxford University Press.

Chetty, Raj, John N. Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan. 2017. "Mobility Report Cards: The Role of Colleges in Intergenerational Mobility." NBER working paper no. 23618. Cambridge, Mass.: National Bureau of Economic Research.

Chua, Amy. 2011. *Battle Hymn of the Tiger Mother*. New York: Bloomsbury Publishing.

Coleman, J. S. 1988. "Social Capital in the Creation of Human Capital." *American Journal of Sociology* 94:S95–120.

Comer, James P. 1995. *School Power: Implications of an Intervention Project*. New York: Free Press.

Costa, Dora L., and Matthew E. Kahn. 2003. "Civic Engagement and Community Heterogeneity: An Economist's Perspective." *Perspectives on Politics* 1(1): 103–11.

Crawford, Susan, and Peggy Levitt. 1999. "Social

Change and Civic Engagement: The Case of the PTA." In *Civic Engagement in American Democracy*, edited by Theda Skocpol and Morris P. Fiorina. Washington, D.C.: Brookings Institution Press.

Cucchiara, Maia B. 2013. *Marketing Schools, Marketing Cities: Who Wins and Who Loses When Schools Become Urban Amenities.* Chicago: University of Chicago Press.

Cucchiara, Maia B., and Erin McNamara Horvat. 2009. "Perils and Promises: Middle-Class Parental Involvement in Urban Schools." *American Educational Research Journal* 46(4): 974–1004.

Currid-Halkett, Elizabeth. 2017. *The Sum of Small Things: A Theory of the Aspirational Class*. Princeton, N.J.: Princeton University Press.

Dee, Thomas. 2012. *School Turnarounds: Evidence from the 2009 Stimulus*. NBER working paper no. W17990. Cambridge, Mass.: National Bureau of Economic Research.

Domina, Thurston. 2005. "Leveling the Home Advantage: Assessing the Effectiveness of Parental Involvement in Elementary School." *Sociology of Education* 78(3): 233–49.

Domina, Thurston, Andrew McEachin, Paul Hanselman, Priyanka Agarwal, NaYoung Hwang, and Ryan Lewis. 2017. "Beyond Tracking and Detracking: The Dimensions of Organizational Differentiation in Schools." RAND working paper WR-1155. Santa Monica, Calif.: RAND Corporation.

Downing, Janelle, and Tim Bruckner. 2019. "Subprime Babies: The Foreclosure Crisis and Initial Health Endowments." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(2): 123–40. DOI: 10.7758/RSF.2019.5.2.07.

Duncan, Greg J., and Richard J. Murnane, eds. 2011. *Whither Opportunity?: Rising Inequality, Schools, and Children's Life Chances*. New York: Russell Sage Foundation.

Epstein, Joyce L. 2001. *School, Family, and Community Partnerships: Preparing Educators and Improving Schools.* Boulder, Colo.: Westview Press.

Goerge, Robert M., and Emily R. Wiegand. 2019. "Understanding Vulnerable Families in Multiple Service Systems." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(2): 86–104. DOI: 10.7758/RSF.2019.5.2.05.

Hallett, Tim. 2010. "The Myth Incarnate: Recoupling Processes, Turmoil, and Inhabited Institutions in an Urban Elementary School." *American Sociological Review* 75(1): 52–74.

Hanushek, Eric A. 1997. "Assessing the Effects of School Resources on Student Performance: An Update." *Educational Evaluation and Policy Analysis* 19(1): 141–64.

Hanushek, Eric A., Steven G. Rivkin, and Lori L. Taylor. 1996. "Aggregation and the Estimated Effects of School Resources." NBER working paper no. W5548. Cambridge, Mass.: National Bureau of Economic Research.

Heissel, Jennifer A., and Helen F. Ladd. 2018. "School Turnaround in North Carolina: A Regression Discontinuity Analysis." *Economics of Education Review* 62 (February): 302–20.

Herrold, Kathleen, Kevin O'Donnell, and Gail Mulligan. 2008. "Parent and Family Involvement in Education, 2006–07 School Year, from the National Household Education Surveys Program of 2007." NCES 2008–050. Washington: U.S. Department of Education.

Horvat, Erin McNamara, Elliot B. Weininger, and Annette Lareau. 2003. "From Social Ties to Social Capital: Class Differences in the Relations between Schools and Parent Networks." *American Educational Research Journal* 40(2): 319–51.

Lareau, Annette. 2000. *Home Advantage: Social Class and Parental Intervention in Elementary Education*. Lanham, Md.: Rowman & Littlefield Publishers.

Lareau, Annette, and Jessica M. Calarco. 2012. "Class, Cultural Capital, and Institutions: The Case of Families and Schools. In *Facing Social Class: How Societal Rank Influences Interaction*, edited by Susan T. Fiske and Hazel Rose Markus. New York: Russell Sage Foundation

Lareau, Annette, and Vanessa Lopes Muñoz. 2012. "'You're Not Going to Call the Shots': Structural Conflicts Between the Principal and the PTO at a Suburban Public Elementary School." *Sociology of Education* 85(3): 201–18.

Lee, Valerie E., Anthony S. Bryk, and Julia B. Smith. 1993. "The Organization of Effective Secondary Schools." *Review of Research in Education* 19:171–267.

Lewis, Amanda E., and John B. Diamond. 2015. *Despite the Best Intentions: How Racial Inequality Thrives in Good Schools.* New York: Oxford University Press.

Lewis-McCoy, L'Heureux. 2014. *Inequality in the*

Promised Land: Race, Resources, and Suburban Schooling. Stanford, Calif.: Stanford University Press.

McFarland, Daniel A., James Moody, David Diehl, Jeffrey A. Smith, and Reuben J. Thomas. 2014. "Network Ecology and Adolescent Social Structure." *American Sociological Review* 79(6): 1088–121.

McKenna, Laura. 2016. "How Rich Parents Can Exacerbate School Inequality." *The Atlantic*, January 28.

McNeal, Ralph B., Jr. 1999. "Parental Involvement as Social Capital: Differential Effectiveness on Science Achievement, Truancy, and Dropping Out." *Social Forces* 78(1): 117–44.

Meyer, John W. 1977. "Effects of Education as an Institution." *American Journal of Sociology* 83(1): 55–77.

Muller, Chandra, and David Kerbow. 1993. "Parent Involvement in the Home, School, and Community." In *Parents, Their Children, and Schools,* edited by Barbara Schneider and James Coleman. Boulder, Colo.: Westview Press.

Murray, Brittany, Thurston Domina, Amy Petts, and Linda Renzulli. 2018. "We're in This Together: Bridging and Bonding Social Capital in Elementary School PTOs." Unpublished working paper. Chapel Hill: University of North Carolina.

Nelson, Ashlyn Aiko, and Beth Gazley. 2014. "The Rise of School-Supporting Nonprofits." *Education Finance and Policy* 9(4): 541–66.

Noel, Amber, Patrick Stark, Jeremy Redford, and Andrew Zukerberg. 2016. "Parent and Family Involvement in Education, from the National Household Education Surveys Program of 2012." NCES 2013-028.rev23. Washington: U.S. Department of Education.

Noguera, Pedro A. 2001. "Transforming Urban Schools Through Investments in the Social Capital of Parents." *Psychology* 16: 725–50.

Orr, Marion, and John Rogers. 2011. *Public Engagement for Public Education: Joining Forces to Revitalize Democracy and Equalize Schools*. Stanford, Calif.: Stanford University Press.

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

Piketty, Thomas, and Emmanuel Saez. 2014. "Inequality in the Long Run." *Science* 344(6186): 838–43.

Portes, Alejandro, and Erik Vickstrom. 2011. "Diversity, Social Capital, and Cohesion." *Annual Review of Sociology* 37:461–79.

Posey-Maddox, Linn. 2014. *When Middle-Class Parents Choose Urban Schools: Class, Race, and the Challenge of Equity in Public Education.* Chicago: University of Chicago Press.

Putnam, Robert D. 2001. *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon and Schuster.

———. 2016. *Our Kids: The American Dream in Crisis*. New York: Simon and Schuster.

Reardon, Sean F. 2011. "The Widening Academic Achievement Gap Between the Rich and the Poor: New Evidence and Possible Explanations." In *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*, edited by Greg J. Duncan and Richard J. Murnane. New York: Russell Sage Foundation.

Reeves, Richard V. 2017. *Dream Hoarders: How the American Upper Middle Class Is Leaving Everyone Else in the Dust, Why That Is a Problem, and What to Do About It.* Washington, D.C.: Brookings Institution Press.

Saez, Emmanuel, and Gabriel Zucman. 2016. "Wealth Inequality in the United States Since 1913: Evidence from Capitalized Income Tax Data." *Quarterly Journal of Economics* 131(2): 519–78.

Schueler, Beth E., Joshua S. Goodman, and David J. Deming. 2017. "Can States Take Over and Turn Around School Districts? Evidence from Lawrence, Massachusetts." *Educational Evaluation and Policy Analysis* 39(2): 311–32.

Skocpol, Theda, and Morris P. Fiorina, eds. 2004. *Civic Engagement in American Democracy*. Washington, D.C.: Brookings Institution Press.

Small, Mario Luis. 2009. *Unanticipated Gains: Origins of Network Inequality in Everyday Life.* New York: Oxford University Press.

Strunk, Katharine O., Julie A. Marsh, Susan C. Bush-Mecenas, and Matthew R. Duque. 2016. "The Best Laid Plans: An Examination of School Plan Quality and Implementation in a School Improvement Initiative." *Educational Administration Quarterly* 52(2): 259–309.

Vaden-Kiernan, Nancy, John McManus, and Chris Chapman. 2005. "Parent and Family Involvement in Education: 2002–03." NCES 2005-043. Washington: U.S. Department of Education.

van der Meer, Tom, and Jochem Tolsma. 2014. "Ethnic Diversity and Its Effects on Social Cohesion." *Annual Review of Sociology* 40:459–78.

Verba, Sidney, Kay Lehman Schlozman, and Henry E. Brady. 1995. *Voice and Equality: Civic Voluntarism in American Politics*. Cambridge, Mass.: Harvard University Press.

Wirt, John, Susan Choy, Debra Gerald, Stephen Provasnik, Patrick Rooney, Satoshi Watanabe, Richard Tobin, and Mark Glander. 2001. "The Condition of Education 2001." NCES 2001-072. Washington: U.S. Department of Education.

Woyshner, Christine. 2003. "Race, Gender, and the Early PTA: Civic Engagement and Public Education, 1897–1924." *Teachers College Record* 105(3): 520–44.

———. 2009. *The National PTA, Race, and Civic Engagement, 1897–1970*. Columbus: The Ohio State University Press.

Are They Still Worth It? The Long-Run Earnings Benefits of an Associate Degree, Vocational Diploma or Certificate, and Some College

CHANGHWAN KIM AND CHRISTOPHER R. TAMBORINI

Sub-baccalaureate education accounts for most of the expansion in higher education over the last century. Nevertheless, relatively few studies have examined the related long-term financial benefits. Exploiting a rich dataset linking the Survey of Income and Program Participation and administrative earnings records, this study investigates these benefits over a person's early and mid-career and the heterogeneity of these patterns by field of study. We find substantial payoffs, net of an extensive set of demographic covariates and variables indicating high school courses taken. At the same time, we find considerable variation across degree types and fields of study. Several vocational diplomas, certificates, and associate degrees are associated with higher earnings than bachelor's degrees in social science, liberal arts, and education. Implications of these findings are discussed.

Keywords: long-term earnings, community college, associate degree, vocational certificate, field of study

The financial payoffs associated with human capital investments in postsecondary education have received considerable attention in scholarly and policy arenas in recent decades (Davies and Guppy 1997; Hout 2012; Kim, Tamborini, and Sakamoto 2015; Oreopoulos and Petronijevic 2013; Sandefur and Park 2007; Tamborini, Kim, and Sakamoto 2015). To date, researchers have largely focused on degrees from bachelor's and graduate programs when estimating the labor market benefits associated with postsecondary education. An appreciably

ChangHwan Kim is professor in the Department of Sociology at the University of Kansas. **Christopher R. Tamborini** is a senior researcher in the Office of Research, Evaluation, and Statistics at the U.S. Social Security Administration.

 Kim, ChangHwan, and Christopher R. Tamborini. 2019. "Are They Still Worth It? The Long-Run Earnings Benefits of an Associate Degree, Vocational Diploma or Certificate, and Some College." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(3): 64–85. DOI: 10.7758/RSF.2019.5.3.04. The views expressed in this article are those of the authors and do not represent the views of the Social Security Administration (SSA). Access to SSA data linked to Census Bureau survey data is subject to restrictions imposed by Title 13 of the U.S. Code. The data are accessible at a secured site and must undergo disclosure review before their release. For researchers with access to these data, our programs used in this analysis are available on request. Direct correspondence to: ChangHwan Kim at chkim@ku.edu, Department of Sociology, University of Kansas, 1415 Jayhawk Blvd., Room 716, Lawrence, KS 66045; and Christopher R. Tamborini at chris.tamborini@ssa.gov, Senior Researcher, Office of Research, Evaluation, and Statistics, U.S. Social Security Administration, 8th floor, 250 E. St., SW, Washington, D.C. 20254.

smaller body of work has focused on sub-baccalaureate programs, such as associate degrees and vocational certificates (Belfield and Bailey 2011; Dougherty 1987; Gill and Leigh 2003; Grubb 1993; Jepsen, Troske, and Coomes 2014; Marcotte et al. 2005; Monk-Turner 1990).

Despite receiving less attention in empirical work, sub-baccalaureate education has expanded substantially over recent decades. In 1963, 740,000 students were in public two-year institutions, accounting for 24 percent of public higher education (Baum, Kurose, and McPherson 2013). As of 2013, more than seven million students were, representing around 40 percent of all postsecondary students and around half of public college education (Kena et al. 2015). This growth raises questions about the financial benefits associated with sub-baccalaureate education over a person's work life and the heterogeneity of these patterns by field of study.

Systematic study of the financial payoffs associated with sub-baccalaureate education, particularly over the long run, has been hampered by a number of factors. One central factor is a lack of suitable data, particularly on fields of study. This is unfortunate given the vast diversity of programs and credentials at the sub-baccalaureate level, which are related with heterogeneous labor market outcomes (Grubb 1997; Jepsen, Troske, and Coomes 2014).

In this study, we investigate the long-term financial benefits associated with sub-baccalaureate degrees and certificates in the United States. To do so, we use nationally representative data that matches respondents from the 2004 and 2008 Survey of Income and Program Participation (SIPP) to restricted-use administrative files containing their longitudinal earnings from tax records. Specifically, the linked tax data allow us to follow the annual earnings of men and women finishing high school during the period between 1972 and 1993, for up to twenty years after graduation (on the advantage of the survey and administrative records linked data, see Grusky et al. 2019). In turn, SIPP's educational history module permits us to disaggregate between associate degree, vocational-technical diploma or certificate, and more important, by fields of study at the sub-baccalaureate level.

Our analysis explores three main lines of inquiry. The first examines how annual earnings unfold from early to mid-career across sub-baccalaureate groups using a twenty-year observational window starting from high school graduation. The second considers variation in the long-term benefits of sub-baccalaureate education across fields of study. Horizontal stratification by field of study is an important yet often overlooked dimension shaping the financial rewards to education (Kim, Tamborini, and Sakamoto 2015). A handful of studies reveal differential returns at the sub-baccalaureate level by field of study, but more work is needed (Dadgar and Trimble 2015; Grubb 1993, 1997; Kane and Rouse 1995; Liu, Belfield, and Trimble 2015). The third line of inquiry estimates the long-term financial benefits to sub-baccalaureate education by constructing a measure of twenty-year cumulative earnings starting from the year after high school graduation. Altogether, the long-term view presented in our analysis is critical, we argue, because standard accounts in the literature about the economic benefits of sub-baccalaureate education tend to rely on cross-sectional estimates or shorter-term observational windows.

Our results show that annual earnings, cumulative twenty-year earnings, and earnings growth of all sub-baccalaureate groups including vocational certificates and diploma are higher over early to mid-adulthood than comparable high school graduates. At the same time, we uncover important heterogeneity by fields of study at the sub-baccalaureate level. Some fields of study at the sub-baccalaureate level, we find, are associated with higher payoffs than others among bachelor's degree holders over the observational period. Although our aim is not to derive estimates of the causal effects of sub-baccalaureate education, we provide new insights into the shape of earnings trajectories from early to mid-career associated with different sub-baccalaureate groups and their fields of study by assembling novel data that links national survey data with longitudinal tax data. More generally, the article demonstrates one way in which survey data linked to administrative records can be used to better understand long-term economic patterns related to education over the life course.

BACKGROUND

The importance of education in determining the labor market experiences of workers has been widely discussed in economic and sociological research (Baum, Kurose, and McPherson 2013; Brand and Xie 2010; Hout 2012; Tamborini, Kim, and Sakamoto 2015). Despite a growing body of work on the financial returns to sub-baccalaureate education (Jepsen, Troske, and Coomes 2014), the bulk of the literature has focused on the financial rewards flowing from human capital investments in bachelor's degree programs (Rosenbaum, Ahearn, and Rosenbaum 2017). Most existing studies tend to calculate the financial rewards associated with the sub-baccalaureate level under the broad category of "some college" (Tamborini, Kim, and Sakamoto 2015).

A number of factors have complicated examination of the financial benefits related to sub-baccalaureate education. First, the institutions and programs offered at that level vary widely, and may include community colleges, vocational and technical institutes, and proprietary schools (Grubb 1993, 1997; Liu, Belfield, and Trimble 2015). Moreover, students attending such institutions are quite diverse in their goals and family backgrounds. Some attend with the intention of receiving a degree or certificate, whereas others may see a sub-baccalaureate degree as a stepping stone for a bachelor's degree program. Sub-baccalaureate institutions provide second chances for educational mobility, particularly among disadvantaged students. This diversity makes it challenging for researchers to estimate the associated benefits.

Second, as mentioned, the available data for sub-baccalaureate credentials are relatively limited (Belfield and Bailey 2011). Large national surveys such as the American Community Survey only provide cross-sectional information and do not contain detailed fields of study, especially at the sub-baccalaureate level. Panel surveys such as the National Longitudinal Survey of Youth include rich longitudinal information, but often do not have it on the field of study, and sample sizes are limited. Furthermore, these panel data have high sample attrition rates. Recently, administrative data on students enrolled in community colleges matched with unemployment insurance information have been used to study the long-term economic return to sub-baccalaureate education (Dadgar and Trimble 2015; Jepsen, Troske, and Coomes 2014; Liu, Belfield, and Trimble 2015; Stevens, Kurlaender, and Grosz 2018). However, these studies are limited to one or two states and compare only between those who earn an award in a community college and those who enroll in a community college but do not earn any certificate, diploma, or degree. The relative benefits to sub-baccalaureate education relative to other levels of education such as high school graduates or four-year college degree holders are not estimated.

Is the Financial Benefit to Sub-Baccalaureate Education Substantive?

These challenges aside, a number of key studies provide estimates of the labor market benefits associated with sub-baccalaureate education. One strand of work has explored college credits as well as degrees and certificates. In earlier work, Thomas Kane and Cecilia Rouse show net advantages to community college attendance even without finishing, finding around a 4 to 7 percent premium for a year of community college (1995). Norton Grubb, however, finds low economic benefits for those who fail to earn a credential or degree (1993, 1997). Steven Brint argues that community colleges tend to have low financial returns and may not be optimal for disadvantaged students by diverting them from four-year college (2003). Based on the notion of negative or zero returns of sub-baccalaureate higher education, the *cooling out* hypothesis argues that community colleges can discourage disadvantaged or underprepared students from achieving their desirable educational goals such as a bachelor's degree (Dougherty 1994; Rosenbaum et al. 2017).

Although mixed findings might be the story of a quarter century ago, more recent estimates suggest positive returns to different types of sub-baccalaureate education. According to James Rosenbaum, Caitlin Ahearn, and Janet Rosenbaum, the economic return to community college is positive for recent cohorts, though it was not significantly different from zero for older cohorts (2017). The extant litera-

ture since the 1990s confirms more substantial economic returns for those who are awarded an associate degree or vocational certificate, even as estimates of the magnitude of such returns vary sharply due to the different datasets and methodologies being used (Belfield and Bailey 2011). For example, Christopher Jepsen, Kenneth Troske, and Paul Coomes show that earnings rise substantially after receiving an associate degree or diploma for men, around 18 to 30 percent, and women, 40 percent, who entered community college in the early 2000s (2014). Stephanie Cellini and Latika Chaudhary document around a 15 percent gain in earnings associated with an associate degree among young adults age twenty-four to thirty in 2008 (2014). Using the 2000 follow-up of the National Education Longitudinal Survey, Dave Marcotte and his colleagues find substantial returns to an associate degree for cohorts who graduated in the 1990s, especially for women (2005). Looking at the median annual earnings of male full-time wage and salary workers age twenty-five and older, Sandy Baum, Charles Kurose, and Michael McPherson also find higher earnings among some college and associate degree holders, but relatively flat earnings for this group over recent decades (2013).

Although a consensus is developing on the positive payoff to an associate degree in recent decades, estimates of other sub-baccalaureate credentials are less established. For example, using a large administrative dataset reflecting information from the North Carolina Community College System, Vivian Liu, Clive Belfield, and Madeline Trimble report that the wage return to community college certificates are negative for both genders relative to those who enroll in a community college but do not earn any credentials (2015). Analyzing the longitudinal Youth Development Survey, Mike Vuolo, Jeylan Mortimer, and Jeremy Staff also find the negative return to certificates relative to the noncompleters (2016). Other studies, however, document higher earnings associated with certificates (Jepsen, Troske, and Coomes 2014). Clive Belfield and Thomas Bailey compute the average quarterly earnings return to certificates across studies that used state-specific administrative data, finding it is $530 for men and $740 for women (2017). Using administrative data from the California Community Colleges system linked to earnings records, Ann Stevens, Michal Kurlaender, and Michel Grosz estimate individual fixed-effects models and find substantially higher earnings among individuals with technically oriented education offered in community college (2018).

Variation of Sub-Baccalaureate Education by Fields of Study

Parallel with discussions on the economic benefits, another important question is the extent to which labor market benefits vary across fields of study. Analysis of heterogeneity in labor market outcomes by field of study addresses broader issues related to horizontal stratification in higher education (Kim, Tamborini, and Sakamoto 2015). Most work in this area, however, has assessed variation of bachelor's and graduate degrees by fields of study rather than sub-baccalaureate level. Among the few studies on sub-baccalaureate degrees, Grubb reports positive returns to an associate degree among women in business and health-related fields but not in some fields such as education (1997). Other studies also reveal relatively higher returns for men and women with associate degrees in health (Jepsen, Troske, and Coomes 2014; Stevens, Kurlaender, and Grosz 2018). A review article argues that previous studies "uniformly show that field of study matters, with returns that are especially high for awards in health-realted fields, high for directly vocational subjects, but negligible for academically focused associate degrees" (Belfield and Bailey 2017, 13).

One aspect that is not adequately addressed in previous studies is the value of various sub-baccalaureate fields relative to different fields of a bachelor's degree. For example, does an associate degree in health bring in a higher labor market benefit than a bachelor's degree in sociology? To our knowledge, no studies including recent research using state-level administrative data have provided such information.

Long-Term Financial Benefit to Sub-Baccalaureate Education

A dimension that has received surprisingly little attention in the study of sub-baccalaureate education relates to long-term labor market ben-

efits. Because the bulk of existing work uses cross-sectional earnings or average earnings over a period, analysis of the evolution of earnings or employment among sub-baccalaureate degree recipients over long stretches of their lives is rare. As a result, the earnings dynamics associated with sub-baccalaureate credits, degrees, or certificates over the life course are not well understood. One exception uses the Virginia Community College System data matched with unemployment insurance data (Jaggars and Xu 2016). Estimating multilevel growth curve models, Shanna Jaggars and Di Xu uncover substantial jumps in earnings after the completion of a credential. Both the immediate increase after completion and the subsequent earnings growth rates are highest for those who eventually earned a bachelor's degree, followed by associate degree holders and certificate holders. In addition, Liu and her colleagues estimate the return five, seven, and nine years after the first enrollment in a community college and find that the return rapidly increases in the first five years but slows thereafter (2015). In this article, we examine the earnings of individuals, following them for twenty years after high school graduation.

METHODS

We use wave-2 data from the SIPP 2004 and 2008 panels matched to the summary earnings record (SER) and the detailed earnings record (DER) files at the Social Security Administration (SSA). The SIPP data provide demographic and socioeconomic characteristics of a nationally representative sample. Critically, wave 2 includes a one-time topical module (educational history) that provides retrospective information about respondents' education. We use these data to measure field of study at the sub-baccalaureate level and to construct partial proxies for respondents' characteristics related to high school type and academic preparation. We pooled wave 2 of the 2004 and the 2008 SIPP panels to acquire adequate sample size to examine fields of study.

Linkages to the DER administrative file provide SIPP respondents' annual earnings from 1980 to 2014 based on their W-2 tax records. We also use the SER file to observe earnings in the relatively small proportion of years we examine prior to 1980. Both administrative files reflect respondents' tax-recorded earnings for all jobs in a year. One key difference is that the DER file contains respondents' full earnings information beyond the maximum taxable earnings and Social Security–covered employment, whereas the SER contains Social Security earnings that are capped at the taxable maximum under the Social Security program in a given year.[1] To mitigate potential concerns about this difference, we use the SER data only in observations prior to 1980 and switch to earnings from the DER thereafter. Also, the vast majority of workers do not reach the Social Security taxable maximum in any given year and work in covered employment—approximately 90 percent of all U.S. workers in the 1970s. We refer to this matched longitudinal dataset as the SIPP-IRS. More detailed descriptions of the SSA administrative records and survey matches may be found elsewhere (see McNabb et al. 2009; Tamborini, Kim, and Sakamoto 2015).

The central advantage of the SIPP-IRS data file is that it permits the construction of long-term earnings profiles over an age- and year-specific period. These data have several advantages over other longitudinal datasets. The National Longitudinal Survey of Youth is limited to particular birth cohorts, whereas the SIPP-IRS provides a larger range of cohorts and greater sample sizes. Although the Panel Survey of Income Dynamics is not limited to a particular birth cohort, the sample is based on households in 1968 with limited representativeness for more recent cohorts. Moreover, sample attrition is minimal in the SIPP-IRS data relative to other longitudinal datasets, and the linked annual earnings based on tax records are generally thought of as more reliable than self reports (Kim and Tamborini 2014).

A possible drawback of the SIPP-IRS is that

1. The earnings subject to Social Security taxes for a given year is limited by the taxable maximum. The vast majority of workers have earnings under this limit in any given year. In 2018, the taxable maximum was $128,400. Social Security earnings also do not include labor income from a small segment of employment not covered under the Social Security program, such as some local and state government workers.

not all SIPP respondents were successfully matched with the administrative data. The match rate of the 2004 and 2008 panels, however, is high at around 80 to 90 percent. To maintain the national representation of the sample, our analyses use a SIPP weight that adjusts for unmatched respondents.

Target Population

Our main analysis sample consists of those who completed high school between 1972 and 1995. We selected these cohorts to construct twenty-year earnings since high school graduation using the 2004 and 2008 SIPP-IRS matched data. Our target population is those whose highest level of attainment is a vocational-technical diploma or certificate, an associate degree, or some college experience without earning a bachelor's degree (college dropouts) as of wave 2 of the SIPP panels. We compare the twenty-year earnings paths of these groups to high school graduates and individuals with a bachelor's degree of the same high school graduation cohort.

We apply a number of sample restrictions to clarify our analysis. We exclude individuals who did not graduate from a high school and those who have a graduate degree. This helps focus our analysis on a comparison of the earnings trajectories of individuals with a sub-baccalaureate education to those of high school graduates and bachelor's degree holders over the same stage of life. We exclude graduate degree holders because the twenty-year cumulative earnings after high school may not reflect the value of their education accurately. Sensitivity analysis reported later examines the effect of including graduate degree holders in our sample of bachelor's degree holders. We limit our sample to those who acquired a high school diploma by age twenty to exclude nontraditional students who might have unique labor market and schooling experience. We also require that individuals with postsecondary education acquired their final degree within ten years after high school graduation. This restriction is desirable because the benefit to sub-baccalaureate education varies by the age of graduation (Jacobson, Lalonde, and Sullivan 2005). The benefit for nontraditional students may be lower than that for on-time graduates. The tiny share of the sample who have zero tax earnings for twenty years after high school graduation are also dropped.

Analytical Strategy

We assess the earnings streams associated with sub-baccalaureate education in two ways. First, we assess three major subgroups: technical-vocational diploma and certificates, associate degree, and college dropouts, also referred to as some college no degree. Second, we disaggregate further by fields of study. In both instances, we carry out descriptive and multivariate analysis.

With regard to the descriptive work, we map out median annual earnings by age over a twenty-year period after high school graduation, limiting our sample in each year to those with positive earnings. We examine the pattern for the three major subgroups of sub-baccalaureate education and by fields of study. For comparative purposes, we include the earnings trajectories of high school graduates and bachelor's degree holders over the same stage of life.

Multivariate analysis examines the relationship between sub-baccalaureate education (three subgroups and fields of study) and twenty-year cumulative earnings starting with the year of high school graduation. Cumulative earnings is a summary measure that reflects a number of complex life-course processes that affect its value. The measure reflects transitions in and out of labor market over the observational period due to unemployment or withdrawal from the labor force. It is also influenced by differences in labor supply over time, wages, and wage growth. Consequently, observed differentials by education level or field of study could be driven by one or all of these factors. In our analysis, we do not systematically examine the factors that explain differences in cumulative earnings or earnings growth. Instead, our goal is to measure the total cumulative earnings associated with sub-baccalaureate education levels and fields of study and to identify the extent of differences across these subgroups. Equation 1 shows our model specification using ordinary least squares (OLS):

$$y_i = \alpha + \sum_j \sum_k \beta_{jk} Degree_{ijk} + \sum_l \gamma_l X_{il} + e_i , \quad (1)$$

where y_i is log-transformed twenty-year cumulative earnings for individual *i*. Because the cumulative earnings includes years in school and unemployment spells, our estimates uncover the total payoffs to education up to around mid-career. Some of these payoffs can be accounted for by the opportunity cost of going to college (the forgone earnings while in school), the change in employability associated with education, and the change in productivity.

$Degree_{ijk}$ refers to the educational attainment for individual *i*. More specifically, it refers to *j* level of education and *k* field for individual *i*. The five levels of education—high school graduates (HSG); vocational-technical diploma or certificate (Voca); associate degree (AA); college dropouts (ColDrop); and bachelor's degree (BA)—set high school graduate as a reference point. One limitation of our study is that the SIPP does not allow us to separate BA holders who transferred from a community college from those who enter a four-year institution without attending a community college. The earnings for BA holders who transferred from a community college tends to be lower than those who started at a four-year institution (Lockwood Reynolds 2012).

For the analysis on fields of study, we construct eight dummy variables for men who have a vocational-technical diploma or certificate and seven dummies for women with a vocational-technical diploma or certificate. In the case of AAs, fields of study for men are constituted by seven dummy variables. In the case of women with AAs, we used the same dummies as men except for the dummy of engineering and drafting due to a small sample size. For BAs, we use the same set of seven dummy variables. The dummy variables reflecting the fields of study for each educational level are presented in table 3. For some college but no degree, we cannot disaggregate by field of study because the SIPP does not provide information on their study subject in college. In total, we examine twenty-two fields of study for men and twenty-one for women in our model. Because the reference group is HSG, the estimated β_{jk} quantifies the relative (dis)advantage of $Degree_{jk}$ in twenty-year cumulative earnings relative to HSG.

X_l refers to the *L* number of control variables. We control for birth year, survey year, and race-ethnicity. We also introduce a number of educational history variables that are not commonly used. These include six dummy variables measuring whether the following high school courses were taken (yes or no): industry, shop, or home economics; business courses; vocational programs; other specialized programs; advanced math and science courses; and college preparation courses. In addition, we measure whether the respondent's high school diploma is GED, whether the respondent attended private high school, whether the respondents were born in the South, and age at the final degree (or age at the college dropout). Therefore, the estimated β_{jk} is the net effect of education after accounting for these control variables.

EMPIRICAL FINDINGS

Table 1 presents descriptive statistics of our sample. Median twenty-year cumulative earnings since high school graduation for men with some college no degree is $664,248, which is 15 percent higher than high school graduates. For a typical man with a vocational diploma or certificate, it is $624,340, which is 8 percent higher than HSG. AA holders have higher twenty-year cumulative earnings since high school graduation ($710,133) than some college no degree and vocational degree holders, and is 24 percent higher than HSG. Bachelor's degree holders' twenty-year cumulative earnings are substantially higher than those of other degree holders, reflecting 48 percent higher earnings than HSG.

Women's twenty-year cumulative earnings since high school graduation is appreciably lower than men's. For example, AA women's median twenty-year cumulative earnings is 38 percent lower than that of equally educated men. In terms of the rank among the sub-baccalaureate degrees, female AA holders earn the highest, followed by some college and vocational degree holders. Thus, the rank order is the same between men and women.

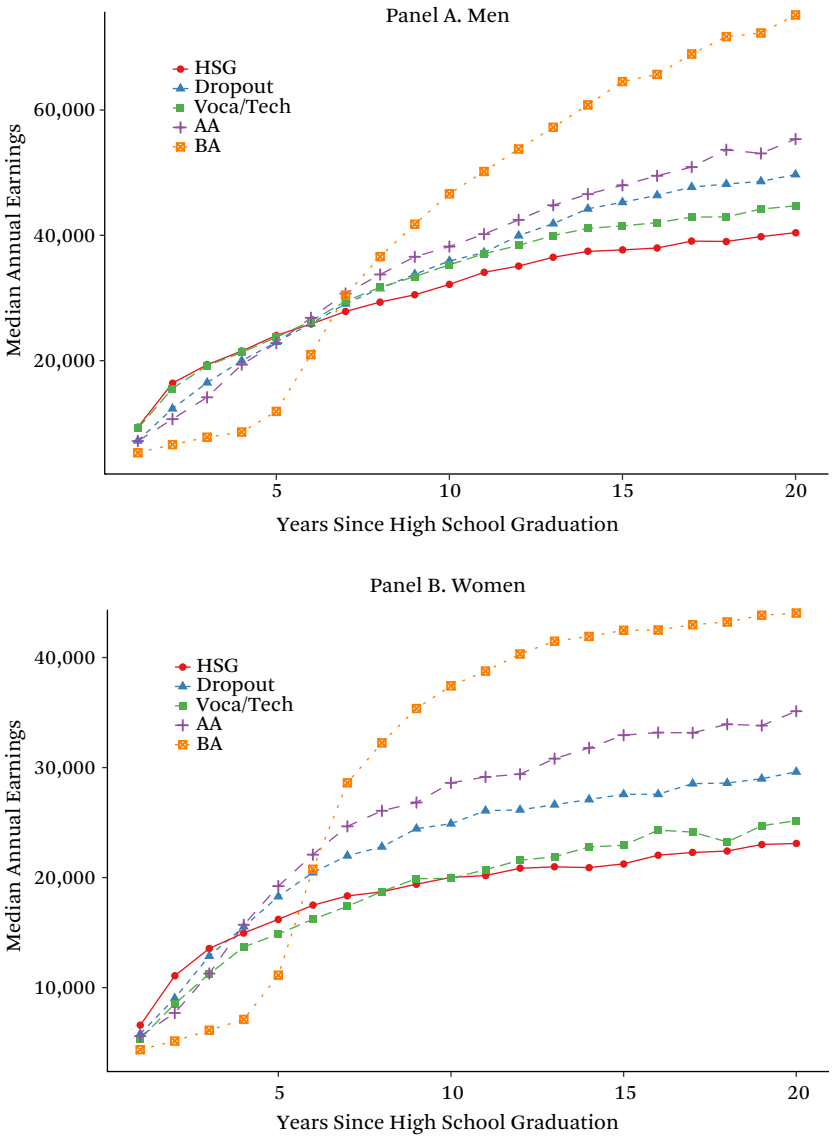
An interesting finding is that the cumulative earnings premium for college dropouts is notably higher than vocational degree holders. This result is consistent with studies showing that taking college credits without credentials

Table 1. Descriptive Statistics

	High School Grad	Some College No Degree (Dropout)	Vocational Degree	Associate Degree	Bachelor's Degree
Panel A. Male					
Twenty-year cumulative earnings (from year of high school graduation)					
Mean ($)	606,716	702,405	668,903	751,955	936,766
Median ($)	576,415	664,248	624,340	710,133	851,174
Median: logged dollar	13.265	13.406	13.345	13.473	13.654
Racial composition					
Whites	75.6	75.1	72.9	81.2	87.0
Blacks	12.8	10.9	13.1	8.3	6.4
Hispanics	8.9	10.7	10.3	7.2	3.5
Other races	2.8	3.3	3.8	3.3	3.1
High school courses and characteristics					
College prep courses	11.8	33.7	14.3	35.9	63.9
Advanced math and science	49.8	68.3	57.5	70.5	86.3
Industrial arts, shop, or home economics	64.8	51.3	64.3	53.5	32.3
Business courses	16.4	24.2	18.5	24.3	24.3
Vocational program	8.7	4.3	15.7	6.1	0.7
Private high school	3.4	7.9	4.3	7.6	14.1
Years taken for degree after high school		3.72	3.24	4.31	5.22
Panel B. Female					
Twenty-year cumulative earnings (from year of high school graduation)					
Mean ($)	313,471	410,686	340,940	470,633	581,544
Median ($)	272,581	367,956	298,581	440,825	529,887
Median: logged dollar	12.516	12.816	12.607	12.996	13.180
Racial composition					
Whites	73.7	76.8	65.6	80.3	84.4
Blacks	14.4	12.7	20.2	10.6	7.8
Hispanics	8.7	7.2	9.9	6.7	4.4
Other races	3.3	3.4	4.3	2.4	3.4
High school courses and characteristics					
College prep courses	14.9	36.3	19.4	42.2	66.6
Advanced math and science	48.6	63.3	55.1	68.4	80.4
Industrial arts, shop, or home economics	46.4	40.3	44.9	39.8	24.8
Business courses	49.8	48.6	47.6	47.4	30.9
Vocational program	3.8	3.1	7.5	3.1	0.70
Private high school	4.6	7.9	4.8	8.1	14.5
Years taken for degree after high school		3.58	3.14	4.03	4.98

Source: Authors' calculations using the SIPP-IRS linked data.
Note: All earnings are adjusted to the 2014 dollars using CPI-U.

Figure 1. Median Annual Earnings Trajectory by Education Level



Source: Authors' calculations using the SIPP-IRS linked data.
Note: Samples are limited to those who reported positive earnings each year.

is associated with higher earnings (Belfield and Bailey 2011).

Courses taken during high school years for HSG and vocational degree holders are also different from those of other degree holders. Only 12 to 14 percent of male and 15 to 19 percent of female HSG and vocational degree holders took college prep courses, whereas more than one-third of AA or BA holders did. Table 1 presents other notable differences across levels of education, including in terms of racial composition, the courses taken during high school, and age of degree completion.

Annual Earnings

Table 1 reveals substantial variation in twenty-year cumulative earnings after high school graduation. To address the life-course pattern of this variation over the early and mid-adulthood, we tracked men's and women's an-

nual earnings over the twenty years after high school graduation, limiting our sample to those with positive earnings in each year. Figure 1 plots the median annual earnings by degree types.

Organizing the data this way provides insights into the shape and speed of earnings growth across education groups by degree type since the year of high school graduation. For the first five years, individuals who end up earning a sub-baccalaureate (except vocational diploma or certificate) or bachelor's degree earn less than high school graduates among men. From the seventh year, however, median earnings for all postsecondary degree holders start to surpass high school graduates. From that point onward, the gaps in median annual earnings grow continuously across degree types.

Among men, the difference in the patterns of earnings growth between bachelor's degrees and all other degrees is noteworthy. For sub-baccalaureate education groups, median earnings are lower than those for HSG in the first several years after high school graduation because they spend their time in school. The earnings of all sub-baccalaureate education groups then start to surpass the HSG, not because of an observable jump or discontinuity in earnings around the period likely to surround degree acquisition, but rather because the earnings growth rates are consistently higher. By contrast, as shown in figure 1, the earnings growth rates for BA holders are almost flat in the first four years after high school, at which point median earnings jump rapidly. These patterns indicates that school-to-work transitions for BA holders is conceivably different those for sub-baccalaureate educations.

Women show a similar pattern as men but with a few important differences as well. The earnings growth rates for female workers seven or eight years since high school graduation become much flatter relative to men. The earnings gaps across levels of education are constant after ten years since high school graduation for women. Another notable gender difference is the trajectory of vocational-technical diploma or certificate. The advantage of vocational or technical diploma over high school graduation seems to be more evident for men than for women.

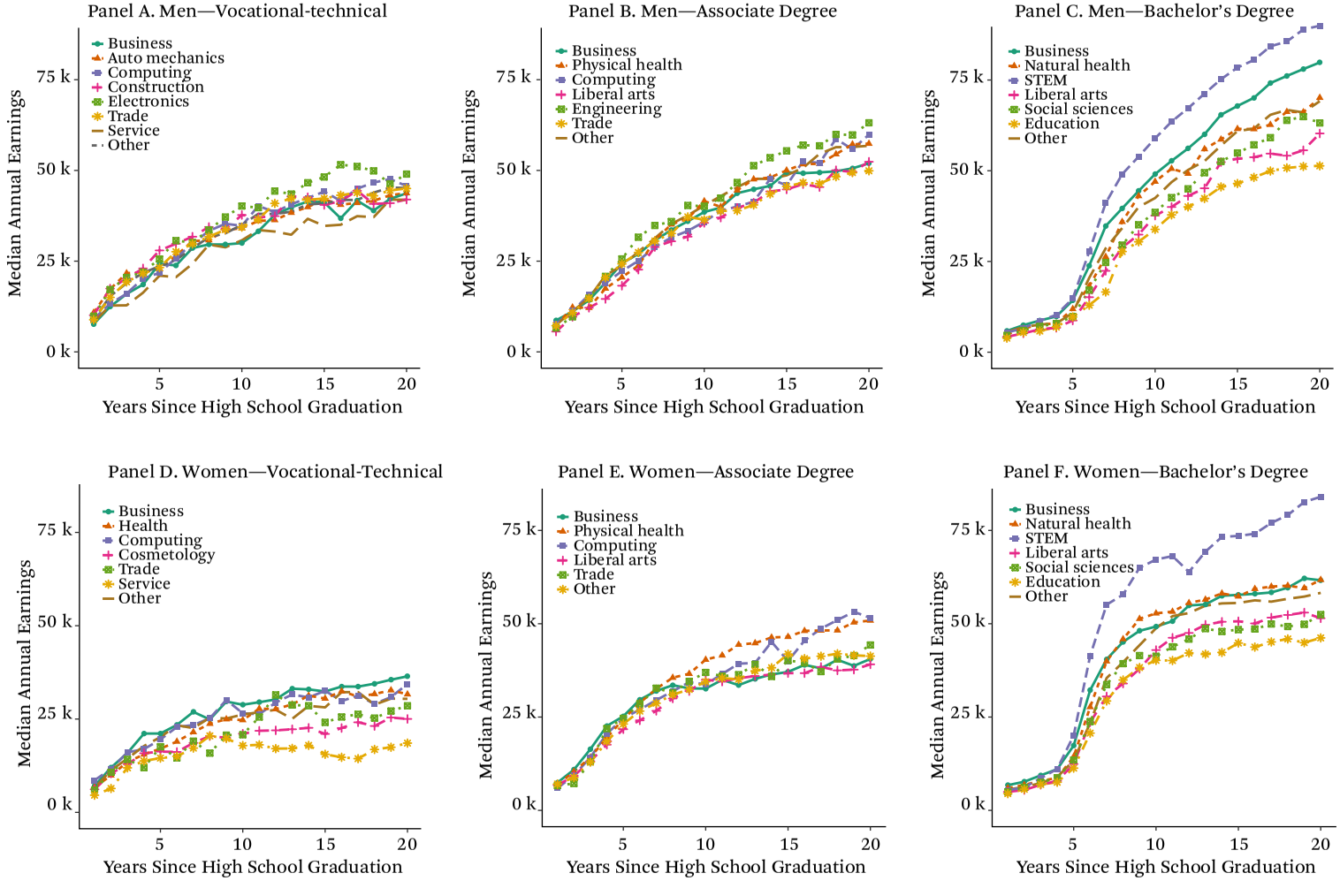
To what extent do the stylized earnings profiles observed in figure 1 represent most subfields in each degree type? Figure 2 addresses this question by tracing out the median annual earnings by fields of study within sub-baccalaureate and bachelor's degrees over twenty years. Several stylized patterns are noteworthy. First, we observe meaningful variation in the evolution of earnings from early to mid-adulthood by fields of study. Among men with vocational degrees or certificates, electronics and computing are associated with higher median earnings, particularly beginning at around ten years after high school graduation. Likewise, the highest median earnings among associate degree holders were associated with those studying engineering, computing, and physical health. Among women with vocational degrees or certificates, studying business and computing is associated with higher annual earnings than those with credentials in service and cosmetology fields. Among AAs, those with degrees in physical health and computing have higher median earnings than those with degrees in liberal arts, namely, toward the end of the observation window.

Second, despite important variation in annual earnings by fields of study within sub-baccalaureate education groups, BA holders appear to show even more heterogeneity. Third, the advantage of BA over sub-baccalaureate education evident in figure 1 is not as clear when these groups are disaggregated by fields of study, particularly for men. Except for business and STEM majors, the earnings trajectories of AA majors appear to overlap with those of BA majors. This suggests that the clear advantage of BAs over AAs in figure 1 is driven by a small number of BA majors.

Finally, the stylized patterns described are more evident for men than for women. Diversity is higher in women's earnings trajectories across fields of study with the same sub-baccalaureate credential. Furthermore, despite clear overlap in earnings trajectories across education levels for men, overlap is much less evident for women.

Figures 1 and 2 are informative but do not account for spells out of the labor force. Consequently, estimates of the earnings gap between the less educated and the highly edu-

Figure 2. Median Annual Earnings Trajectory by Fields of Study



Source: Authors' calculations using the SIPP-IRS linked data.

cated might be downwardly biased over the long term because the less educated are more likely to be out of the labor force. At the same time, those who earned AAs or BAs are more likely to be out of the labor force in their early twenties. Thus, the earnings advantage of the less educated over the highly educated gap during the first five years after high school graduation could be larger on a cumulative basis. To address these concerns, we estimate a series of OLS models to explore the relationships between sub-baccalaureate degrees and fields of study with log twenty-year cumulative earnings after controlling for demographic and high school education covariates.

Regression Analysis by Sub-Baccalaureate Degree: Cumulative Twenty-Year Earnings

Table 2 summarizes the OLS results by level of education using gender-stratified models. The first model is the baseline specification that includes birth year. The second model adds controls for race, high school education, and age of degree acquisition covariates.

Relative to HSG, all three sub-baccalaureate education groups exhibit substantially higher twenty-year cumulative earnings even after controlling for the extensive set of high school education covariates. A vocational degree yields a 17 percent premium (=e.158) for men relative to HSG and a 20 percent premium for women when demographic and high school covariates are controlled for. The financial benefit for college dropouts (some college no degree) is also substantial for both genders. As discussed, studies that used cross-sectional survey data or state-level administrative data do not consistently show the positive earnings premium for all sub-baccalaureate education. However, our analyses with nationally representative data reveal substantial premiums for all sub-baccalaureate education including vocational certificates when demographic and high school covariates are taken into account. As expected, the benefits for sub-baccalaureate education are smaller than those for bachelor's degrees.

Thus far, our main results are interesting but not particularly surprising. We find that BAs earns more than AAs, and AAs earn more than vocational degrees or those with some college but no degree. A growing interest is whether this stylized fact holds if we disaggregate educational attainment by fields of study. Thanks to recent studies using state-level administrative data, consensus is increasing that subjects matter (on data use, Jepsen, Troske, and Coomes 2014; Liu, Belfield, and Trimble 2015; Stevens, Kurlaender, and Grosz 2018; Xu and Trimble 2016; on consensus, Belfield and Bailey 2011, 2017). We contribute to this literature by examining the extent to which field of study in sub-baccalaureate education is associated with long-term earnings within the same sub-baccalaureate group and relative to fields of study among BA holders.

Regression Analysis by Field of Study

We estimate the net value of some college education on twenty-year long-term earnings from early to mid-adulthood across fields of study after controlling for demographic and high school–educated related covariates in table 3. Like table 2, the dependent variable is log-transformed twenty-year cumulative earnings after high school graduation. The reference group for the coefficients estimated in table 3 is HSG. To facilitate the comparison across educational levels and fields, we present the results of models 6 and 8 in figure 3.

Table 3 and figure 3 reveal a surprisingly high degree of variation within the same level of education when demographic and high school educated related covariates are controlled for. More important, men with an associate degree or even a vocational diploma in technical fields earn more in the first twenty years after high school graduation than what BAs in liberal arts or humanities earn over the same period. The latter finding corroborates that the financial benefits of sub-baccalaureate education is as high as bachelor's degrees, depending on majors.

For example, an AA in engineering has cumulative earnings that are 18 percent (=exp(.296-.132)–1)) higher than a BA in liberal arts and humanities, and the gap is statistically significant. The estimated difference in the twenty-year cumulative earnings between a BA and an AA in liberal arts and humanities is tiny and statistically not different from zero. All seven

Table 2. OLS Regression of Education Level on Twenty-Year Cumulative Earnings After High School Graduation

	Male				Female			
	Model 1		Model 2		Model 3		Model 4	
	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
Control variables								
Birth year	o		o		o		o	
Survey year	o		o		o		o	
Race			o				o	
High school courses[a]			o				o	
GED			o				o	
Private high school			o				o	
Born in the South			o				o	
Age of final degree			o				o	
(Reference = high school graduate)								
Some college no degree	0.185	(0.019)***	0.159	(0.020)***	0.392	(0.029)***	0.329	(0.032)***
Vocational diploma-certificate	0.141	(0.024)***	0.158	(0.024)***	0.193	(0.030)***	0.203	(0.032)***
Associate degree	0.297	(0.021)***	0.248	(0.025)***	0.578	(0.028)***	0.495	(0.032)***
Bachelor's degree	0.488	(0.015)***	0.401	(0.024)***	0.786	(0.022)***	0.664	(0.028)***
R^2	0.069		0.117		0.104		0.125	
N	15,418		15,418		16,351		16,351	

Source: Authors' calculations using the SIPP-IRS linked data.

Note: Adjusted weights are applied to reflect the complex survey design and varying match rates by groups.

[a] High school courses include six dummy variables indicating whether the respondent took industrial, shop, or home economics courses; business courses; a vocational program; another program; Adavanced Placement math or science courses; and college prep courses.

*p < .05; **p < .01; ***p < .001 (two-tailed test)

Table 3. OLS Regression of Education Level and Fields of Study on Twenty-Year Cumulative Earnings After High School Graduation

	Male				Female			
	Model 5		Model 6		Model 7		Model 8	
	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
Control variables								
Birth year	o		o		o		o	
Survey year	o		o		o		o	
Race			o				o	
High school courses[a]			o				o	
GED			o				o	
Private high school			o				o	
Born in the South			o				o	
Age at final degree			o				o	
(Reference = high school graduate)								
Some college no degree	0.185	(0.019)***	0.160	(0.020)***	0.392	(0.029)***	0.335	(0.032)***
Vocational diploma or certificate								
Trade	0.189	(0.029)***	0.200	(0.032)***	0.183	(0.085)*	0.206	(0.087)*
Business, office manager	0.084	(0.087)	0.085	(0.087)	0.306	(0.048)***	0.309	(0.048)***
Communications, information services	0.122	(0.126)	0.141	(0.115)	0.273	(0.085)**	0.290	(0.085)**
Services	-0.203	(0.035)	-0.173	(0.119)	-0.234	(0.157)	-0.172	(0.149)
Construction trades	0.194	(0.055)***	0.216	(0.055)***	—		—	
Electronics	0.294	(0.056)***	0.286	(0.055)***	—		—	
Auto mechanics	0.129	(0.047)**	0.147	(0.049)**	—		—	
Cosmetology	—		—		0.055	(0.053)	0.061	(0.054)
Health care	—		—		0.226	(0.048)***	0.246	(0.052)***
Other	0.118	(0.035)***	0.142	(0.033)***	0.191	(0.052)***	0.211	(0.055)***

(*continued*)

Table 3. (*continued*)

	Male				Female			
	Model 5		Model 6		Model 7		Model 8	
	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
Associate degree								
Business, office manager	0.304	(0.039)***	0.243	(0.045)***	0.563	(0.046)***	0.483	(0.046)***
Physical sciences, health	0.323	(0.063)***	0.296	(0.067)***	0.792	(0.041)***	0.714	(0.044)***
Communications, information services	0.331	(0.058)***	0.313	(0.060)***	0.655	(0.107)***	0.569	(0.107)***
Liberal arts, humanities	0.167	(0.067)*	0.140	(0.066)*	0.452	(0.050)***	0.382	(0.054)***
Other vocation, trade	0.260	(0.034)***	0.198	(0.035)***	0.529	(0.080)***	0.453	(0.082)***
Engineering, drafting	0.382	(0.048)***	0.304	(0.051)***	—		—	
Other	0.325	(0.039)***	0.282	(0.035)***	0.481	(0.048)***	0.415	(0.050)***
Bachelor's degree								
Business	0.591	(0.021)***	0.509	(0.030)***	0.920	(0.031)***	0.797	(0.035)***
STEM	0.713	(0.025)***	0.619	(0.030)***	1.156	(0.058)***	1.058	(0.060)***
Natural sciences, health	0.459	(0.041)***	0.375	(0.047)***	0.933	(0.035)***	0.817	(0.048)***
Social sciences, history, communications	0.386	(0.030)***	0.324	(0.034)***	0.690	(0.040)***	0.528	(0.041)***
Education	0.075	(0.076)	-0.004	(0.068)	0.499	(0.034)***	0.396	(0.038)***
Liberal arts, humanities	0.218	(0.048)***	0.132	(0.048)**	0.679	(0.032)***	0.562	(0.037)***
Other	0.416	(0.024)***	0.326	(0.027)***	0.827	(0.036)***	0.718	(0.039)***
R^2	0.086		0.132		0.114		0.135	
N	15,418		15,418		16,351		16,351	

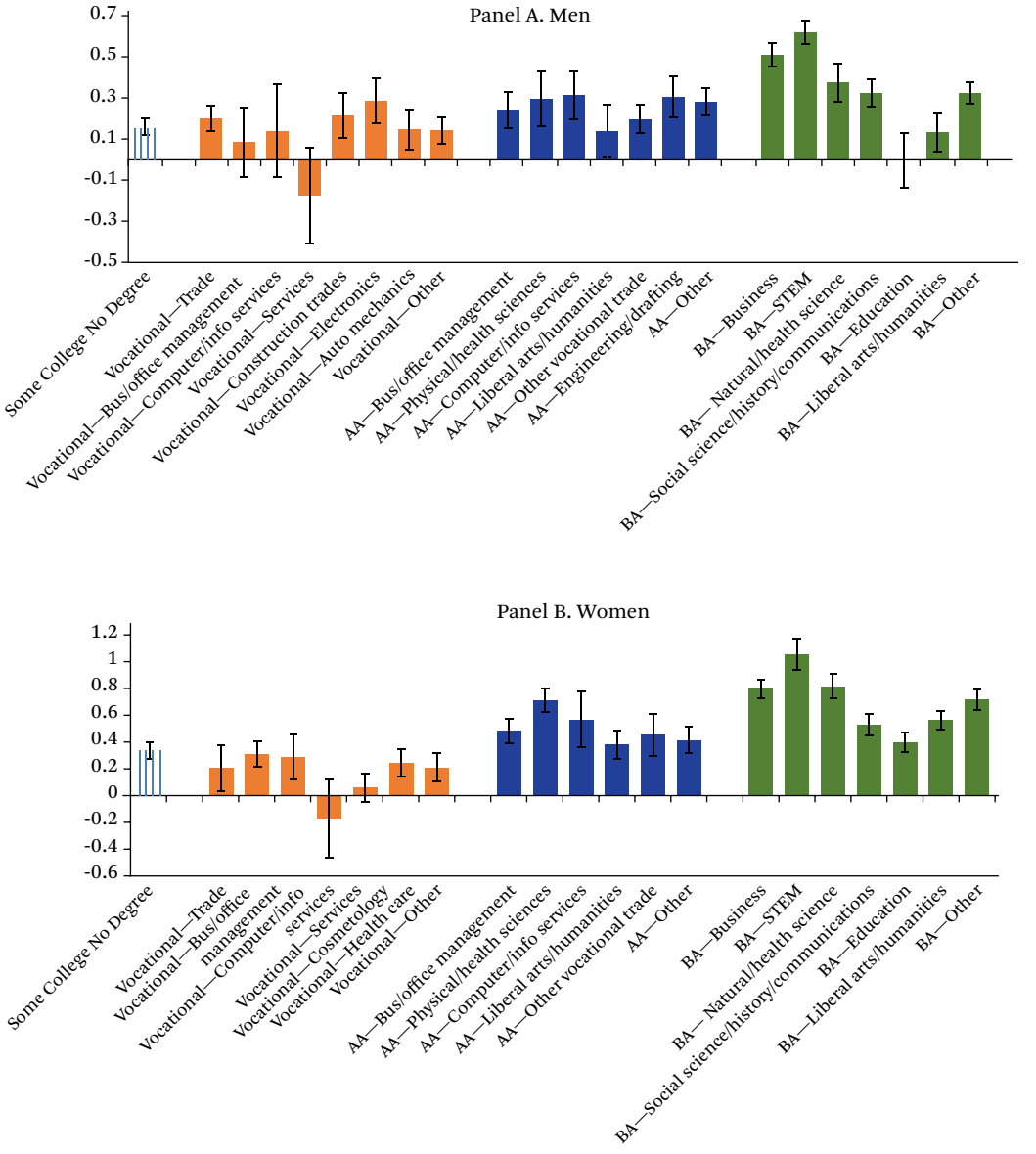
Source: Authors' calculations using the SIPP-IRS linked data.

Note: Adjusted weights are applied to reflect the complex survey design and varying match rates by groups.

[a] High school courses include six dummy variables indicating whether the respondent took industrial, shop, home economics courses; business courses; vocational program, other program, Adavanced Placement math or science courses, and college prep courses.

*p < .05; **p < .01; ***p < .001 (two-tailed test)

Figure 3. Relative Return to Higher Education in Twenty-Year Cumulative Log Earnings Relative to High School Graduates



Source: Authors' compilation based on table 3.
Note: Y-axis indicates the log earnings gap relative to high school graduation. The results are based on the estimates of models 6 and 8 of table 3.

majors of AA assessed in this study show higher average cumulative earnings than BAs in liberal arts and humanities (four are statistically significantly higher). This finding implies that the value of AAs is not limited to health-related majors and engineering. The coefficient of an AA in liberal arts and humanities is not statistically different than that of a BA. That is, the value of a BA is not significantly higher than the value of an AA in regard to twenty-year cumulative earnings after high school graduation. Of the seven BA fields, only business and STEM majors

show statistically significantly higher cumulative earnings than all majors in AA or vocational diploma or certificates. Natural and health sciences show higher earnings than all majors in AA fields or vocational diploma or certificates. But the differences between BAs and AAs in the natural and health sciences are not always statistically significant.

Even among the eight vocational fields assessed in this study, none of the estimated coefficients are statistically significantly lower than that of the BA in the liberal arts and humanities except the vocational diploma in services. Comparing AA and vocational diplomas or certificates, AAs are worth more on average, but those who have a vocational diploma in electronics earn as much as any other AA majors, if not higher. The net earnings of some college without a degree are obviously higher than HSG and not lower than those who have a BA in the liberal arts and humanities.

Gender Differences in Financial Benefits Associated with Sub-Baccalaureate Education

Several gender differences are noteworthy. The value of higher education for women is more clearly hierarchical by level than it is for men. Relative to HSG, higher education is associated with a greater twenty-year cumulative earnings advantage for women than for men. The gap between HSG and AAs is almost twice as large for women (0.495) as it is for men (0.248), as shown in table 2. The only level of education that does not show a substantial gender difference is vocational-technical diploma or certificate. The result that the premium for women is higher than for men is consistent with other studies (Dadgar and Trimble 2015; Gill and Leigh 2003; Liu, Belfield, and Trimble 2015). These results do not mean that college-educated women earn more than equally educated men. However, the relative benefit of college education is almost twice as large for women as that for men. This large relative benefit for women might be a partial explanation on the rise of women in higher education (DiPrete and Buchmann 2013).

Unlike that for men, the clearly vertical hierarchy by educational level for women is evident again when disaggregated by fields of study. Among the six AA majors assessed in this study, two—physical and health services and computer and information services—exhibit a higher coefficient than that of BAs in the liberal arts and humanities, and only one—physical and health services—shows a statistically significantly difference. The earnings advantages of health-related AA majors reported in previous studies is evident for women. Of the seven BA fields we examined, four—business, STEM, natural and health sciences, and other—seem to earn more on a twenty-year cumulative basis than all other AA or vocational training fields on average, although some of differences are not statistically significant. The value of associate degrees for women looks higher than vocational training regardless of field. Put simply, horizontal stratification by fields of study seems to be more important for men than for women in regard to long-term earnings covering early to middle career. Conversely, vertical stratification seems to be more important for women than for men. These findings warrant further investigation.

Yet we hasten to add caveats on the interpretation of our findings. Our models do not include hours worked due to limitations in the linked tax data. To be sure, many of the trends observed for women are influenced by the timing of marriage and motherhood and its attendent impacts on labor supply. As noted, one aim of this article was to document the earnings patterns among female workers by education level rather than to disentangle the causal impact of degree types and field of study that may be comingled with selection processes related to fertility, education investments, and work patterns.

Robustness Checks

Some may wonder whether the relative earnings of BAs are underestimated in our models, given that graduate degree holders are excluded from the primary analyses. It is debatable whether BAs-only or BAs or higher should be a comparison group in assessing the relative value of AAs. Nevertheless, we conducted a series of sensitivity tests that reestimate models of tables 2 and 3 with a sample including grad-

uate degree holders to address the concern of underestimation.[2] As expected, the value of a BA or higher is a bit larger in the new estimation, but the increase is small. For example, when we estimate the same model with model 1 of table 2, the coefficients of BA+ become 0.522 for male and 0.861 for female, which are only 0.034 and 0.075 higher respectively than what we report in table 2. When other covariates such as race and high school performances are controlled for (the same model specification with model 2 of table 2), the estimated coefficient of BA+ for men (0.378) becomes smaller than that of BA-only (0.401 in table 2).

The estimations including fields of study also are not substantively different either when we include graduate degree holders in the sample of BAs. Net of demographic and high school performance covariates, none of the seven undergraduate majors among bachelor's or higher degree holders become more significant than reported in table 3. For some majors, the coefficients become smaller as graduate degree holders are added in our analyses. This counterintuitive result is partially because many graduate degree holders are out of the labor force in their twenties and even their early thirties, and thus have a relatively shorter span to accumulate labor income over the twenty-year observational period beginning after high school graduation.

One noteworthy change when graduate degree holders are included in our analyses is the increase in the value of an education major. When we include graduate degree holders, the coefficient of education major among men becomes highly significant (0.218) without controlling for race and high school performance. However, when race and high school performance are controlled for, the coefficient becomes substantially lower (0.059), which is statistically not different from zero.

CONCLUSIONS

In this study, we use data from the SIPP and matched administrative tax information to examine the long-term labor market benefits associated with sub-baccalaureate degrees and certificates. Following men and women for twenty years after high school graduation, we provide new estimates of the cumulative earnings associated with awards to various programs within the sub-baccalaureate level and field of study. We also add to the growing, yet still small, body of literature on the labor market outcomes associated with sub-BA education by documenting how earnings unfold from early to mid-career by educational level and fields of study therein.

Several notable findings emerge from our analysis. Our estimates showing higher annual and cumulative twenty-year earnings up to mid-adulthood among individuals with sub-baccalaureate education at all levels (associate degree, vocational diploma or certificate, and college dropout) relative to high school graduates are consistent with the notion that sub-baccalaureate education has long-term economic benefits. We shed related new light on some important gender differences. Consistent with earlier work, the value of an AA over vocational training is higher for women than for men (Dougherty 2005). We also provide strong evidence of substantial variation by fields of study. The sub-baccalaureate field with the highest long-term, twenty-year cumulative earnings net of control variables were in fields associated with health, technical, or craft skills, which is consistent with research (Jepsen, Troske, and Coomes 2014). For vocational diplomas or certificates, we find relatively higher payoffs for men with awards in construction, electronics, and auto mechanics.

A somewhat surprising finding is that the earnings of some sub-baccalaureate majors are high relative to some types of bachelor's degrees over early to mid-adulthood. For example, the cumulative twenty-year earnings from high school graduation for individuals with associate degrees in physical and health sciences, computer and information services, and engineering and drafting is equal to or higher than those BA fields except business and STEM. Consider also that the earnings levels associated with vocational training in electronics and construction trades for men is higher than those

2. Results are not shown here but can be obtained from the authors on request.

with BAs in liberal arts and the humanities. These patterns suggest that a bachelor's degree is not necessarily associated with a higher financial benefit to the investment in education than an associate degree, at least considering the twenty years from high school graduation that incorporate years before and after degree acquisition. This pattern is driven, in part, by substantially greater heterogeneity in earnings by fields of study within bachelor's degree holders, especially among men. The cumulative earnings gap between BA STEM majors and BA liberal arts or humanities majors is 63 percent among men, whereas that between AA engineering or drafting majors and AA liberal arts and humanities majors is 17 percent.

Our results speak to broader themes related to human capital investments in postsecondary education. Education has become an increasingly key determinant of a person's life chances (Fischer and Hout 2006). Horizontal stratification within higher education (fields of study) has received relatively less attention than vertical stratification (levels of education) as a mechanism of economic stratification (Gerber and Cheung 2008). This study's results are consistent with a growing body of studies showing the importance of horizontal stratification (Kim, Tamborini, and Sakamoto 2015; Ma and Savas 2014; Torche 2018). In light of recent technological changes and economic trends, some students who studied engineering or health services may benefit more from community college programs than from studying nontechnical BA majors (Oreopoulos and Petronijevic 2013). Kim, Tamborini, and Sakamoto present strong evidence that horizontal stratification in higher education (BA, graduate degree) is a stronger determinant than vertical stratification in terms of forty-year lifetime earnings (2015). In this context, recent reports that the proportion of students who major in humanities and social sciences has decreased, whereas the proportion of STEM majors has increased after the Great Recession might reflect a response to the growing importance of horizontal stratification (Blom, Cadena, and Keys 2015).

Our findings also are relevant to the cooling out hypothesis. Kevin Dougherty labels community college as the contradictory college because it appears to structure both aspirations and failure (1994). Contrary to this argument, our results show that except for a small number of subfields, almost all sub-baccalaureate education brings in substantial financial benefits relative to high school graduates. Given these findings, it is hard to say that sub-baccalaureate education is an institutional process that structures and legitimates failure.

We note several important caveats throughout this article, but several points are worth underscoring. To begin with, our results should not be interpreted as evidence indicating that the total value of a bachelor's degree is low. The financial benefits to a bachelor's degree tend to be larger at the advanced mid-career and later stage of work life, which were not examined in this study (Kim, Tamborini, and Sakamoto 2015; Tamborini, Kim, and Sakamoto 2015). Moreover, other financial compensations such as pensions, health insurance, and longer vacation times are tied to higher education (Kristal, Cohen, and Mundlak 2011; Tamborini and Kim 2017; Zajacova, Montez, and Herd 2014). Furthermore, the total benefit of education is not restricted to direct monetary compensations (Hout 2012; Oreopoulos and Salvanes 2011).

One concern might be whether the relatively high value of sub-BA education in our estimations is driven by the inclusion of older cohorts in our sample (1970s graduates) such that the value of sub-BA could be lower in recent cohorts. The number of community colleges in the 1970s was small and concentrated in a small number of states but expanded rapidly after that (Dougherty 1988). To address this concern, we disaggregated our sample into two cohorts (those who completed high school between 1972 and 1984 and those who completed between 1985 and 1995) and estimated the same models contained in tables 2 and 3. The results (available on request) are almost identical between the two cohorts. Nevertheless, we cannot rule out the possibility that those who obtained a sub-baccelaureate degree in the twenty-first century have different experiences than previous generations. Future research on this topic is warranted.

Another concern relates to the effects of unobserved heterogeneity. Selection into four-year institutions is more likely to be positive than

selection into sub-baccalaureate education. Therefore, the actual gap between BA and sub-BA net of uncontrolled heterogeneity might be smaller than our estimates. On the flip side, selection into sub-baccalaureate education is more likely to be positive than among those who stop formal education at high school graduation. Thus, our estimates showing the earnings advantages of sub-baccalaureate education relative to HSG could be upwardly biased.[3] Our estimated coefficients in this study should not be interpreted as causal.

Altogether, our study illustrates one way that linking survey and administrative data can enable researchers to tackle new questions and produce estimates that would be otherwise unfeasible by making use of information provided by both data sources. Survey data linked to the IRS or SSA earnings information provide ample opportunity to study individuals' long-term earnings histories and to track the changes in annual earnings for the same individuals over long stretches of their lives. Federal statistical research data centers are one place researchers can access such data linkages.[4]

REFERENCES

Baum, Sandy, Charles Kurose, and Michael McPherson. 2013. "An Overview of American Higher Education." *The Future of Children* 23(1): 17–39.

Belfield, Clive R., and Thomas Bailey. 2011. "The Benefits of Attending Community College: A Review of the Evidence." *Community College Review* 39(1): 46–68.

———. 2017. "The Labor Market Returns to Sub-Baccalaureate College: A Review." CAPSEE working paper. New York: Teachers College, Columbia University. Accessed November 6, 2018. https://ccrc.tc.columbia.edu/media/k2/attachments/labor-market-returns-sub-baccalaureate-college-review.pdf.

Blom, Erica, Briand C. Cadena, and Benjamin J. Keys. 2015. "Investment over the Business Cycle: Insights from College Major Choice." IZA discussion paper no. 9167. Bonn: Institute of Labor Economics.

Brand, Jennie E., and Yu Xie. 2010. "Who Benefits Most from College? Evidence for Negative Selection in Heterogeneous Economic Returns to Higher Education." *American Sociological Review* 75(2): 273–302.

Brint, Steven. 2003. "Few Remaining Dreams: Community Colleges Since 1985." *Annals of the American Academy of Political and Social Science* 586(1): 16–37.

Cellini, Stephanie Riegg, and Latika Chaudhary. 2014. "The Labor Market Returns to a for-Profit College Education." *Economics of Education Review* 43: 125–40.

Dadgar, Mina, and Madeline Joy Trimble. 2015. "Labor Market Returns to Sub-Baccalaureate Credentials: How Much Does a Community College Degree or Certificate Pay?" *Educational Evaluation and Policy Analysis* 37(4): 399–418.

Davies, Scott, and Neil Guppy. 1997. "Fields of Study, College Selectivity, and Student Inequalities in Higher Education." *Social Forces* 75(4): 1417–38.

DiPrete, Thomas A., and Claudia Buchmann. 2013. *The Rise of Women: The Growing Gender Gap in Education and What It Means for American Schools*. New York: Russell Sage Foundation.

Dougherty, Christopher. 2005. "Why Are the Returns to Schooling Higher for Women than for Men?" *Journal of Human Resources* XL(4): 969–88.

Dougherty, Kevin J. 1987. "The Effects of Community Colleges: Aid or Hindrance to Socioeconomic Attainment?" *Sociology of Education* 60(2): 86–103.

———. 1988. "The Politics of Community College Expansion: Beyond the Functionalist and Class-Reproduction Explanations." *American Journal of Education* 96(3): 351–93.

———. 1994. *The Contradictory College: The Conflicting Origins, Impacts, and Futures of the Community College*. Albany: State University of New York Press.

Fischer, Claude S., and Michael Hout. 2006. *Century of Difference: How America Changed in the Last One Hundred Years*. New York: Russell Sage Foundation.

Gerber, Theodore P., and Sin Yi Cheung. 2008. "Horizontal Stratification in Postsecondary Education:

3. To address this concern, we estimate annual earnings growth rates after controlling for individual fixed effects. The average earnings growth rates after degree completion are twice as high for AAs as they are for HSGs.

4. See "Federal Statistical Research Data Centers," U.S. Census Bureau, https://www.census.gov/fsrdc (accessed November 7, 2018).

Forms, Explanations, and Implications." *Annual Review of Sociology* 34(1): 299–318.

Gill, Andrew M., and Duane E. Leigh. 2003. "Do the Returns to Community Colleges Differ Between Academic and Vocational Programs?" *Journal of Human Resources* 38(1): 134–55.

Grubb, W. Norton. 1993. "The Varied Economic Returns to Postsecondary Education: New Evidence from the Class of 1972." *Journal of Human Resources* 28(2): 365–82.

———. 1997. "The Returns to Education in the Sub-Baccalaureate Labor Market, 1984–1990." *Economics of Education Review* 16(3): 231–45.

Grusky, David B., Michael Hout, Timothy M. Smeeding, and C. Matthew Snipp. 2019. "The American Opportunity Study: A New Infrastructure for Monitoring Outcomes, Evaluating Policy, and Advancing Basic Science." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(2): 20–39. DOI: 10.7758/RSF.2019.5.2.02.

Hout, Michael. 2012. "Social and Economic Returns to College Education in the United States." *Annual Review of Sociology* 38(1): 379–400.

Jacobson, Louis, Robert J. Lalonde, and Daniel Sullivan. 2005. "The Impact of Community College Retraining on Older Displaced Workers: Should We Teach Old Dogs New Tricks?" *ILR Review* 58(3): 398–415.

Jaggars, Shanna Smith, and Di Xu. 2016. "Examining the Earnings Trajectories of Community College Students Using a Piecewise Growth Curve Modeling Approach." *Journal of Research on Educational Effectiveness* 9(3): 445–71.

Jepsen, Christopher, Kenneth Troske, and Paul Coomes. 2014. "The Labor-Market Returns to Community College Degrees, Diplomas, and Certificates." *Journal of Labor Economics* 32(1): 95–121.

Kane, Thomas, and Cecilia Rouse. 1995. "Labor Market Returns to Two- and Four-Year Colleges: Is a Credit a Credit and Do Degrees Matter?" *American Economic Review* 85(3): 600–14.

Kena, Grace, Lauren Musu-Gillette, Jennifer Robinson, Xiaolei Wang, Amy Rathbun, Jijun Zhang, Sidney Wilkinson-Flicker, Amy Barmer, and Erin Dunlop Velez. 2015. *The Condition of Education 2015*. Washington: National Center for Education Statistics.

Kim, ChangHwan, and Christopher R. Tamborini. 2014. "Response Error in Earnings: An Analysis of the Survey of Income and Program Participation Matched With Administrative Data." *Sociological Methods & Research* 43(1): 39–72.

Kim, ChangHwan, Christopher R. Tamborini, and Arthur Sakamoto. 2015. "Field of Study in College and Lifetime Earnings in the United States." *Sociology of Education* 88(4): 320–39.

Kristal, Tali, Yinon Cohen, and Guy Mundlak. 2011. "Fringe Benefits and Income Inequality." *Research in Social Stratification and Mobility* 29(4): 351–69.

Liu, Vivian Y. T., Clive R. Belfield, and Madeline J. Trimble. 2015. "The Medium-Term Labor Market Returns to Community College Awards: Evidence from North Carolina." *Economics of Education Review* 44 (February): 42–55.

Lockwood Reynolds, C. 2012. "Where to Attend? Estimating the Effects of Beginning College at a Two-Year Institution." *Economics of Education Review* 31(4): 345–62.

Ma, Yingyi, and Gokhan Savas. 2014. "Which Is More Consequential for Income Disparity: Fields of Study or Institutional Selectivity?" *Review of Higher Education* 37(2): 221–47.

Marcotte, Dave E., Thomas Bailey, Carey Borkoski, and Greg S. Kienzl. 2005. "The Returns of a Community College Education: Evidence From the National Education Longitudinal Survey." *Educational Evaluation and Policy Analysis* 27(2): 157–75.

McNabb, Jennifer, David Timmons, Jae Song, and Carolyn Puckett. 2009. "Uses of Administrative Data at the Social Security Administration." *Social Security Bulletin* 69(1): 75–84.

Monk-Turner, Elizabeth. 1990. "The Occupational Achievements of Community and Four-Year College Entrants." *American Sociological Review* 55(5): 719–25.

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

Oreopoulos, Philip, and Kjell G. Salvanes. 2011. "Priceless: The Nonpecuniary Benefits of Schooling." *Journal of Economic Perspectives* 25(1): 159–84.

Rosenbaum, James E., Caitlin E. Ahearn, and Janet E. Rosenbaum. 2017. *Bridging the Gaps: College Pathways to Career Success*. New York: Russell Sage Foundation.

Sandefur, Gary D., and Hyunjoon Park. 2007. "Educational Expansion and Changes in Occupational

Returns to Education in Korea." *Research in Social Stratification and Mobility* 25(4): 306–22.

Stevens, Ann Huff, Michal Kurlaender, and Michel Grosz. 2018. "Career Technical Education and Labor Market Outcomes: Evidence From California Community College." *Journal of Human Resources*. Published online April 4, 2018. DOI:10.3368/jhr.54.4.1015.7449R2.

Tamborini, Christoper R., and ChangHwan Kim. 2017. "Education and Contributory Pensions at Work: Disadvantages of the Less Educated." *Social Forces* 95(4): 1577–606.

Tamborini, Christopher R., ChangHwan Kim, and Arthur Sakamoto. 2015. "Education and Lifetime Earnings in the United States." *Demography* 52(4): 1383–407.

Torche, Florencia. 2018. "Intergenerational Mobility at the Top of the Educational Distribution." *Sociology of Education* 91(4): 266–89.

Vuolo, Mike, Jeylan T. Mortimer, and Jeremy Staff. 2016. "The Value of Educational Degrees in Turbulent Economic Times: Evidence from the Youth Development Study." *Social Science Research* 57 (May): 233–52.

Xu, Di Xu, and Madeline Trimble. 2016. "What About Certificates? Evidence on the Labor Market Returns to Nondegree Community College Awards in Two States." *Educational Evaluation and Policy Analysis* 38(2): 272–92.

Zajacova, Anna, Jennifer Karas Montez, and Pamela Herd. 2014. "Socioeconomic Disparities in Health Among Older Adults and the Implications for the Retirement Age Debate: A Brief Report." *Journals of Gerontology: Series B* 69(6): 973–78.

PART V

Hiring

Network Recruitment and the Glass Ceiling: Evidence from Two Firms

ROBERTO M. FERNANDEZ AND BRIAN RUBINEAU

Does network recruitment contribute to the glass ceiling? We use administrative data from two companies to answer the question. In the presence of gender homophily, recruitment through employee referrals can disadvantage women when an old boys' network is in place. We calculate the segregating effects of network recruitment across multiple job levels in the two firms. If network recruitment is a factor, the segregating impact should disadvantage women more at higher levels. We find this pattern, but also find that network recruitment is a desegregating force overall. It promotes women's representation strongly at all levels, but less so at higher levels. This article shows how administrative data can be used to tackle the complex problem of gender inequality in organizations to counter the glass ceiling.

Keywords: glass ceiling, gender segregation, network recruitment, administrative data

Many studies have adopted the glass ceiling metaphor to describe the phenomenon where gender inequality in outcomes is more severe at the top of the reward distribution. Although many other factors are theorized to be at play, some recent research has sought to understand the external sources of the glass ceiling pattern (Gorman and Kmec 2009; Fernandez and Abraham 2010, 2011). One factor that is often invoked to account for gender differences in outcomes is firms' reliance on hiring through social networks. Numerous studies have documented the segregated nature of social networks (see, for example, Brass 1985; Campbell 1988; Ibarra 1992; Lincoln and Miller 1979; Marsden 1987, 1988; Moore 1990; Straits 1996). This line of reasoning states that in the presence of gender homophily, recruitment through employee referrals is likely to further disadvantage women in hiring. This logic is often connected to the glass ceiling by arguing that recruitment to higher levels of the organization is likely dominated by an old boys' network. Even in the absence of gender biases, homophily among successful men can effectively restrict access to the upper echelons of an organization in a man-

Roberto M. Fernandez is William F. Pounds Professor in Management and professor of organization studies at the MIT Sloan School of Management. **Brian Rubineau** is associate professor of organizational behavior at the Desautels Faculty of Management at McGill University.

 Fernandez, Roberto M., and Brian Rubineau. 2019. "Network Recruitment and the Glass Ceiling: Evidence from Two Firms." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(3): 88–102. DOI: 10.7758/RSF.2019.5.3.05. Direct correspondence to: Roberto M. Fernandez at robertof@mit.edu, MIT Sloan School of Management, Massachusetts Institute of Technology, Room E62–387, 100 Main St., Cambridge, MA 02142.

ner that obstructs women's success. The logic of this argument is clear. But is there empirical evidence for its presence and operation?

In this article, we look for evidence for whether network recruitment contributes to a glass ceiling phenomenon using administrative data from two distinct organizations. Although politically quite sensitive, the data required for this exercise is commonly available at most mid- to large-size organizations. Prior scholarship has documented a method for measuring the segregating effects of network recruitment (Rubineau and Fernandez 2015). Organizations with referral bonuses commonly collect the application, hiring, and referring data needed to use this method. The authors have worked in collaboration with two large pharmaceutical organizations to identify and collect relevant data to help these firms in pursuit of their gender equity goals. We leverage their existing human resources (HR) data to measure the segregating effects of network recruitment. Specifically, we estimate these segregating effects at multiple organizational levels in each of the two companies. We examine whether these estimated segregating effects are consistent with the predicted glass ceiling pattern and discuss the practical and theoretical implications of our findings.

We first bring empirical evidence to the prior assertion-based arguments regarding a possible role of network processes in contributing to a glass ceiling effect. In doing so, we demonstrate how existing private-sector organizational HR data resources can be leveraged to understand and manage the segregating effects of network recruitment. The article thus serves as a powerful example of how administrative data—the residuum of modern bureaucratic processes (Reardon 2019)—can be harvested and put to use in addressing important science and policy questions.

Both firms studied yield the same paradoxical results. First, we find evidence of a clear glass ceiling pattern: network recruitment produces a decreasing percentage of women as one looks across levels within the firms' hierarchies. Second, we find that network recruitment tends to act as a desegregating force. That is, network recruitment tends to push firms to become more female than they would be without network recruitment. Although network recruitment is a desegregating force overall, it promotes women's representation more strongly at lower levels within the firms, and less strongly at higher levels.

THE GLASS CEILING AND NETWORK RECRUITMENT

The term *glass ceiling* has become a general metaphor applied to the topic of gender inequality in the workplace and in organizational governance (Hymowitz and Schellhardt 1986). Put simply, the glass ceiling is meant to imply that women encounter invisible barriers when attempting to move up corporate hierarchies. Women may be able to enter the workforce with little resistance, and even attain near- or beyond-parity representation in lower-level jobs. But when a glass ceiling is in place, women seeking positions of leadership, power, and status in their organizations face a gendered obstacle. Although coined more than thirty years ago, the metaphor has retained its currency in the face of women's enduring relative absence in high positions in corporations (Cotter et al. 2001; Glass and Cook 2016; Tinsley et al. 2017; Yap and Konrad 2009).

The Glass Ceiling and the Labor Market

Much scholarship has sought to identify the mechanisms that produce this gendered pattern of hierarchical advantage (see Petersen and Saporta 2004). Although many glass ceiling studies invoke the image of an internal labor market or a within-firm job ladder, recent work argues that the pattern applies to the labor market more broadly (Fernandez and Campero 2017; Fernandez-Mateo and Fernandez 2016). In this work, the glass ceiling is seen as a vertical form of more general gendered job segregation patterns (Fernandez and Campero 2017; Fernandez and Abraham 2010, 2011; Fernandez-Mateo 2009). In this respect, insight into the phenomenon can be gained by drawing on the rich research addressing job sex segregation (for example, Bielby and Baron 1986; Jacobs 1989). Some of these mechanisms are posited to operate on the demand side of the labor market (Cejka and Eagly 1999; Kmec 2005). Others argue that supply-side processes are also likely to contribute to job sex segregation (Correll 2001; Fer-

nandez and Friedrich 2011; Barbulescu and Bidwell 2013; Fernandez and Campero 2017).

Whatever the sources of this segregation, addressing the enduring inequality in gendered organizational attainment requires revealing and resolving the mechanisms that generate it (Reskin 2003). Complicating the task are the many mechanisms interacting simultaneously and working toward preserving the inequality (Ridgeway 1997). This article seeks to contribute to the scholarly understanding of the glass ceiling phenomenon by making a detailed empirical examination of one commonly discussed mechanism seen as contributing to the glass ceiling—network recruitment.

Network Recruitment and the Old Boys' Network

One of the factors often invoked to account for gender differences in labor market outcomes is firms' reliance on hiring through social networks. Network recruitment continues to be both a dominant and an effective recruitment method (Breaugh 2014; Marsden and Gorman 2001). In relying on social networks, however, employers inherit the biases within them. Numerous studies have documented the segregated nature of social networks generally (Brass 1985; Campbell 1988; Ibarra 1992; Lincoln and Miller 1979; Marsden 1987, 1988; Moore 1990; Straits 1996). Just as many have documented the gender segregation of workplace networks specifically (Belle, Smith-Doerr, and O'Brien 2014; Fernandez and Sosa 2005; Kleinbaum, Stuart, and Tushman 2013; Lawrence 2006). Because of gender homophily in workers' networks—the tendency for networks to be same-sex—network recruitment has been argued to disadvantage women in hiring (Acker 2009; Beaman, Keleher, and Magruder 2018; Leicht and Marx 1997; McDonald 2011; Reskin, McBrier, and Kmec 1999).

The expected gender-segregating effects of network recruitment sometimes explicitly invokes the old boys' network (see, for example, McDonald 2011, 317). Despite the strong conventional wisdom that network recruitment would be expected to have segregating effects, empirical studies seeking to evaluate these effects have sometimes yielded contrary findings. For example, a study of hospitals in the northwestern United States finds that organizations that engaged in higher levels of network recruitment exhibited lower levels of gender segregation (Kmec 2008). In another example, a big-data study examining the segregating effects of network recruitment in Stockholm finds that increased use of network recruitment was associated with a decrease in the level of gender segregation in the labor market (Collet, Hedström, and Johansson 2014). These and a variety of other empirical findings challenge the idea that network recruitment necessarily perpetuates or exacerbates segregation (see Rubineau and Fernandez 2015).

To better understand the segregating effects of network recruitment, we developed a line of research-building theory regarding the dynamics relating network recruitment to segregation. This effort yielded two articles examining how network recruitment affects hiring outcomes (Rubineau and Fernandez 2013, 2015). The articles show, first, that referrers play an important role in determining the segregating effects of network recruitment (2013), and, second, that measures taken from HR data involving job applications and referring ties can reveal the equilibrium composition toward which network recruitment will push a job's gender composition (2015). This equilibrium indicates the segregating effects of network recruitment in isolation—network recruitment's push toward or pull from segregation. The key determinants are not the initial composition of the job or firm but instead the gender differences in two referring-related rates: one at which employees refer generally and another at which referring employees specifically refer same-sex contacts to apply to the firm. In the presence of homophilous referring, if men and women differ in their rates of producing referral applicants, then, ceteris paribus, the equilibrium gender composition from referring will favor the group that produces more referral applicants. Similarly, if the men and women who refer differ in their rates of referring same-sex contacts, then the equilibrium gender composition from network recruitment will favor the group that generates a higher proportion of same-sex referral applicants.

In addition to providing analytical bite on the general issue of network recruitment, these articles yield important practical implications

for firms seeking to affect the gender compositions of their workforce. The method requires that the firm follow formalized HR practices and data recording protocols for both the hiring process, beginning with job applications, and the referral bonus administration that links current employee referrers with their referral applicants (Rubineau and Fernandez 2015). Although rarely systematically analyzed, most mid-sized to large firms routinely track the data required to shed light on how network recruitment is affecting their gender composition.[1] Given enough observations, the segregating effects of network recruitment could be calculated not only on a firm level, but also subdivided into useful subsets. In this article, we look at the segregating effects of network recruitment for lower-level jobs within the firm relative to those for higher-level jobs to examine whether and to what extent network recruitment may contribute to a glass ceiling effect. By applying this method, we can test for evidence that network recruitment contributes to the glass ceiling phenomenon.

Specifying Expectations: A Glass Ceiling Pattern

If network recruitment does contribute to a glass ceiling effect in organizations, how would the contribution be manifest and how could it be documented? The method just described enables the calculation of the segregating effects of network recruitment in terms of the equilibrium composition of women toward which network recruitment pushes the current incumbent composition.

If a glass ceiling effect derives in part from network recruitment dynamics, then we would expect a specific pattern in the relative calculated equilibrium compositions of women produced by network recruitment at different ranks within a firm. More specifically, the equilibrium composition of women produced by network recruitment operating at higher levels of the firm would be expected to be lower than that at lower levels of the firm. By measuring the segregating effects of network recruitment for multiple job levels within a single firm, we can evaluate whether network recruitment dynamics contribute the expected glass ceiling pattern.

We undertake this analysis for two organizations and in doing so attempt an empirical existence proof. Evidence that network recruitment yields the expected glass ceiling pattern would provide empirical support allowing us to rule-in the claim that this is one of the many mechanisms contributing to a glass ceiling. Absence of such evidence could not be taken as evidence that this mechanism never contributes to a glass ceiling. Instead, such absence of evidence would indicate that although network recruitment could potentially contribute to a glass ceiling in other contexts, the current empirical contexts would not provide unambiguous evidence for such a contributory role. Although recognizing the logical status of the evidence presented here is important for scientific progress in this domain, it is also important to understand that from the perspective of the cooperating firms, these analyses are inherently of interest for furthering their own understanding of how their policies around network recruitment contribute—or not—in their pursuit of gender equity in their organizations.

DATA AND METHODS

We approached the companies studied here through professional connections developed in the course of our teaching business classes at our respective universities. In both cases, the contacts suggested that their companies were interested in deeper understanding of issues of gender diversity. After a period of discussion with the firms' management, we presented a research plan and developed a mutually acceptable confidentiality and nondisclosure agreement. We then worked with company personnel to identify the relevant databases and to match and merge these various sources using anonymized data.

The estimation of the segregating effects of network recruitment requires a specific set of HR data regarding organizational recruitment and referring. These data requirements are summarized here (for a full description, see Rubineau and Fernandez 2015). As numerous other articles in this volume demonstrate, when dealing with administrative records, mul-

1. In the United States, firms of fifty or more employees are legally required to maintain such records.

tiple data sources often need to be matched and combined to triangulate on the information of interest (for example, Penner et al. 2019). This is true in our case as well. The required data for these analyses are coming from several disparate databases that tend to be maintained by different organizational subunits, and in our private-sector cases, these subunits are housed under the umbrella of the HR department of these organizations.[2]

First, to characterize the hiring pipeline into the firm, the method requires data on all job candidates' gender, not simply the gender of the people hired (for a discussion of "start with hire" analyses, see Fernandez and Weinberg 1997). This is because relative to post-hire data, the prehire pool more accurately reflects the supply-side biases affecting nonreferral applicants' decisions to apply for a defined job, as well as the referring behavior of employee referrers.

Second, we need to know the referral status for each candidate. This information is typically included in a database produced by applicant tracking software managed by the recruitment arm of the human resources department. However, for each of the referral candidates, we also need to know the identity of their corresponding employee referrer so as to identify the referrer's gender and level in the organization. In our cases, information on referrers is generated by an HR team that administers the firms' employee referral bonus programs.

In addition to data on the hiring pipeline, the method also requires data for current employees and employee referrers. Specifically, we need to know all employees' gender. These are contained in the companies' human resources information system (HRIS). We combine the HRIS data with information from the employee referral bonus program to identify whether employees have attempted to refer someone; for those who have, we identify their corresponding referral applicants in the applicant tracking system.

To provide better understanding of the gender patterns across the organizational hierarchy, this analysis also requires grouping these data elements by job level within the firm. For these reasons, we also need to identify the job levels of the job opportunity for which applicants apply, and the job levels of all current employees. These too are taken from the companies' HRIS databases.

As described, we identified two organizations (Firm 1 and Firm 2) willing to cooperate and provide the needed data elements. Both are large U.S. pharmaceutical industry firms and provided data on their job recruitment efforts and referral bonus program data for the 2014 calendar year. The choice of these firms was made based primarily on our ability to establish data-sharing and collaborative research relationships with the firms. We neither claim that pharmaceutical firms are particularly likely or unlikely to exhibit a glass ceiling phenomenon, nor would such a claim be required for conducting the current analysis.

The jobs in each of these two organizations are divided into four groups of job levels. Contacts at Firm 1 provided the needed data already grouped by job-level groups based on their understanding of meaningful division lines among job levels. Firm 2 provided detailed job-level data, and the researchers divided these observations into four groups based on a combination of job categories (for example, nonsalaried nonexempt jobs, hourly, and unionized jobs, as level 1, and salaried exempt jobs as progressively higher levels), and the sizes of the specific job levels to allow for approximately similarly sized job level groups. The job-level groupings for the Firm 2 data were based on the job level of the referrer alone, as the job applicant data did not allow for clear matching of job levels. We discuss the implications of this assumption later. For Firm 2, the gender of job applicants was coded probabilistically by comparing applicants' first names against a database of first names used by people in the United States and their probability of being associated with men and women as provided by Genderize.io.[3] Table 1 provides a summary of the nature of the recruiting and referring processes for these two organizations.

2. In theory, it is possible for the data for these disparate processes to be contained in a single relational database. In practice, however, we have yet to encounter examples of such integrated end-to-end HR databases.

3. See "Determine the gender of a first name," https://genderize.io/#overview (accessed October 15, 2018).

Table 1. HR Recruitment and Referring Data

	Level 1		Level 2		Level 3		Level 4		Whole Firm	
	Count	Percent Female	Count	Percent Female	Count	Percent Female	Count	Percent Female	Count	Percent Female
Firm 1										
External applicants	17,419	48	33,089	50	36,802	39	6,690	32	94,000	43.7
External hires	187	51	286	58	426	41	136	26	1,035	45.7
Referral applicants	677	57	2,568	59	2,816	45	264	33	6,325	46.7
Employees	162	49	2,455	58	3,826	49	773	41	7,216	51.5
Referrers	33	79	479	66	599	56	65	51	1,176	60.5
Firm 2										
External applicants	—	—	—	—	—	—	—	—	60,893	38.2
External hires	155	34	335	49	336	52	268	50	1,094	47.8
Referral applicants	806	52	874	53	1,545	50	615	48	3,840	50.6
Employees	278	40	1,832	46	3,721	53	3,893	42	9,724	47.0
Referrers	98	58	247	48	634	54	311	53	1,290	53.0

Source: Authors' tabulation.

Table 2. Measures and Calculations for Equilibrium Gender Composition from Network Recruiting

	Level 1	Level 2	Level 3	Level 4	All
Firm 1					
A: Proportion male among applicants referred by men	0.46	0.42	0.52	0.30	0.50
B: Proportion female among applicants referred by women	0.61	0.61	0.49	0.36	0.57
C: Proportion referrers among men at firm	0.09	0.16	0.14	0.08	0.13
D: Proportion referrers among women at firm	0.31	0.22	0.18	0.12	0.19
f: Proportion female of nonreferral applicants	0.48	0.50	0.39	0.32	0.42
s: Supply-side effects parameter = *f/(1 – f)*	0.92	0.99	0.63	0.46	0.72
m: Men's same-sex referring rate = *As/(1–A+As)*	0.44	0.42	0.41	0.16	0.42
w: Women's same-sex referring rate = *B/(s–Bs+B)*	0.63	0.61	0.60	0.55	0.65
d: Demand-side effects parameter = *D/C*	3.38	1.40	1.31	1.47	1.44
Equilibrium percent female	60	60	49	48	55
Incumbents percent female	49	58	49	41	52
Percentage-point change from inputs (*f*) to equilibrium	12	10	10	16	13
Hiring preference equivalent: female hiring odds	1.27	1.22	1.23	1.40	1.30
Firm 2					
A: Proportion male among applicants referred by men	0.58	0.51	0.59	0.58	0.56
B: Proportion female among applicants referred by women	0.61	0.58	0.57	0.52	0.57
C: Proportion referrers among men at firm	0.25	0.13	0.17	0.07	0.09
D: Proportion referrers among women at firm	0.51	0.14	0.17	0.10	0.12
f [a]*:* Proportion female of nonreferral applicants	0.40	0.46	0.53	0.42	0.39
s: Supply-side effects parameter = *f/(1 – f)*	0.66	0.84	1.13	0.73	0.64
m: Men's same-sex referring rate = *As/(1–A+As)*	0.47	0.47	0.62	0.51	0.45
w: Women's same-sex referring rate = *B/(s–Bs+B)*	0.70	0.62	0.54	0.60	0.67
d: Demand-side effects parameter = *D/C*	2.09	1.08	1.05	1.55	1.23
Equilibrium percent female[a]	55	54	49	47	51
f [b]*:* Proportion female of nonreferral applicants	0.39	0.39	0.39	0.39	0.39
Equilibrium percent female[b]	55	54	49	47	51
Incumbents percent female	40	46	53	42	47
Percentage-point change from input (*f* [a]) to equilibrium	15	8	-4	5	12
Hiring preference equivalent: female hiring odds	1.35	1.17	0.92	1.11	1.28

Source: Authors' tabulations.

[a] Percentage female among nonreferral by level was not available – only the percentage female among all nonreferral applicants. Calculations are based on percentage female among incumbents at that level.

[b] Calculations are based upon percent female among all nonreferral applicants.

RESULTS

The variables estimated from the two datasets are described as follows (letter designations from Rubineau and Fernandez 2015). Their estimates for the four job levels in the two organizations are presented in table 2.

(A) Percent male among applicants referred by men

(B) Percent female among applicants referred by women

(C) Percent referrers among men at firm

(D) Percent referrers among women at firm

(f) Percent female of nonreferral applicants

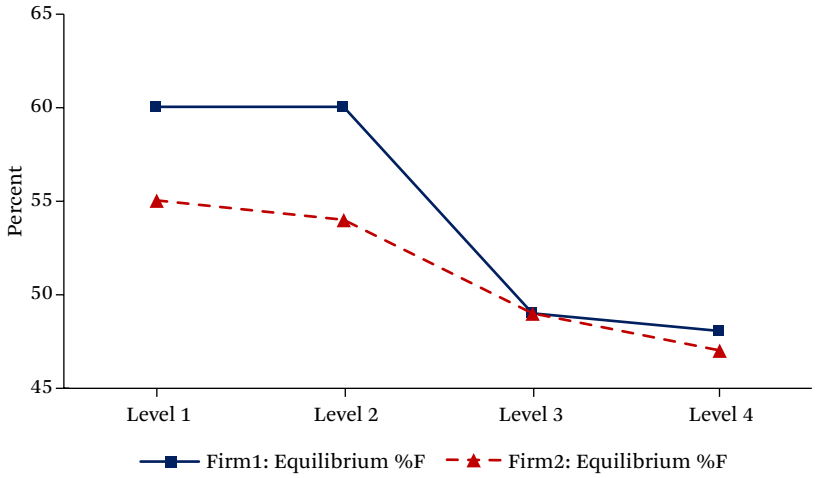
(s) Supply-side effects parameter = $f/(1 - f)$

(d) Demand-side effects parameter = D/C

(m) Men's same-sex referring rate = $As/(1 - A + As)$

(w) Women's same-sex referring rate = $B/(s - Bs + B)$

Figure 1. Comparing Equilibrium Gender Compositions Produced by Network Recruiting



Source: Authors' tabulations.

The first five variables are estimated from the data provided by the two firms, and the next four terms are calculated using the previous five. These last four terms (*s, d, m, w*) are entered into the following equation (Rubineau and Fernandez 2015, 1654). The equation gives the equilibrium percent female toward which network recruitment is pushing the composition of incumbents.

$$f^{*} = (ws - m + \sqrt{w^2s^2 + 2mws + m^2 + 4s - 4ws - 4ms}) \cdot (2 - 2w + ws - m + \sqrt{w^2s^2 + 2mws + m^2 + 4s - 4ws - 4ms})^{-1}.$$

The last column of table 2 shows the aggregated results (without respect to organizational level) for the two firms as a whole. Overall, both firms are near gender parity: females in Firm 1 are 52 percent of incumbent employees; Firm 2 is 47 percent female. Interestingly, by itself, network recruitment dynamics are pushing both firms to be even more female than they already are. For Firm 1, the equilibrium percent female is 55 percent (versus 52 percent of incumbents), and at Firm 2 the equilibrium is 51 percent (47 percent of incumbents).

Equilibriums Across Job Levels

We next look at how these dynamics play out at each level of these organizations. The columns of table 2 provide the results of these calculations along with the gender composition of incumbents in each of the four job levels in the two firms. For Firm 2, the *f* term—percent female among nonreferral applicants—was not available separately for each of the job levels. This percentage was available only for the entire applicant pool (39 percent). Because of this limitation, as a type of sensitivity check, we calculated two sets of equilibriums for the four job levels of Firm 2. One set used the 39 percent value for all four job levels as the *f* term. A second set used the current composition of incumbents. These two sets of calculations yielded almost identical values, showing that the sensitivity of the equilibriums to the value of this variable is relatively low.

Table 2 shows that for both Firm 1 and Firm 2, we find the same paradoxical result. First, we do find evidence of a strikingly clear glass ceiling pattern: the effects of network recruitment are a decreasing percentage of women with increasing levels within the firms' hierarchies. In Firm 1, network recruitment pushes the lowest-level jobs toward a 60 percent female composition, but the highest-level jobs toward a below-parity 48 percent female composition. The two job levels between these ends also have intermediate equilibrium compositions, resulting in a monotonic decline in the gender integration effects of network recruitment. This monotonic decline is also illustrated in the solid line in figure 1.

In Firm 2, we see an almost identical pattern.

Network recruitment pushes the lowest-level jobs toward a 55 percent female composition, but the highest-level jobs toward a below-parity 47 percent. The trend across the four job levels in Firm 2 is also a monotonic decline, illustrated graphically in the dashed line in figure 1. Results from both firms examined are consistent in exhibiting the pattern of results that would be expected if network recruitment contributed to a glass ceiling effect.

These glass-ceiling-consistent results do not mean, however, that network recruitment is acting as a segregating force at every level of the two firms. By comparing the calculated equilibrium compositions with the composition of incumbents, it is clear that network recruitment tends to act as an integrating force even within each level. As shown in the final rows of the two panels of table 2, for all job levels except level 3 (coincidentally for both firms), incumbents are less female than the equilibrium entailed by network recruitment. This result means that network recruitment acts to make these jobs more female than they would be otherwise. This integrating effect of network recruitment is strongest for the lowest job levels and weaker for the higher job levels. In this way, we have network recruitment exhibiting the curious behavior of simultaneously contributing to the gender integration of the firms but also contributing to the glass ceiling patterns in both firms.

Segregating Effects of Network Recruitment

The equilibriums in table 2 illustrate not the projected future composition of each job level for each firm, but instead the conceptual target composition toward which network recruitment—in isolation from other composition-altering mechanisms—is pushing those jobs. That is, if the HR processes in the firms could reach a state at which the only mechanisms affecting the compositions of the job levels were the rate and gender composition of nonreferral applicants, network recruitment, and otherwise gender-unbiased hiring rate and exit rates, then the equilibriums would be the projected future composition of each job level. This hypothetical state is not an actual goal but a useful conceptual baseline to illustrate the segregating effects of network recruitment.

The calculated equilibriums themselves are not the sole indicator of the segregating effects of network recruitment. If network recruitment were pushing two sets of jobs both toward a 60 percent female equilibrium composition, and the first set had an input composition of 30 percent female, and the second set a composition of 55 percent, we would know that the network recruitment is exerting a larger push on the composition of the first set of jobs than on the second set. The direction and magnitude of the difference between the input composition and the calculated equilibrium composition from network recruitment serves as one way to quantify the composition-changing effects of network recruitment.[4] Table 2 provides these percentage-point difference measures for the four job levels and the overall firm for both firms.

We also provide a second, and possibly more easily interpreted, measure of the composition-changing effects of network recruitment: the magnitude of another well-understood segregating mechanism—an explicit bias in screening applicants—that would be needed to yield the equivalent compositional changes as produced by network recruitment (first described in Rubineau and Fernandez 2013). Using the examples from the previous paragraph, how much of an explicit preference for hiring women over men would be needed to move one set of jobs with a steady input composition—30 percent female—to yield a final stable job composition that is 60 percent female, and how much of an explicit preference would be needed to move another set of jobs from 55 percent female to 60 percent?

We provide this answer by solving a probability problem. If the entering composition of applicants is p (in this example, 30 percent female or 55 percent), what odds for choosing a female for hire are needed to yield the target equilibrium composition (in this example, 60

4. For the input composition, we use the percent female of nonreferral applicants. In the hypothetical case where all other composition-altering mechanisms were removed, the equilibrium gender composition of the firm would necessarily be the composition of its inputs—nonreferral job applicants.

percent female)? When the odds are 1, then the equilibrium composition necessarily equals the input composition. Greater than 1 odds for choosing a female for hire will push the composition more female than the input composition, and below 1 odds for choosing a female for hire will push the composition more male. The two odds in the example are 1.11, and 1.87, respectively.

We solve this problem for the four sets of job levels as well as for all jobs in the two firms. The calculated odds shown in table 2 represent the odds for choosing a female for hire necessary to push the input compositions (measures *f* in the table) to the equilibrium compositions from network recruitment (indicated in the table in bold). This odds measure is a useful summary metric for the strength of the segregating effects of network recruitment. The magnitudes of the odds shown in table 2, ranging from 0.92 to 1.40, are substantial. These substantial pushes to the composition of jobs in the two firms illustrate the power of network recruitment to alter the gender compositions of jobs in firms. This power can also be wielded as a tool for achieving organizational diversity goals. Our analysis of these administrative hiring and referring data provide additional insights regarding the likely effectiveness of using referring to achieve organizational diversity goals.[5]

Effects of Different Referring Behaviors

Additional insights can be gleaned from close scrutiny of the terms in table 2 for the two organizations. With respect to policy choices, which can affect the equilibrium toward which network recruitment is pushing the firm, two basic levers are available. First, firms can seek to encourage more referring from the underrepresented group relative to the overrepresented group. In terms of the parameters, this implies that D (the referring rate for females) should be made greater than C (the referring rate for males). Second, firms can encourage referrers to target the underrepresented group in their referring.[6] In these cases, this corresponds to having males produce relatively more female referrals (a *low* level of A), and have females produce a high level of female referrals (a *high* level of B).

Comparisons between Firm 1 and Firm 2 at low and high levels help reveal the relative strength of these levers. At the lowest level of both firms, the first lever is clearly active: D is substantially greater than C. This contributes to pushing the equilibrium composition of females at the lowest level to be substantially higher than the current percent female among incumbents. In contrast, at higher organizational levels, the gap between D and C is much smaller, and thus cannot serve to differentiate these two cases.

At the high end of Firm 1—but not Firm 2—we see evidence of the second lever at work: the majority of referrals produced by males are female (A is < 0.5), and the majority of referrals produced by females, are also female (B is > 0.5). Although the contrast between the two firms in these rates is striking, the resulting equilibriums are not at all different across the two firms, and not as high as the equilibriums for the lower-level jobs. From the percentage-point change from input to equilibrium, and from the equivalent odds measure of the strength of the compositional push from network recruitment, it is clear that this second lever can produce substantial changes. These changes, however, cannot overcome the very low composition of the inputs, and the net effect remains that higher-level jobs are pushed toward a composition that is less female than lower-level jobs. This suggests that at least in these cases, lever 1—getting the underrepresented group to refer more—may be the more reliably beneficial approach

5. These integrating processes are from the firms' behaviors affecting the flow of candidates into the firm, and not from any screening preference. Indeed, as discussed in Rubineau and Fernandez (2013, 2015), this is one of the appealing features of using network recruitment as an integrating lever. In an era when firms are shy about implementing affirmative action as hiring preferences, "pipeline stoking" processes are particularly attractive.

6. We have come across two examples where companies have sought to encourage this second lever in their diversity efforts. Pinterest and Intel both have instituted higher referral bonuses whenever people refer a woman or a minority than when they refer a white male. Interestingly, we had judged the use of this kind of targeted referring as ethically and legally questionable (Rubineau and Fernandez 2015).

when using network recruitment to ameliorate the glass ceiling.

DISCUSSION AND IMPLICATIONS

Although network recruitment processes have often been invoked as a contributor to the glass ceiling phenomenon, little empirical evidence has been presented to demonstrate such an association. Prior to this project, the relationship has largely been putative. This article leverages detailed HR data regarding recruitment and referring processes in two firms to empirically evaluate whether network recruitment contributes to a glass ceiling effect in the two firms.

This evaluation process first calculates the segregating effects of network recruitment for the firm overall. We then divide the data from each firm into four job-level categories, representing increasingly elite echelons within each firm. If network recruitment were to contribute to a glass ceiling phenomenon, then the calculated effects of network recruitment would be a declining gender composition at higher job levels within the firm. We find precisely and unambiguously this pattern in both firms.

Grouping the Data by Job Level: Approaches and Implications

Grouping data from the two firms by job level was accomplished in potentially different ways. Job referrers are not necessarily at the same job level as that of the job opportunity to which they refer. Grouping firm referring data into job levels is done in one of three ways. One approach is to limit referrers and referral applicants who are referring from and applying to the same job level. A second approach is to define the job level by that of the referrer. A third approach is to define the job level by that of the job to which the referral applicant applies. Firm 1 calculated the parameters we requested, but we do not know for certain which of the three approaches was used. For Firm 2, we necessarily determined job level by the job level of the referrer—the second approach. We examine the implications of these approaches.

Might the job grouping approach affect the resulting pattern we observe? Although we cannot test the assumption using our data directly, we consider its possible implications. First, extensive scholarship shows that people tend to refer homophilously, including on characteristics such as educational background, professional behavior, job-related aptitude, and corporate division (Fernandez, Castilla, and Moore 2000; Rider 2012; Burks et al. 2015; Hensvik and Skans 2016; Brown, Setren, and Topa 2016). Also, our job-level categories are broad. For these reasons, we do expect the cases where referrer and applicant are at the same broad job-level category to have the most observations.

Second, both theoretical and empirical network recruitment scholarship supports the idea that referrers tend to refer for job opportunities at their own level or lower more than for job opportunities at a higher level than the referrer (Leicht and Marx 1997; Brown, Setren, and Topa 2016). Building on this finding, we examine the likely effects of the grouping-by-referrer and grouping-by-applicant approaches. The grouping-by-referrer approach to define the job level (as used in the Firm 2 data) should be most precise for the lower job levels, and most noisy for higher job levels. This expectation comes from the likelihood that referrers at the lowest job level rarely refer applications for jobs at higher levels. Conversely, the most noisy parameter calculations would be for the higher job levels as referrers at the highest job level could refer applicants for all levels in the firm. What would the noisier parameter calculations at the highest level mean for our study? It would mean that the calculations for the highest job levels would be biased toward the firm-wide outcome rather than the job-level-specific outcome. And what would this bias mean for our results? This bias would mean that our observed pattern of a declining percentage female from network recruitment is conservative, and the true pattern would show a steeper negative slope. This steeper negative slope would be the only way we could observe our results in the context of mean-biased values for the higher job levels. That is, consider the lowest job-level calculation fixed (if it is the most precise), and that the highest job-level calculation has been pushed toward the firm mean through the assumptions underlying our calculations. Because the observed slope is negative, the only way the higher job-level values could have been pushed toward the firm mean would be if their true values were even lower than the current

calculations—that is, a more steeply negative slope. This grouping-by-referrer was the case for Firm 2. If this method was also used for Firm 1, the pair of results would be similarly conservative.

Conversely, if Firm 1 used a grouping-by-applicant approach, then the highest job-level calculations would be the most precise because referral applicants to the highest job levels would be likely to receive those referrals only from others in the highest job levels, and least likely from those at lower job level). The lowest job-level calculations would be pushed toward the overall firm mean because the applicants at the lowest levels could have been referred by referrers at any level. In this case, the highest job-level calculations could be considered fixed, and the calculations of the lowest job levels would have been pushed toward the firm mean. Because our negative sloping pattern for Firm 1 has the highest job levels yielding the lowest percentage female equilibrium levels, the only way the lowest job levels could be at their calculated levels and biased toward the firm mean would be if the actual equilibrium for the lowest job levels were higher than what was calculated. Again, this would mean that the true slope would be more negative than the calculated trend for Firm 1.

We know the assumption underlying the parameter calculations for Firm 2. As we argue, we have reason to believe that the observed negative slopes are conservative estimates. This conclusion, plus the fact that we observe a similar negative slope with our data from Firm 1 both lend strength to our conclusion that network recruitment yields a glass ceiling pattern.

Theoretical and Practical Implications

The two goals were to bring empirical evidence to the prior assertion-based arguments regarding a possible role of network processes in contributing to a glass ceiling effect, and to demonstrate how existing organizational HR data resources can be leveraged to understand and manage the segregating effects of network recruitment. These two goals map well onto the theoretical and practical implications of this article.

In theoretical terms, we provide empirical evidence that network recruitment can contribute to the glass ceiling. But the nature of this evidence also challenges and changes the conventional wisdom explaining why network recruitment contributes to the glass ceiling. Network recruitment does not contribute to the glass ceiling simply because of homophily and the old boys' network. Instead, network recruitment contributes to the glass ceiling because of gendered differences in the rates of who refers, and gendered differences in the rates in which referrers refer their same-sex contacts (on these dynamics, see Rubineau and Fernandez 2015). In the two firms we examined, the lowest-level jobs had the highest gender differences in rates of referring—both heavily favoring women engaging in referring. These differences contributed greatly to the strong integrating effects of network recruitment for the lowest-level jobs. These system-based insights into the segregating effects of network recruitment also have important practical implications for managing these effects.

In practical terms, we show that the HR data commonly collected by large firms in administering their recruitment efforts and referral bonus programs is enough to reveal the segregating effects of network recruitment not only for the firm overall, but even for narrower subsets of jobs within the firm. We show that the data need not be perfect or complete to be able to make these calculations. In the absence of detailed data, reasonable approximations along with sensitivity analyses can reveal whether the results would be sensitive to the more detailed data. Thus, firms seeking to evaluate the effects of their network recruitment practices on their goals for diversity can readily do so.

These methods can even help leverage network recruitment as a tool for improving and targeting diversity outcomes. For example, the current finding is that network recruitment acts as an integrating force, but more strongly for lower-level jobs. Without intervening in the network recruitment processes for higher-level jobs, the expectation may be a gradual gender integration of the firm. This integration, however, would be more substantial for lower-level jobs and would be unlikely to ameliorate the gendered barriers apparent for higher-level jobs. Moreover, our analyses suggest that policies designed to encourage the underrepre-

sented group to participate in referring are more likely to harness the integrating powers of network recruitment than are policies that seek to specifically target the underrepresented group.

We conclude with confidence that the pattern of network recruitment does contribute to a glass ceiling effect in these two firms. As noted, our empirical setting is not representative of any larger set of firms, so our conclusion does not imply such a pattern in other firms. The firms selected were a kind of convenience sample of firms collecting and willing to share their recruitment and referral data. They were not identified based on their being more or less likely to exhibit a glass ceiling phenomenon or to have such a phenomenon associated with their referring practices. Still, given that ours is the first study to empirically assess whether network recruitment contributes to the glass ceiling, we find it striking that we find strong and consistent results in the first two settings we examine. This result could be due to chance. It could also suggest that network recruitment commonly contributes to the glass ceiling. Determining which of these two explanations is more correct requires replicating this study in more organizational settings.

Regardless of whether these results generalize to other settings, the cooperating firms find clear value in assessing whether and how network recruitment dynamics contribute to gender disparities across levels of their organizations. As the firms craft their policies on these issues, that these analyses correspond to their circumstances is an appealing rather than a limiting feature. In their cases, one size needn't fit all when developing their policy prescriptions.

These analyses isolate and clarify the contribution of recruitment network dynamics affecting the glass ceiling. However, other biases and composition-altering mechanisms likely operate in real-world recruitment processes. Although the calculated equilibriums for network recruitment effects represent the segregating effects of network recruitment in isolation, we cannot claim that these affects are a complete explanation for why we so often find glass ceiling patterns in organizations. We can, however, offer some general guidance about the relative importance of such network recruitment factors.

Although other forces may affect a given firm's recruitment process, we can say that, for firms that do not rely heavily on network recruitment, the segregating effects of network recruitment are expected to be very weak. At the same time, for firms that do rely heavily on network recruitment, the segregating effects of this mode of recruitment are expected to be stronger. To build intuition about the relative magnitude of these network recruitment factors, we have developed a way of representing the size of these effects in terms of one of those other biasing factors, that is, sex biases in screening (see Rubineau and Fernandez 2013). The calculated equilibriums we present for network recruitment in these two firms provides a measure of the direction and magnitude of the composition-altering effects of network recruitment—the segregating or desegregating push or pull of network recruitment. The magnitude of the network factors we uncovered in these two settings are equivalent to an overall preference for female applicants that resembles a 56:44 or 57:43 chance (the equivalent of odds of 1.28 to 1.30) for hiring women relative to men, rather than a gender-neutral 50:50 chance.

In conclusion, although other factors may be at work at producing the glass ceiling, at least in these settings, network recruitment dynamics are a meaningful part of their stories. Revealing these network recruitment dynamics can be an important step toward managing these processes to support and achieve firms' diversity goals. These insights are made possible only by partnering with these firms to gain access to these commonly available but rarely analyzed administrative HR data.

REFERENCES

Acker, Joan. 2009. "From Glass Ceiling to Inequality Regimes." *Sociologie du travail* 51(2): 199–217.

Barbulescu, Roxanna, and Matthew Bidwell. 2013. "Do Women Choose Different Jobs from Men? Mechanisms of Application Segregation in the Market for Managerial Workers." *Organization Science* 24(3): 737–56.

Beaman, Lori, Niall Keleher, and Jeremy Magruder. 2018. "Do Job Networks Disadvantage Women? Evidence from a Recruitment Experiment in

Malawi." *Journal of Labor Economics* 36(1): 121–57.

Belle, Deborah, Laurel Smith-Doerr, and Lauren M. O'Brien. 2014. "Gendered Networks: Professional Connections of Science and Engineering Faculty." In *Gender Transformation in the Academy*, edited by Vasilikie Demos, Catherine White Berheide, and Marcia Texter Segal. Bringley, UK: Emerald Group Publishing.

Bielby, William T., and James N. Baron. 1986. "Men and Women at Work: Sex Segregation and Statistical Discrimination." *American Journal of Sociology* 91(4): 759–99.

Brass, Daniel J. 1985. "Men's and Women's Networks: A Study of Interaction Patterns and Influence in an Organization." *Academy of Management Journal* 28(2): 327–43.

Breaugh, James. 2014. "Employee Recruitment." In *Meeting the Challenge of Human Resource Management: A Communication Perspective*, edited by Vernon D. Miller and Michael E. Gordon. London: Routledge.

Brown, Meta, Elizabeth Setren, and Giorgio Topa. 2016. "Do Informal Referrals Lead to Better Matches? Evidence from a Firm's Employee Referral System." *Journal of Labor Economics* 34(1): 161–209.

Burks, Stephen V., Bo Cowgill, Mitchell Hoffman, and Michael Housman. 2015. "The Value of Hiring Through Employee Referrals." *Quarterly Journal of Economics* 130(2): 805–39.

Campbell, Karen E. 1988. "Gender Differences in Job-Related Networks." *Work and Occupations* 15(2): 179–200.

Cejka, Mary Ann, and Alice H. Eagly. 1999. "Gender-Stereotypic Images of Occupations Correspond to the Sex Segregation of Employment." *Personality and Social Psychology Bulletin* 25(4): 413–23.

Collet, Francoise Herve, Peter Hedström, and Anders Johansson. 2014. "Micro Dynamics and Macro Outcomes: A Study of the Role of Networks in Segregation Processes." *Academy of Management Proceedings* 1 (January): 15828.

Correll, Shelley J. 2001. "Gender and the Career Choice Process: The Role of Biased Self-Assessments." *American Journal of Sociology* 106(6): 1691–730.

Cotter, David A., Joan M. Hermsen, Seth Ovadia, and Reeve Vanneman. 2001. "The Glass Ceiling Effect." *Social Forces* 80(2): 655–81.

Fernandez, Roberto M., and Mabel Abraham. 2010. "From Metaphors to Mechanisms: Gender Sorting into an Organizational Hierarchy." MIT Sloan research paper no. 4779-10. Cambridge, Mass.: Massachusetts Institute of Technology. Accessed October 25, 2018. http://papers.ssrn.com/abstract=1589012.

———. 2011. "Glass Ceilings and Glass Doors? Internal and External Hiring in an Organizational Hierarchy." MIT Sloan research paper no. 4895-11. Cambridge, Mass.: Massachusetts Institute of Technology. Accessed October 25, 2018. http://papers.ssrn.com/abstract=1804896.

Fernandez, Roberto M., and Santiago Campero. 2017. "Gender Sorting and the Glass Ceiling in High-Tech Firms." *Industrial and Labor Relations Review* 70(1): 73–104.

Fernandez, Roberto M., Emilio J. Castilla, and Paul Moore. 2000. "Social Capital at Work." *American Journal of Sociology* 105(5): 1288–356.

Fernandez, Roberto M., and Colette Friedrich. 2011. "Gender Sorting at the Application Interface." *Industrial Relations: A Journal of Economy and Society* 50(4): 591–609.

Fernandez, Roberto M., and M. Lourdes Sosa. 2005. "Gendering the Job: Networks and Recruitment at a Call Center." *American Journal of Sociology* 111(3): 859–904.

Fernandez, Roberto M., and Nancy Weinberg. 1997. "Sifting and Sorting: Personal Contacts and Hiring in a Retail Bank." *American Sociological Review* 62(6): 883–902.

Fernandez-Mateo, Isabel. 2009. "Cumulative Gender Disadvantage in Contract Employment." *American Journal of Sociology* 114(4): 871–923.

Fernandez-Mateo, Isabel, and Roberto M. Fernandez. 2016. "Bending the Pipeline? Executive Search and Gender Inequality in Hiring for Top Management Jobs." *Management Science* 62(12): 3636–55.

Glass, Christy, and Alison Cook. 2016. "Leading at the Top: Understanding Women's Challenges Above the Glass Ceiling." *Leadership Quarterly* 27(1): 51–63.

Gorman, Elizabeth H., and Julie A. Kmec. 2009. "Hierarchical Rank and Women's Organizational Mobility: Glass Ceilings in Corporate Law Firms." *American Journal of Sociology* 114(5): 1428–74.

Hensvik, Lena, and Oskar N. Skans. 2016. "Social Networks, Employee Selection, and Labor Market Outcomes." *Journal of Labor Economics* 34(4): 825–67.

Hymowitz, Carol, and Timothy D. Schellhardt. 1986. "The Glass Ceiling: Why Women Can't Seem to Break the Invisible Barrier that Blocks Them from the Top Jobs." *Wall Street Journal* 24(1): 1573–92.

Ibarra, Herminia. 1992. "Homophily and Differential Returns: Sex Differences in Network Structure and Access in an Advertising Firm." *Administrative Science Quarterly* 37(2): 422–47.

Jacobs, Jerry A. 1989. *Revolving Doors: Sex Segregation and Women's Careers*. Stanford, Calif.: Stanford University Press.

Kleinbaum, Adam M., Toby E. Stuart, and Michael L. Tushman. 2013. "Discretion Within Constraint: Homophily and Structure in a Formal Organization." *Organization Science* 24(5): 1316–36.

Kmec, Julie A. 2005. "Setting Occupational Sex Segregation in Motion: Demand-Side Explanations of Sex Traditional Employment." *Work and Occupations* 32(3): 322–54.

———. 2008. "The Process of Sex Segregation in a Gender-Typed Field: The Case of Male Nurses." *Sociological Perspectives* 51(2): 259–79.

Lawrence, Barbara S. 2006. "Organizational Reference Groups: A Missing Perspective on Social Context." *Organization Science* 17(1): 80–100.

Leicht, Kevin T., and Jonathan Marx. 1997. "The Consequences of Informal Job Finding for Men and Women." *Academy of Management Journal* 40(4): 967–87.

Lincoln, James R., and Jon Miller. 1979. "Work and Friendship Ties in Organizations: A Comparative Analysis of Relation Networks." *Administrative Science Quarterly* 24(2): 181–99.

Marsden, Peter V. 1987. "Core Discussion Networks of Americans." *American Sociological Review* 52(1): 122–31.

———. 1988. "Homogeneity in Confiding Relations." *Social Networks* 10(1): 57–76.

Marsden, Peter V., and Elizabeth H. Gorman. 2001. "Social Networks, Job Changes, and Recruitment." In *Sourcebook of Labor Markets*, edited by Ivar Berg and Arne L. Kalleberg. New York: Springer US.

McDonald, Steve. 2011. "What's in the 'Old Boys' Network? Accessing Social Capital in Gendered and Racialized Networks." *Social Networks* 33(4): 317–30.

Moore, Gwen. 1990. "Structural Determinants of Men's and Women's Personal Networks." *American Sociological Review* 55(5): 726–35.

Penner, Emily K., Jane Rochmes, Jing Liu, Sabrina Solanki, and Susanna Loeb. 2019. "Differing Views of Equity: How Prospective Educators Perceive Their Role in Closing Achievement Gaps." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(3): 103–27. DOI: 10.7758/RSF.2019.5.3.06.

Petersen, Trond, and Ishak Saporta. 2004. "The Opportunity Structure for Discrimination." *American Journal of Sociology* 109(4): 852–901.

Reardon, Sean F. 2019. "Educational Opportunity in Early and Middle Childhood: Using Full Population Administrative Data to Study Variation by Place and Age." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(2): 40–68. DOI: 10.7758/RSF.2019.5.2.03.

Reskin, Barbara F. 2003. "Including Mechanisms in Our Models of Ascriptive Inequality." *American Sociological Review* 68(1): 1–21.

Reskin, Barbara F., Debra B. McBrier, and Julie A. Kmec. 1999. "The Determinants and Consequences of Workplace Sex and Race Composition." *Annual Review of Sociology* 25(1): 335–61.

Rider, Christopher I. 2012. "How Employees' Prior Affiliations Constrain Organizational Network Change: A Study of US Venture Capital and Private Equity." *Administrative Science Quarterly* 57(3): 453–83.

Ridgeway, Cecilia L. 1997. "Interaction and the Conservation of Gender Inequality: Considering Employment." *American Sociological Review* 62(2): 218–35.

Rubineau, Brian, and Roberto M. Fernandez. 2013. "Missing Links: Referrer Behavior and Job Segregation." *Management Science* 59(11): 2470–89.

———. 2015. "Tipping Points: The Gender Segregating and Desegregating Effects of Network Recruitment." *Organization Science* 26(6): 1646–64.

Straits, Bruce C. 1996. "Ego-Net Diversity: Same- and Cross-Sex Coworker Ties." *Social Networks* 18(1): 29–45.

Tinsley, Catherine H., James B. Wade, Brian G. Main, and Charles A. O'Reilly. 2017. "Gender Diversity on US Corporate Boards: Are We Running in Place?" *ILR Review* 70(1): 160–89.

Yap, Margaret, and Alison M. Konrad. 2009. "Gender and Racial Differentials in Promotions: Is There a Sticky Floor, a Mid-Level Bottleneck, or a Glass Ceiling?" *Relations Industrielles/Industrial Relations* 64(4): 593–619.

Differing Views of Equity: How Prospective Educators Perceive Their Role in Closing Achievement Gaps

EMILY K. PENNER, JANE ROCHMES, JING LIU, SABRINA M. SOLANKI, AND SUSANNA LOEB

Hiring is an opportunity for school districts to find educators with values and beliefs that align with district goals. Yet beliefs are difficult to measure. We use administrative data from more than ten thousand applications to certificated positions in an urban California school district in which applicants submitted essays about closing achievement gaps. Using structural topic modeling (STM) to code these essays, we examine whether applicants systematically differ in their use of these themes and whether themes predict hiring outcomes. Relative to white applicants, Hispanic and African American applicants are more likely to identify structural causes of inequities and discuss educators' responsibilities for addressing inequality. Similar differences in themes emerge between applicants to schools with different student populations. Techniques like STM can decipher hard-to-measure beliefs from administrative data, providing valuable information for hiring and decision making.

Keywords: teacher hiring, structural topic modeling, educational equity

Schools across the country serve an increasingly diverse student body, and California's schools lead the nation in diversity. Currently some 75 percent of public school students in California identify as a race-ethnicity other than white and nearly 43 percent speak a language other than English at home (California Department of Education 2016a, 2016b). Yet these students have not had equal educational opportunities or successes. Despite repeated

Emily K. Penner is assistant professor of education at the University of California, Irvine. **Jane Rochmes** is assistant professor of sociology at Christopher Newport University. **Jing Liu** is a postdoctoral research associate in the Annenberg Institute at Brown University. **Sabrina M. Solanki** is a doctoral candidate in the School of Education at the University of California, Irvine. **Susanna Loeb** is director of the Annenberg Institute and professor of education and international and public affairs at Brown University.

 Penner, Emily K., Jane Rochmes, Jing Liu, Sabrina M. Solanki, and Susanna Loeb. 2019. "Differing Views of Equity: How Prospective Educators Perceive Their Role in Closing Achievement Gaps." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(3): 103–27. DOI: 10.7758/RSF.2019.5.3.06. Support for this project was provided by grants from the National Science Foundation (IBSS-L1620419), Institute for Education Sciences (R305B130017), and the Stanford Graduate School of Education Incentive Fund. Direct correspondence to: Emily K. Penner at emily.penner@uci.edu, 3200 Education, University of California, Irvine, Irvine, CA 92697; Jane Rochmes at jane.rochmes@cnu.edu, Department of Sociology, Social Work, and Anthropology, Christopher Newport University, 1 Avenue of the Arts, Newport News, VA 23606; Jing Liu at jing_liu3@brown.edu, Annenberg Institute for School Reform, Brown University, 383 Benefit St., Providence, RI 02903; Sabrina M. Solanki at ssolanki@uci.edu, 3200 Education, University of California, Irvine, Irvine, CA 92697; and Susanna Loeb at loeb@brown.edu, Annenberg Institute for School Reform, Brown University, 383 Benefit St., Providence, RI 02903.

efforts to equalize educational experiences and outcomes for students, school segregation and racial achievement gaps persist and economic achievement gaps have widened (Harris and Herrington 2006; Reardon 2011; Reardon and Bischoff 2011; Reardon and Owens 2014).

Hiring presents school districts with an opportunity to identify educators who have not only relevant experience, but also the beliefs, attitudes, and pedagogical skills that enable them to promote engagement and achievement among students with varied experiences and backgrounds. However, districts often struggle to identify candidates with the characteristics they most desire, and identifying candidates whose beliefs align best with district policy strategies and goals is challenging. Although recent evidence suggests that well-developed applications and interview processes can provide valuable insights into applicant quality, such hiring methods can be cost prohibitive for many districts (Jacob et al. 2016; Rockoff et al. 2011).

This article examines how educator applicants address issues of equity and diversity in their application materials and considers how such beliefs and attitudes influence application decisions and hiring outcomes. We use seven years of unique essay data from more than ten thousand applications to our partner district in California to examine how educator applicants express their beliefs and how the district evaluates them in its hiring process.[1] Because equity is central to the district's mission, it explicitly asks all applicants to write a short essay discussing how they would address the achievement gap in their classrooms if hired. We use structural topic modeling (STM) to code the ten most common essay themes from applicant responses to this prompt. We describe variation in the prevalence of each essay theme across educators and contrast response themes by educator race. We then use regression analyses to test whether applicants that articulate the most common beliefs, values, and strategies about how to address achievement inequities apply to schools with more disadvantaged students, rate more highly in application review, and are ultimately more likely to be hired by the district.

Educator short-essay responses cover a wide range of topics, some expressing more general sentiments about the achievement gap and others describing specific strategies and beliefs about how they would address it. Essay content varies by race-ethnicity: relative to white applicants, African American and Hispanic applicants focus more on naming structural causes of inequities facing students, advocating solutions rooted in educators' responsibilities to address inequity and their own biases and prioritizing cultural and linguistic diversity; Asian applicants focus more on building supportive classrooms and family and community engagement. Several essay themes differentially predict applicants' rubric scores on evaluations of their applications and the likelihood that the district hires them. Relative to differentiated instruction, a common theme, applicant essays that describe experience with cultural and linguistic diversity and educators' responsibilities to address inequities receive the highest rubric scores and are associated with the highest likelihood of being hired in the district. Further, several themes are positively related to applying to and being hired in schools with the largest populations of traditionally underserved students—particularly Hispanic students and English-language learners. These results are robust to a variety of model specifications, including those that control for the writing quality of the essays, essay sentiment, educator de-

1. We include applicants to all certificated positions, including classroom teachers, principals, therapists, and others, because some applicants apply to both teaching positions as well as some other type of certificated position for which they believe they are qualified, making the search processes for both positions contingent on one another. For parsimony, throughout the article we refer to all applicants as educators or teachers, reflecting the fact that the majority of the existing literature on educator hiring focuses on teachers. Results discussed are largely consistent whether we use the full sample of educator applicants or only the subsample of applicants to teaching positions (81 percent of the full applicant pool and 85 percent of the hired pool). Nearly all coefficients are similar in sign, significance, and magnitude. A small subset lose or gain significance with the teacher-only sample.

mographics, experience, and credentials, and fixed effects for the specific jobs to which applicants applied.

Structural topic modeling techniques are a promising tool for examining administrative data from school districts as well as other settings, particularly qualitative and long-form data that are typically very resource-intensive to analyze. In this application and elsewhere, STM can help examine hard-to-measure practices and beliefs, providing more information at the point of hire and informing decision making in other areas.

BACKGROUND AND MOTIVATION

Although somewhat diminished in recent decades, racial achievement gaps remain an enduring feature of schools nationwide (Gamoran 2001; Murphey 2014; Reardon 2011). Economic achievement gaps, meanwhile, are expanding markedly, especially between those in the middle of the distribution and their highest income peers (Reardon 2011). School segregation remains entrenched and perpetuates unequal learning opportunities along racial-ethnic and economic lines (Reardon and Owens 2014; Reardon and Bischoff 2011). Our society expects educators to address these persistent challenges and to create classrooms that provide opportunities for all students. However, because we struggle to agree on the goals of schooling, we have no singular blueprint for how to do so (Labaree 1997).

Instead, teachers have a variety of goals for their teaching, and some may not prioritize reducing inequality (Rimm-Kaufman et al. 2006). If educators in some schools think addressing inequality is a key priority of their work but their counterparts in other schools do not, they might create very different learning environments for their students, further exacerbating inequality. Educators also have differing levels of comfort and experience in planning and enacting solutions to address inequality in their work. Understanding this variation is important because teachers' beliefs are an important indicator of what they are likely to do in their classrooms (Buehl and Beck 2015; Pajares 1992; Wilkins 2008; Opfer and Pedder 2011). Despite the central role we expect educators to play in addressing inequality, we know little about what educators think about inequality and how they should work to address it.

We need a better understanding of what educators think about inequality because educators' identities and beliefs are consequential for learning environments, opportunities, and student outcomes, particularly for marginalized students. We know that teachers' identities shape students' schooling experiences and outcomes. For example, teacher-student race and gender congruence positively affects student achievement, grades, their academic perceptions and attitudes, and discipline outcomes (Dee 2004; Egalite and Kisida 2018; Egalite, Kisida, and Winters 2015; Fox 2016; Gershenson et al. 2017; Lindsay and Hart 2017). These effects are often particularly large for black students. The mechanisms driving these effects—typically presumed rather than empirically tested—are that same-race or same-gender educators not only offer representation but also convey values and attitudes that recognize and affirm the unequal experiences of students from marginalized backgrounds.

Underneath demographic similarities, educators have a wide variety of experiences and knowledge that ultimately shape their beliefs about their work and their students. Teacher beliefs are an integral part of teacher practice, learning, development, and identity (Hollingsworth 1989; Opfer and Pedder 2011; Smagorinsky et al. 2004). Teachers' perceptions affect how they treat their students, the types of help and support they provide, and teachers' empathy toward students (Calarco 2011; Ferguson 2003; Okonofua, Paunesku, and Walton 2016). Teacher perceptions, expectations, and stereotypes can influence disciplinary practices and students' later course trajectories and achievement, contributing to the growth of achievement gaps (Ferguson 2003; Okonofua and Eberhardt 2015; Baker et al. 2015; Alvidrez and Weinstein 1999). Teachers' beliefs about overcoming students' social disadvantage influence student achievement (Rochmes 2018). Moreover, teachers' expectations for student ability and achievement matter most for the outcomes of students from disadvantaged groups (Dee 2005; Downey and Pribesh 2004; Ferguson 2003; McGrady and Reynolds 2013; Rist 1970).

Understanding what educators think about

how to address inequality is important not only because of how beliefs influence educator practice, but also because if educators' goals are misaligned with school or district goals, policy solutions to reduce inequality are unlikely to work. Teachers can promote district goals but can also impede and even derail them if they are not well aligned with the goals of individual teachers or the collective goals of their instructional teams (Coburn 2001, 2004; Coburn, Hill, and Spillane 2016; Golann 2018; Spillane 1999).

Because educators play such an important role in carrying out district priorities, the alignment between district goals and educators' goals is an important consideration for educator hiring. Once we know what educators think about how to address inequality, we can also examine how well aligned these beliefs are with district priorities. Districts might use such data in a range of human resource decisions, including hiring. For districts that are particularly concerned with reducing opportunity and achievement gaps for diverse students, hiring presents an important opportunity to identify educators equipped to meet the needs of marginalized students.

In response to national priorities and district interests, many districts prioritize equity-focused beliefs in their recruitment, demonstrating this commitment by prioritizing the diversification of their teaching force (Bireda and Chait 2011; Brown and Boser 2017; Villegas and Irvine 2010). Despite concerted efforts to increase educator diversity through recruitment, training, and retention efforts, the pace of diversification is slow; in 2016, only 18 percent of the nation's teachers identified as teachers of color, an increase of only 1 percent from 2004 (King, McIntosh, and Bell-Ellwanger 2016; National Center for Education Statistics 2004).

Other avenues are possible for making learning opportunities and outcomes more equitable by carefully considering the beliefs and values of all prospective educators, regardless of demographic background. Districts interested in prioritizing candidates with particular beliefs and values can use interview and application essays to collect this information. Earlier research indicated that applications and interviews could not help districts make effective judgments to fulfill hiring priorities; more recent evidence, however, suggests that thoughtful, well-targeted recruitment efforts can provide meaningful information about applicants (Balter and Duncombe 2006; DeArmond, Gross, and Goldhaber 2010; Jacob et al. 2016; Rockoff et al. 2011). In particular, careful consideration of applicant characteristics through rigorous screening methods and interview processes can identify teachers with desired characteristics and behaviors who have a positive impact on student achievement (Goldhaber, Grout, and Huntington-Klein 2016; Jacob et al. 2016).

If districts want to develop a teaching force that has competencies, beliefs, and mindsets that align with district goals for addressing inequality, then they need to create a recruitment process that can identify these qualities among its applicants. This article draws on data from a partner district in California that uses its application to examine educators' approaches to equity goals during its recruiting process. In particular, it asks all applicants to respond to a short-answer essay prompt asking about the approaches they will use to address achievement gaps in the district. The district's human resources department scores these essays using a rubric during an initial screening process, reviewing for general evidence and understanding of equity commitments rather than specific beliefs, attitudes, or teaching strategies. This limited consideration is understandable given that application review is a time-intensive process, particularly with a large volume of applications.

Given resource constraints, it is difficult to measure educators' beliefs for a large number of individuals in other ways. It is time consuming and expensive to conduct multistage interviews or surveys. Such methods may also provide only a narrow slice of what educators think and can be prone to response bias or adverse reactions to particular types of interviewers. For districts wanting to learn more about applicants' beliefs, such methods can be time and cost prohibitive.

Recent advances in text mining paired with improvements in district administrative data collected from application materials provide new avenues for learning about applicant beliefs outside of extensive interviews. Social sci-

ence research is just beginning to use computational approaches to analyze text and dialogue to reveal additional features about individuals, their conversations, and verbal and written statements (Bettinger, Liu, and Loeb 2016; Gentzkow, Kelly, and Taddy 2017; Kelly et al. 2018; Liu 2018; McFarland, Jurafsky, and Rawlings 2013). The economist Jens Ludwig highlights the utility of data-mining techniques to help solve complex policy problems (2018). The sociologists Roberto Fernandez and Brian Rubineau identify important insights that applicant pool and referral data can provide about hiring dynamics (2019). Given important insights provided by hiring data (Fernandez and Rubineau 2019), we argue that hiring data provide a particularly exciting opportunity for using these techniques.

This study provides an important first case of applying machine-learning techniques to text-rich application data to examine the role of applicant beliefs in hiring processes and outcomes. It examines these relationships for more than ten thousand applicants over seven years in a large, urban district that has often struggled to attract and retain educators, particularly in its most disadvantaged schools. In addition to these methodological and data advances, this study has the novel purpose of focusing on educator attitudes toward and solutions for addressing the achievement gap, which is difficult to measure in a large sample. This study tests whether a district that expresses a deep commitment to equity and justice actually selects applicants that articulate attitudes and solutions that cohere with the district's priorities. Together, these results shed light on two often-neglected pieces of the inequality puzzle—what educators think and whether districts act in line with their values.

Our investigation of the role that educator beliefs about inequality play in hiring examines the following research questions:

> What themes predominate in applicants' attitudes about achievement gaps?
>
> Do the themes discussed vary by applicant race?
>
> Are individuals who write about particular themes more likely to apply to schools with particular characteristics?
>
> Do the essay themes and scores predict hiring outcomes, overall and at particular schools?

DATA AND METHODS

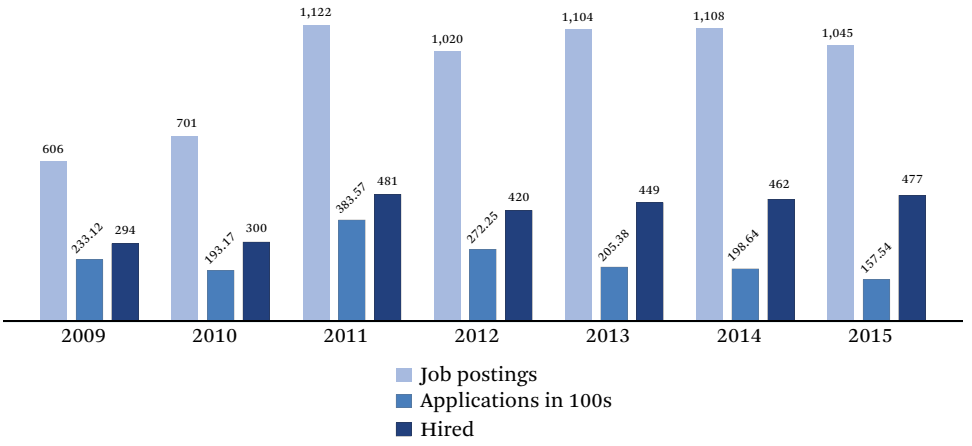
This article uses school-district administrative data to describe educator attitudes and examine how these attitudes affect hiring outcomes. Our source for these attitudes is short-answer responses on job applications. These data are conventional in the sense that many districts screen applicants with some type of writing exercise (Jones 2012). Our use of them, however, is novel because we code them for applicants' attitudes and dispositions using structural topic modeling. These data provide the first use of machine-learning coded applicant essays to learn more about the role of applicants' beliefs in hiring. This district's screening question is particularly useful for examining educators' attitudes because it asks applicants to wrestle with a problem that will directly affect their daily teaching experiences. As a result, it provides a rich lens to uncover educators' feelings about this challenge.

Our partner district is an urban, public school district in California that employs more than 3,500 educators and administrators to serve a diverse student body of more than fifty thousand students (California Department of Education 2015). It is also a district with highly unequal outcomes for its students. In recent years, among California's large urban districts, it had both the highest average achievement and the widest gap between that average and the district's lowest-performing students, as well as large racial-ethnic achievement gaps. Although the district has made progress, gaps in suspensions, graduation rates, and achievement are ongoing challenges.[2]

To examine these relationships, we use applicant data from our partner district's human resources department. Applicants to positions in the district applied through a proprietary online interface. Data from this interface were then linked to other administrative records

2. This information is from the district website and other local sources published between 2008 and 2017, withheld here to preserve the anonymity of the district. For more information, please contact the authors directly.

Figure 1. Number of Job Postings, Applications, and Applicants Hired



Source: Authors' compilation based on partner district data.
Note: Number of hires = 2,883.

from human resources by the research team. Current school district and state databases increasingly include detailed information about employees, but it is far less common that data from applications to district positions are linked to other district data sources (but, for notable exceptions, see Goldhaber, Grout, and Huntington-Klein 2016; Jacob et al. 2016; Saavedra et al. 2017). This type of connection allows for new research approaches that link the applicant pool to the pool of employed educators and other employees within a district. This approach has provided novel insights about hiring pathways, sorting, and discrimination in other sectors (Fernandez and Friedrich 2011; Petersen and Togstad 2006). However, research has not used applicant data that includes long-form writing and would typically be considered qualitative. By applying structural topic modeling to applicant essays, this study is the first to link attitudes within the applicant pool to hiring outcomes.

We examine data for all applicants to certificated positions from March 2009 to October 2015. Figure 1 shows the number of job postings, applications (in thousands), and applicants hired by year.

After an initial acceleration as the system came online, postings, applications, and number of applicants hired stayed fairly consistent across the seven years of our panel, save for an uptick toward the end of the financial crisis in 2011. During this window, this district posted 11,599 unique jobs, of which we use the 6,706 positions for which one or more applications were submitted. Postings received eighty-seven applications on average (from 1 to 544). The district received 218,196 applications for all certificated positions from 14,421 unique individuals. Our analytic sample consists of the 10,188 applicants (13,016 applicant-year observations) that completed essays (164,367 applications). Applicants submitted an average of sixteen applications each over this period (an average of 22.97 at the applicant-year level). The district ultimately hired 2,883 individuals from this pool.[3] Table 1 presents descriptive characteristics, experience, and qualifications for the district applicant pool to all certificated positions between 2009 and 2015, and compares these

3. Some applicants (27 percent) applied in multiple years, and some applicants applied and were hired in multiple years. These applicants have distinct application data from each job search and are thus treated separately in all analyses. Our sample of 2,883 hired individuals includes 2,442 unique applicants, some hired multiple times.

Table 1. Descriptive Statistics of Applicant Sample by Hired Status

	Full Sample		Hired		Not Hired		Group Diffs
Variable	Mean	SD	Mean	SD	Mean	SD	*p*-value
Hired	0.22	0.42	1	0	0	0	
Applicant characteristics							
White	0.52	0.50	0.49	0.50	0.53	0.50	0.000
Black	0.05	0.22	0.04	0.20	0.05	0.23	0.007
Asian	0.14	0.35	0.17	0.37	0.14	0.34	0.000
Hispanic	0.14	0.34	0.16	0.37	0.13	0.34	0.000
Other race	0.04	0.20	0.05	0.22	0.04	0.20	0.048
Decline to state race	0.10	0.30	0.10	0.30	0.11	0.31	0.177
Female	0.67	0.47	0.72	0.45	0.65	0.48	0.000
Years of prior K–12 experience	4.80	6.29	4.58	5.78	4.87	6.43	0.032
Number of jobs on application	4.22	2.59	4.24	2.51	4.21	2.62	0.637
Multiple subject credential	0.36	0.48	0.41	0.49	0.34	0.47	0.000
Special education credential	0.19	0.40	0.22	0.41	0.19	0.39	0.000
STEM credential	0.15	0.36	0.15	0.36	0.16	0.36	0.341
Humanities credential	0.10	0.30	0.07	0.26	0.11	0.32	0.000
Foreign language credential	0.02	0.15	0.02	0.14	0.02	0.15	0.167
Social science credential	0.07	0.26	0.06	0.23	0.08	0.27	0.001
Physical education	0.02	0.13	0.02	0.15	0.02	0.12	0.011
Early childhood credential	0.02	0.13	0.01	0.11	0.02	0.13	0.034
Administrative credential	0.01	0.08	0.01	0.07	0.01	0.09	0.249
Supplemental credential	0.03	0.17	0.03	0.18	0.03	0.17	0.801
Substitute credential	0.03	0.18	0.01	0.12	0.04	0.19	0.000
GPA	3.02	0.99	3.16	0.85	2.98	1.03	0.000
No degree declared	0.03	0.17	0.02	0.13	0.04	0.18	0.000
Associate degree	0.00	0.05	0.00	0.05	0.00	0.06	0.344
Bachelor's degree	0.40	0.49	0.39	0.49	0.41	0.49	0.056
Master's degree	0.53	0.50	0.57	0.50	0.52	0.50	0.000
PhD	0.03	0.18	0.03	0.16	0.03	0.18	0.033
N applications submitted	22.97	40.94	23.33	40.73	22.87	40.99	0.591
Essay themes							
Special services	0.08	0.08	0.08	0.09	0.08	0.08	0.000
Family and community engagement	0.14	0.09	0.12	0.08	0.14	0.10	0.000
Believe to overcome	0.10	0.08	0.09	0.08	0.10	0.08	0.000
Experience with cult./ling. diversity	0.09	0.09	0.10	0.09	0.08	0.08	0.000
Standards and assessment	0.11	0.10	0.11	0.09	0.11	0.10	0.274
Educators' responsibilities	0.08	0.06	0.09	0.07	0.08	0.06	0.000
Supportive classroom	0.17	0.10	0.17	0.09	0.17	0.10	0.006
Naming structural causes	0.08	0.08	0.08	0.07	0.08	0.08	0.006
Cross-subject strategies	0.05	0.07	0.05	0.07	0.05	0.07	0.342
Differentiated instruction	0.11	0.09	0.11	0.08	0.11	0.09	0.023
Writing quality							
Sentiment measure: Syuzhet	12.02	6.36	13.26	6.21	11.67	6.36	0.000
Automated readability index	12.89	3.50	13.20	3.18	12.81	3.58	0.000
Type-token ratio (lexical complexity)	0.59	0.10	0.57	0.08	0.59	0.10	0.000
Number of words	241.33	126.71	264.83	129.24	234.64	125.18	0.000
Number of misspelled words	1.14	2.47	1.31	3.18	1.09	2.22	0.000
Total score: all application essays	5.97	1.73	6.63	1.55	5.76	1.73	0.000
Achievement gap essay score	1.89	0.81	2.19	0.73	1.79	0.81	0.000
N (unique individuals)	13,016		2,883		10,133		

Source: Authors' tabulations.

Note: The total number of unique submissions is 164,367.

characteristics for successful and unsuccessful applicants.

Hired applicants differ on many dimensions from those who were not hired. More of the hired applicants were Asian or Latino and fewer were white.[4] The hired applicants had fewer years of experience, but were more likely to have worked in the district before applying for the job for which we observe them being hired. More hired teachers had single-subject credentials (eligible to teach in a particular content area in grades six through twelve) and fewer had multiple-subject credentials (eligible to teach all subjects in grades pre–K through eight) relative to the nonhired teachers.

Coding Applicant Essays to Identify Most Common Themes

The district's application also asks for responses to three short-answer essays detailing how applicants would address particular social issues and problems of practice that are relevant for teaching in the district context. Because of the district's equity emphasis, we focus solely on the following short-essay question:

> Superintendent [Name] has stated that the achievement gap is the greatest civil rights issue facing our district today and closing that gap is the foundation and vision for the critical work of our teachers, staff, and administrators every day. As an educator, what is your role in working towards closing the achievement gap in [district]? In considering the demographics of our student population, what experiences or skills make you well-positioned to close the achievement gap in the context of a diverse district such as [district]?

This study uses responses to this question to indicate job candidates' attitudes toward equity and approaches to addressing inequality. Notably, the district leaves the specific meaning of the achievement gap to the interpretation of the applicant, not defining priorities about racial achievement gaps, income achievement gaps, or simply gaps between high and low achievers.

We use structural topic modeling, an unsupervised machine-learning technique, to detect the most common topics applicants discuss and their distributions across all of the essays (see, for example, Roberts, Stewart, and Tingley 2014). STM assumes that each document is a mixture of topics and that each topic is a mixture of a set of representative words. STM is particularly useful for evaluating massive textual data, as in our case, and the output of STM can provide metrics about the text content to be used in quantitative analysis.

Before we estimated the STM, we conducted standard preprocessing to prepare the texts for analysis. We first removed all stop words and punctuation.[5] We then transformed all the words to lower case and reduced words to their root form, a procedure called stemming. As an unsupervised method, STM requires us to set the number of topics before we run the model. The optimal number of topics needs to balance model fit and substantive interpretation. We estimated the model multiple times to identify the ten, fifteen, twenty, and thirty most common, unique topics. We also controlled for whether an applicant previously worked in the district and their total years of K–12 teaching experience in our topic estimation models.

Our research team then examined the output from each model to classify the topics into essay themes. Four human coders examined model output, including top words, the most common words for each theme; *FREX,* the frequency and the exclusivity of each word to each theme; *lift*, which weights words by their frequency in a specific topic and other topics; and *score*, the top words from an index commonly used in linear discriminant analysis. Coders independently reviewed example essays that were the most emblematic of each topic to assign a topic label. From this, we discussed disagree-

4. The applicant pool is less diverse than the current teaching force in the district where approximately 50 percent of teachers identify as nonwhite.

5. Stop words are considered the most common words in a language, such as *the*, *about,* and *own*. In addition to removing these words, we remove many of the words used in the question itself, particularly the name of the superintendent.

ments as a team and developed harmonized theme labels for the various topics within each coding set. Many model themes were consistent regardless of the number of distinct topics. In models with fewer topics, several of the themes identified in the more numerous coding schemes combined into a single theme. The coders agreed that the most concise variant with only ten distinct topics captured the majority of the themes from across the variants with a larger number of topics. Diagnostic measures of essay topic classification (not shown) indicated that we lose little information by using ten topics rather than fifteen, twenty, or thirty. In addition, our interrater reliability of theme naming was highest with the smallest set of topics (80 percent exact agreement on labels with ten topics as opposed to 50 percent with twenty topics). We use this model to describe the prevalence of each theme across the pool of applicant essays and use the themes as key predictors in our analyses.

In addition, the STM algorithm codes each essay for the proportion of the text that addresses each theme and gives a rating of the density with which the essay covers each theme. We standardize the proportion of their essays that address each theme. We then use these values in regression models examining which applicants write most about which themes and which themes predict hiring outcomes most strongly. Theme ratings are not mutually exclusive and a given essay can have elements that correspond to multiple themes simultaneously.

Because the district's evaluation of an essay might be influenced by other features of the writing, we measure several characteristics of writing quality, including sentiment, readability, lexical density, number of misspelled words, and text length. We evaluated essays based on whether their overall sentiment or tone is more positive or negative using three off-the-shelf dictionaries to score the sentiment. Correlations between the sentiment measures were fairly high (r = 0.73 to 0.78), and thus we include only the Syushet sentiment scores (Jockers 2017). Readability, or understandability, refers to the level of education one needs to be able to read a piece of text easily. Readability scores are calculated using a grade-level scale, which roughly corresponds to the number of years of education one needs to read a given text. We tested the readability of the essays using the four most popular readability measures. Correlations between the readability measures were very high (r = 0.86 to 0.97), and thus we control for only the automated readability index in our analysis (Senter and Smith 1967). Lexical density refers to the number of lexical, or content, words in a sentence divided by the total number of words. Lexical words are those that give a sentence its meaning and include the nouns, verbs, adjectives, and adverbs. A high lexical density indicates a large number of information-carrying words, which is generally more difficult to read. Lexical density is negatively correlated with the length of a text. We use a type-token ratio to measure the lexical density of the applicant essays (Templin 1957). Finally, because essay screeners may be influenced by overall essay length and obvious errors, we control for the total number of words applicants used in their essays and the percentage of those words that are misspelled.

Examining Variation in Essay Theme Content

After identifying the most common themes discussed in applicants' essays, we test whether theme coverage varies by race-ethnicity. Because we are interested in whether applicants of color and white applicants share similar beliefs about solutions for inequality and achievement gaps, we focus our discussion of the results on this dimension.[6] To examine how essay themes vary by race, we estimate ordinary least squares regression models predicting the degree to which an essay includes content about each of the ten themes. These models take the form

$$T_{int} = \alpha + Race/Ethnicity\beta_{it} + X_{it} + \gamma_t + \varepsilon_{it}, \quad (1)$$

where T_{int} is an indicator for the proportion of applicant i's essay that covered a specific theme n, in year t which is a function of their race-ethnicity β_{it}, a vector of controls X_{it} for other

6. We also test for differences based on variation among other demographic, experience, and quality dimensions (not shown).

applicant demographics, qualifications, and experience, year fixed effects γ_t, and an error term. We estimate separate models predicting variation in the coverage of each of the ten themes.

Application and Hiring Outcomes

To be hired in the district, an applicant must first upload their entire application content, including the essays, and then apply for a particular position. Applicants may submit applications for multiple positions. More than 83 percent of applicants applied for two or more jobs, but they used the same application for all district positions for which they apply.

The essay responses are likely indicative of some type of social desirability bias on the part of the applicant, given that they are seeking employment. This is likely true of all data collected about applicants because of the very nature of seeking a job. This might lead them to write responses that they believe will elicit specific reactions on the part of principals in the district. However, because applicants cannot change their essay, they are unlikely to write content tailored to different positions. Moreover, job postings for specific schools do not reference attitudes or values that principals are seeking; typically, an applicant applying to multiple positions would simply see variation in the job description or subject area needed for each posting to which they submit an application.

We do not have a good understanding of what motivates educators to apply for some positions and not others. Research indicates that teachers move to higher-achieving, whiter, more-advantaged schools over time, but it is also possible that their beliefs and attitudes influence the initial school placements they seek in a district (Boyd et al. 2005). To test this, we examine the degree to which writing about particular essay themes predicts whether applicants apply to schools with student populations that are above the annual district average for different characteristics. We code each school as having student populations that are above district averages for three characteristics based on the district's public reports of school composition: race-ethnicity, English-language learners (ELLs), and free or reduced-price lunch enrollment. To examine how applicant essays predict whether applicants apply to schools with different student populations, we estimate a series of linear probability models that take the following basic form:

$$Y_{it} = \alpha + Essay\ Elements\ \beta_{it} + X_{it} + \gamma_t + \varepsilon_{it}, \quad (2)$$

where Y_{it} is an indicator for whether applicant i applied to a school with a student population that was above the district average in that year (for example, above-average percentage ELL) in year t which is a function of several essay elements β_{it} including essay themes, indicators of writing quality, and essay scores, depending on the model, a vector of controls X_{it} for applicant demographics, qualifications, and experience, year fixed effects, γ_t, and an error term. In this analysis, we exclude applications that individuals submit to positions not located at a specific school site.[7]

After applicants submit their application for a specific position in the district through the online system, human resources staff conduct an initial screening of the application. This consists of reviewing the materials for completeness, confirming that the appropriate certification documents and transcripts have been submitted with the application, and reviewing and scoring the application essays. This screening is done with a rubric (not shared to preserve district confidentiality) that evaluates each of the three essays for specific criteria and competencies, awarding zero to three points for no evidence, mixed or limited evidence, satisfactory evidence, or strong evidence of each competency. These scores are then aggregated across essays. Applicants with a score of four or lower are removed from the pool; those with scores of five or higher are passed on to an internal database for principals and central office administrators to review.

7. Supplemental models examine all applications per applicant and include fixed effects for the general job category (such as social studies teacher, counselor, bilingual Chinese teacher), or fixed effects for the specific position an applicant applied to. As the results are qualitatively similar, we present models in the paper without the subject area or job-specific fixed effects. Alternative versions are available on request.

We use the rubric scores on each of the three application essays as an indicator of applicant quality. We create a global score for the three essays; we use the score on the achievement gap essay as a stand-alone; and we aggregate the two non-achievement gap essay scores for control variables in some models. Mean scores for these essays for the full sample and separately for hired and nonhired applicants are shown in table 1.

To examine whether discussing a specific essay theme predicts rubric scores or being hired by the district, we estimate analogous models to equation (2), where *rubric score* or *hired* is the dependent variable. The same set of predictors is included in these models. We fit the same models as equation (2) to predict whether applicants are hired by schools with student populations that are above reported district averages across several demographic categories.

RESULTS

Educators provide a great variety of responses to the essay question, suggesting that they have meaningfully different values and attitudes about the social and achievement inequities facing district students regardless of the image they attempt to project toward their prospective employer. The content covered and omitted, as well as the general tenor, argumentation, and structure help classify applicants based on their perspectives on the underlying causes and challenges involved in addressing achievement gaps. Clear distinctions appear in the approaches to addressing this challenge and in the attitudes and biases of these applicants.

We summarize this rich detail into the ten most common themes that emerge from our structural topic model coding of applicants' essays. Table 2 lists these themes. Each label summarizes the approach that educators espouse to complete the phrase, "use _____ to reduce inequality." For example, some applicants would "use *special services* to reduce inequality and close achievement gaps." For example texts most emblematic of each of the ten themes, see the appendix.[8]

Several themes describe pedagogical approaches and experiences that applicants believe equip them to address achievement gaps. Two themes speak to the ways in which applicants adapt and customize their instruction to the needs of diverse learners by using *differentiated instruction* and *cross-subject strategies*, integrating content across subject areas and adapting it for different skill-levels to reduce inequality. One theme draws heavily upon recent school reform approaches and advocates using *standards and assessment* to improve and monitor instruction thereby reducing inequality. Another argues that educators are best equipped to reduce inequality by drawing upon their personal and instructional *experiences with cultural and linguistic diversity*.

Other applicants focus more on creating a supportive, inspiring classroom and engaging school, family, and community resources in their efforts to reduce inequality. Two of the themes describe resources outside the classroom that the applicant would access. Applicants who indicate that they would access *special services* write about working with school services such as therapy and speech pathologists to support diverse learners. Those who discuss *family and community engagement* write about working with the broader community and families and actively inviting them into the school. Essays also describe a more general idea of having a *supportive classroom* rather than referring to a specific strategy. Another theme focuses on being supportive, but relies heavily on general ideas about inspiring and *believing in students to overcome* challenges without clearly connecting these ideas to specific behaviors.

Finally, two themes highlight the societal power structures that create and maintain unequal conditions and achievement gaps in the district. Essays *naming structural causes* detail the structural challenges facing marginalized students and the ways societal inequities contribute to achievement and opportunity gaps, arguing that these inequities must be interrogated and addressed to make progress in reducing inequality. In some cases, this theme also includes language typically associated with deficit orientations, emphasizing group and fam-

8. Applicants could cover multiple themes in their essays, although correlations between themes (not shown) were low to moderate, and the largest correlations of approximately 0.45 were negative.

ily conditions that reinforce societal inequities. A similar theme identifies societal causes for achievement gaps but discusses *educators' responsibilities* to work through these challenges to reduce inequality. Educators discussing this social justice–oriented theme decry opportunity gaps, want to challenge biases, and discuss drawing on students' funds of knowledge and using critical and culturally relevant pedagogies to promote equity.

Table 1 includes the average fraction of the essay text that discusses each topic and comparisons of the degree to which hired and nonhired applicants' essays discuss each of the ten topics. Tests of differences in the use of each topic by group are also shown in table 1, suggesting significant differences in topic use between hired and nonhired applicants on nearly every essay topic. Table 1 also includes descriptive comparisons of the other essay features, including readability, lexical complexity, and number of words for the full sample and between hired and nonhired applicants.

On average, the degree to which essays address each theme may seem low, but this is in part because each essay also includes text that is either not specific enough to correspond to a particular theme or addresses another topic that was not part of these ten most common themes. The most prevalent themes are supportive classroom (covered in 17 percent of essays) and family and community engagement (covered in 14 percent of essays). The least prevalent theme is cross-subject strategies, which is covered in only 5 percent of essays.

In addition to variation in percentage of applicants who cover each theme across the body of essays, applicants also vary in the degree of their essay text that is devoted to covering a specific theme. This ranges from educators' responsibilities, about which an essay has as much as 40 percent of its content covering this theme, to standards and assessment, about which an essay has as much as 68 percent of its content about this theme. As noted elsewhere, applicants often have some coverage of multiple themes in their essays, one theme frequently dominating one or more others.

Figure 2 presents a graphical representation of theme density within the entire corpus of essays, showing how two themes overlap, and displays portions of example texts that are among the most emblematic of standards and assessment and educators' responsibilities. The essay emblematic of the standards and assessment theme indicates the applicant's strong belief in standards-aligned curriculum and lesson plans, frequent assessment of student progress, and accountability to help address the achievement gap. The example educators' responsibilities essay highlights the need to use culturally relevant teaching and learning and to validate student cultures and experiences, and advocates for systemic, justice-oriented reform. These applicants clearly have different values and beliefs about educators' roles in addressing inequality and preferred strategies for addressing achievement gaps.

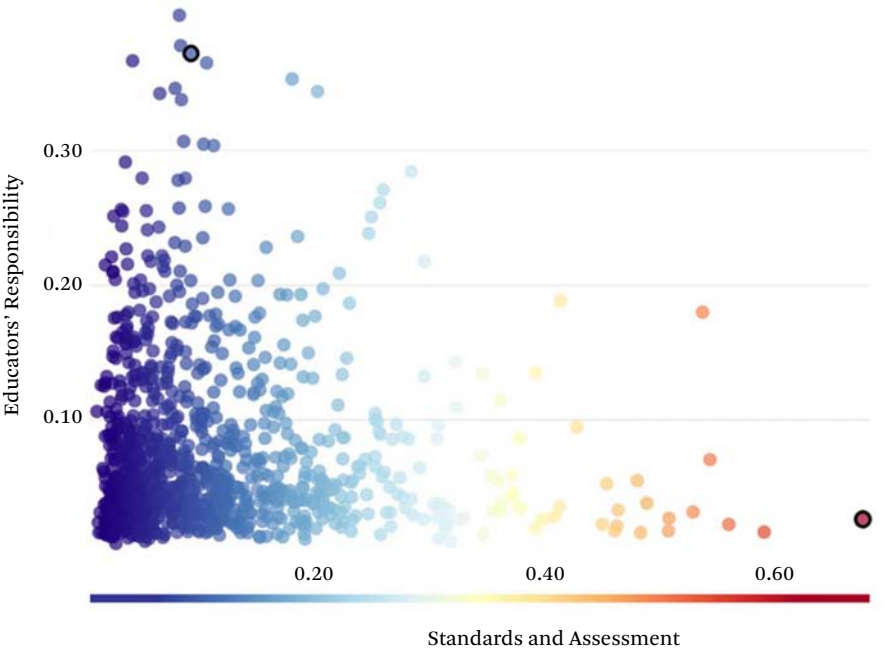
Essay themes vary by the race-ethnicity of the applicants. Table 2 displays results from models predicting theme content based on applicants' race-ethnicity, controlling for other demographic characteristics, experience, and credentials. A number of differences emerge. Applicants from different racial-ethnic groups cohere around particular sets of themes, relative to white applicants. Black applicants focus more on naming structural causes of inequality, such that their essays include 0.23 standard deviations more text about structural inequities than white applicants' essays do. They also call for greater educator responsibility than white applicants do. Asian applicants tend to focus more than white applicants on themes that engage community, family, and special services and build supportive classrooms. Compared with white applicants, Hispanic applicants focus more on creating supportive classrooms, invoke their experiences with cultural and linguistic diversity, and frame their approaches to the achievement gap by naming the structural causes of inequality and identifying educators' responsibilities to address them.[9]

Application and Hiring Outcomes

Applicant essays are predictive of candidates' application decisions. Table 3 presents results from models predicting the characteristics of

9. Differences are also notable by gender, credential type, experience, and education. Results available on request.

Figure 2. Theme Coverage Comparison: Standards and Assessment and Educators' Responsibilities Themes



Source: Authors' compilation.
Note: Strong standards and asssessment essay: "First, educational standards at both the national and state level are necessary to closing the Achievement Gap. Administrators, teachers, and parents should be aware of the standards a student must meet to pass their grade level."
Strong educators' responsibilities essay: "I not only have an obligation to provide my students with an equitable education, but also to be an advocate for change. We need to move toward a curriculum and pedagogy that is culturally responsive and relevant to our students. . . . Closing the achievement gap, creating schools that are equitable and just, changing the way we view education, and ensuring that every educator reflects on their own experiences and biases, will take time and is a journey that I am completely dedicated and committed to. I strive to teach for social justice."

schools to which applicants apply. These models include controls for the applicants' demographic characteristics, experience, credentials, and measures of writing quality. In each case, the outcome is whether the applicant applied to a position in a school that was above the district-average demographic composition for that student group in that year. Given the specificity and ubiquity of differentiated instruction in teacher preparation programs and widespread use in classrooms, and because it is the theme that most closely resembles a solution for gaps between high and low achievers rather than race or income gaps, it is the reference category theme throughout.

Table 3 shows notable differences in the essay topics covered by applicants to schools with different demographic compositions. In particular, applicants who apply to schools that have above-average Hispanic populations write least about differentiated instruction and family and community engagement. In contrast, in above-average Asian schools, applicants write less about nearly every theme than differentiated instruction, although the coefficients are only significantly different for standards and assessment. Applicants to above-average ELL schools focus most on experiences with cultural and linguistic diversity and standards and assessment, which is consistent both with the

Table 2. Predictors of Themes

	Special Services	Family and Community Engagement	Believe to Overcome	Experience with Cultural and Linguistic Diversity	Standards and Assessment	Educators' Responsibilities	Supportive Classroom	Naming Structural Causes	Cross-Subject Strategies	Differentiated Instruction
Black	0.018	−0.017	0.012	−0.030	0.063	0.089*	−0.095*	0.231***	−0.129**	−0.112**
	(0.037)	(0.039)	(0.040)	(0.035)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)
Asian	0.085***	0.178***	−0.110***	0.025	−0.078**	−0.031	0.090***	−0.022	−0.152***	−0.046
	(0.024)	(0.025)	(0.026)	(0.023)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
Hispanic	0.011	−0.021	−0.055*	0.097***	−0.033	0.075**	0.071**	0.101***	−0.138***	−0.111***
	(0.024)	(0.026)	(0.026)	(0.023)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
Other race	0.034	−0.071	−0.003	0.073	−0.031	0.119**	−0.036	0.130**	−0.040	−0.121**
	(0.040)	(0.042)	(0.043)	(0.038)	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)
Decline to state race	−0.003	−0.038	−0.126***	−0.021	−0.010	0.085*	0.042	0.102**	0.048	−0.045
	(0.035)	(0.037)	(0.038)	(0.033)	(0.037)	(0.037)	(0.037)	(0.038)	(0.038)	(0.038)
Constant	−0.180*	0.848***	0.549***	−0.459***	−0.376***	−0.474***	0.341***	−0.401***	−0.243**	0.120
	(0.072)	(0.075)	(0.078)	(0.068)	(0.077)	(0.077)	(0.077)	(0.078)	(0.078)	(0.078)
Observations	13,016	13,016	13,016	13,016	13,016	13,016	13,016	13,016	13,016	13,016
r^2	0.194	0.106	0.050	0.270	0.064	0.065	0.065	0.042	0.052	0.050

Source: Authors' tabulations.

Note: Themes standardized to have a mean of 0 and SD of 1. Race is white is omitted category. All models include controls for teacher demographics, education, experience, credentials, and year in which the applicant applied.

*$p < .05$; **$p < .01$; ***$p < .001$

Table 3. Do Essay Themes Predict Applying to Schools with Particular Characteristics?

	(1)	(2)	(3)	(4)	(5)	(6)
	(Location applied to is above the district average percent student enrollment)					
	Black	Hispanic	Asian	White	ELL	FRPL
Special services	−0.002	0.007	0.007	−0.011**	−0.001	0.001
	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
Family and community engagement	−0.000	0.004	−0.004	0.002	0.007	0.001
	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)
Believe to overcome	0.006	0.010**	−0.001	−0.010**	−0.003	0.009*
	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)
Experience with cultural-linguistic diversity	−0.003	0.016***	−0.006	−0.003	0.016***	0.008*
	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)
Standards and assessment	0.005	0.014***	−0.009*	−0.006	0.009*	0.010*
	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)
Educators' responsibilities	0.002	0.009**	−0.003	−0.001	0.003	0.001
	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)
Supportive classroom	0.000	0.012***	−0.004	−0.005	0.005	0.007
	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)
Naming structural causes	0.005*	0.009**	−0.006	−0.008**	0.001	0.011***
	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)
Cross-subject strategies	−0.002	0.011***	0.001	−0.003	0.003	0.007
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
Constant	0.498***	0.394***	0.489***	0.321***	0.637***	0.608***
	(0.030)	(0.037)	(0.036)	(0.030)	(0.036)	(0.038)
Observations	130,551	130,551	130,551	130,551	130,424	99,651
r^2	0.027	0.025	0.060	0.039	0.047	0.055

Source: Authors' tabulations.

Note: Race is white and multiple subject credential are omitted categories. Standard errors clustered by applicant. All models include controls for writing quality, teacher demographics, education, experience, credentials, and year in which the applicant applied. Schools missing on school characteristics omitted from the analysis. Models also exclude applications to centralized district positions.

*$p < .05$; **$p < .01$; ***$p < .001$

needs these schools have and the recent policy prescriptions for schools whose students struggle with English proficiency. Applicants who apply to above-average free or reduced-price lunch schools discuss themes somewhere between those who apply to above-average Hispanic schools and above-average ELL schools. Only one theme, naming structural causes, significantly predicts applying to above-average black schools.

After an applicant submits their application, screeners in the district's human resources department read it and score each of the three essays using an internal rubric. In models (not shown), we examine the relationship between overall scores and specific essay scores. Higher essay scores positively predict being hired, but the achievement gap essay score matters much more for being hired than scores on the other two essays. An additional point on the achievement gap essay increases an applicant's likelihood of being hired by 6.6 percentage points but the other two essays combined by only 3.7 percentage points (22 percent of applicants were eventually hired). This suggests that information in the achievement gap essay beyond simple evaluation scores benefits candidates. Given the importance of the essay scores and themes,

Table 4. How Essay Themes Predict Achievement Gaps Essay Scores and Being Hired

	(1) Essay Score	(2) Hired
Special services	0.058***	0.011*
	(0.012)	(0.005)
Family and community engagement	0.014	–0.004
	(0.012)	(0.005)
Believe to overcome	–0.009	0.002
	(0.012)	(0.005)
Experience with cultural-linguistic diversity	0.099***	0.026***
	(0.013)	(0.006)
Standards and assessment	0.079***	0.018**
	(0.014)	(0.006)
Educators' responsibilities	0.103***	0.025***
	(0.010)	(0.005)
Supportive classroom	0.078***	0.016**
	(0.012)	(0.006)
Naming structural causes	0.057***	0.009*
	(0.010)	(0.004)
Cross-subject strategies	0.042***	0.006
	(0.012)	(0.005)
Constant	1.570***	0.095
	(0.133)	(0.053)
Observations	13,016	13,016
r^2	0.247	0.105

Source: Authors' tabulations.
Note: Race is white and multiple subject credential are omitted categories.
Standard errors clustered by applicant. All models include controls for writing quality, teacher demographics, education, experience, credentials, and year in which the applicant applied. Models also inlcude position-specific fixed effects.
*p < .05; **p < .01; ***p < .001

table 4 displays results from models examining the relationship between essay themes, essay scores given by the initial application screening, and being hired by the district.

Discussing almost every one of the themes is associated with higher essay scores than discussing differentiated instruction is (except family and community engagement and believing in students to overcome). This suggests that the district prioritizes many other beliefs and responses to achievement gaps over something that is part of typical educator training and practice, such as differentiated instruction. The themes scored most highly are educators' responsibilities and experience with cultural or linguistic diversity, but standards and assessment and supportive classroom also scored fairly high. This pattern holds even with controls for racial-ethnic identity and is thus not an artifact of the district simply preferring candidates of color, who are more likely to discuss these themes in their essays.

Many of these themes are also associated with being hired in the district. In general, the themes that received higher scores in the initial screening process are associated with the greatest likelihood of being hired. Although the coefficients might seem small, it is important to consider them relative to the means and standard deviations for the sample. For example, the average achievement gap essay score was 1.9 (on a 0 to 3 scale) with a standard deviation of 0.8. In discussing educators' responsibilities more than differentiated instruction, an appli-

cant increases their achievement gap essay score by 0.1 points, which is 13 percent of a standard deviation. Moreover, that same content increases the likelihood that they are hired by 2.5 percentage points, regardless of the position sought, which is roughly a 10 percent increase. Additional models (not shown) omit different themes to serve as the reference category to test the relative importance of each theme. Although writing about almost any of the themes improves the likelihood of getting hired more than discussing differentiated instruction, two themes relate to the highest likelihood of being hired relative to all others: experience with cultural-linguistic diversity and educators' responsibilities. The two themes are not significantly different from one another, although the coefficients for educators' responsibilities are positive relative to those of cultural-linguistic diversity.

An additional concern is the distribution of educators across schools. In particular, if districts think that having educators with particular pedagogical orientations and experiences might be better at promoting equity, they might also want to concentrate such individuals in traditionally underserved schools. However, such candidates might be attractive to all types of schools and, by virtue of a decentralized hiring process, may end up with placements at schools with relatively more advantaged students. Table 5 presents results from models investigating whether applicant essay themes are differentially associated with being hired in schools with different student populations, conditional on having applied. As in table 3, the outcomes in this table are whether the school in which the educator was hired has above district-average concentrations of students from particular demographic groups.

The results in table 5 have some patterns in common with table 3, but a few notable differences and smaller coefficients as well. Two themes positively predict being hired in schools with large concentrations of many types of students: special services and educators' responsibilities. Only one other theme, experience with cultural-linguistic diversity, predicts a higher probability of being hired in schools with above-average Asian populations, and supportive classroom predicts higher probability of being hired in schools with above-average numbers of students enrolled in free or reduced-price lunch programs. As with the likelihood of applying, writing about nearly every essay theme increases an applicant's likelihood of being hired in schools with above-average populations of Hispanic students over writing about differentiated instruction, excepting family and community engagement. Although structural causes is the only theme to significantly predict applying to above-average black schools, successful applicants to the same schools discuss a blend of classroom climate themes (family and community engagement, supportive classroom, and special services) and educator strategy themes (standards and assessment and educators' responsibilities).[10]

DISCUSSION

This study uses school-district administrative data to examine how educators' beliefs and attitudes about inequality discussed in application essays impact application behavior and hiring outcomes. It uses machine-learning techniques to identify the most common themes across over ten thousand essays and tests whether applicants who write about particular themes are more likely to apply to schools with certain student populations, are rated more highly by the district, and are more likely to get hired. This multifaceted investigation makes three contributions. First, it extends work examining current and preservice educators' attitudes about inequality by investigating equity attitudes among a large sample of educator applicants. Second, it contributes to literature evaluating educator-hiring processes and hiring in other settings to consider the ways in which attitudes that applicants express in their

10. One might worry that prior experiences in the district or in the specific school to which a teacher was reapplying might lead them to write especially well-tailored essays. Supplemental results indicate that the themes discussed do vary somewhat between prior employees and completely new applicants. However, similar themes predict higher screening scores and successful hiring outcomes regardless of prior employment in the district or specific school.

Table 5. Do Essay Themes Predict Teacher Hire in Schools with Particular Characteristics?

	(1)	(2)	(3)	(4)	(5)	(6)
	(Location hired is above district average percent student enrollment)					
	Black	Hispanic	Asian	White	ELL	FRPL
Special services	0.003*	0.003**	0.003*	0.003**	0.004***	0.003**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Family and community engagement	−0.000	−0.000	−0.000	−0.001	0.001	−0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Believe to overcome	0.002*	0.002	0.001	0.002	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Experience with cultural-linguistic diversity	0.001	0.002*	0.002*	0.001	0.003**	0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Standards and assessment	0.002*	0.003**	0.001	0.001	0.003***	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Educators' responsibilities	0.003***	0.004***	0.002	0.003**	0.003***	0.003**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Supportive classroom	0.002*	0.003**	0.001	0.000	0.003***	0.002*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Naming structural causes	0.000	0.001	0.000	0.000	0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Cross-subject strategies	0.001	0.001	−0.000	0.000	0.002*	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Constant	0.011	0.021*	0.011	0.011	0.017	0.014
	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Observations	71,165	66,117	50,676	41,950	72,651	61,158
r^2	0.009	0.010	0.011	0.007	0.008	0.011

Source: Authors' tabulations.

Note: Models condition on having applied to positions at one or more schools with the identified student population. Race is white and multiple subject credential are omitted categories. Standard errors clustered by applicant. All models include controls for writing quality, teacher demographics, education, experience, credentials, and year in which the applicant applied. Schools missing on school characteristics omitted from the analysis. Models also exclude applications to dentralized district positions.

*p < .05; **p < .01; ***p < .001

applications affect hiring outcomes. Third, it harnesses existing administrative data and uses text-as-data empirical techniques to glean new knowledge about beliefs and attitudes, which are difficult to measure and evaluate systematically.

Although one might worry that the essays do not reveal applicants' true beliefs, attitudes, and values about equity and achievement gaps, the application essays do provide insight into applicants' perceptions of what they think an employer would like to hear about such topics. Most applicants write thoughtful statements about how they might address a very real problem facing the district. The wide variety of responses suggests that educators have differing perspectives about how to address inequality, even when trying to impress an employer.

The essays show pronounced differences in themes that are differentially related to hiring outcomes. Distinct profiles emerge that vary by applicant race-ethnicity as well as the student population they aim to teach. Black and Hispanic applicants—groups traditionally marginalized in educational settings—write more about issues around structural inequities facing students, advocate for solutions that stress educators' responsibilities for addressing ineq-

uities, and draw on experiences with cultural and linguistic diversity. Many of these themes are positively related to essay scores and hiring outcomes, regardless of the race of the applicant, relative to differentiated instruction, perhaps because differentiated instruction among marginalized populations may be perceived as a way to separate, track, and withhold educational opportunities. Some of the themes also predict an applicant's interest in teaching in schools with marginalized student populations, but a somewhat different set are associated with actually being hired in such schools.

In contrast, Asian and white applicants—groups traditionally advantaged in this district and schools more broadly—focus more on family and community engagement and differentiating instruction. Applicants seeking to work in schools with more-advantaged students talk less about structural inequity and instead focus on meeting individual student needs through differentiated instruction. However, almost all of the other themes are more strongly related to being hired than differentiated instruction is. In other words, educators' approaches to this issue, and their underlying values, attitudes, and beliefs that inform their responses matter in district hiring decisions.

The degree to which educators cultivate distinct equity-oriented personas or profiles in response to this question is striking. It parallels work by the sociologist Lauren Rivera that examines how students in elite colleges work to cultivate the cultural capital needed to successfully match to high-status private-sector jobs (2016). In our setting, educator applicants work not to brandish their elite cultural bona fides but instead to demonstrate that their dispositions position them to work as change-agents in a district seeking to combat systemic inequality (Bourdieu 2000). Yet, although multiple profiles appear to be desired by the district, not all applicants are equally successful at projecting this particular image. Educators from underrepresented racial backgrounds are particularly good at characterizing themselves in a manner consistent with the district's social-justice aims, but many white applicants also reflect thoughtfully on the ways they would address inequality, and the district appears to value this change-agent cultural capital regardless of the demographic background of the applicant who displays it.

The limitations of this study should be kept in mind when interpreting the results. In particular, we do not observe all stages in the hiring process. Between the district's initial screening and the time that applicants are hired for a position, hiring procedures are largely left to the discretion of individual principals. This makes it difficult to determine what type of interview process occurred and how much additional information principals had beyond the application materials with which to make their hiring decisions. This opacity in turn makes it difficult to determine what information principals used to select one candidate over another. Another limitation is that although structural topic modeling does allow for a careful screening of a large volume of text, the algorithms may overlook important insights that human readers would identify. Human coders might see meaning from particular phrases or topics that machine-learning techniques cannot.

Finally, this study takes place in a district that is extremely cognizant of structural inequality and particularly mindful of its role in creating and combating these inequities. Not all districts have this focus or awareness. Thus, the types of candidates attracted to work in this setting and the types of responses applicants provide to this prompt are likely different than they would be if many other districts asked for the same information. Although the generalizability of these results might be limited, they do provide helpful insights about what individuals think (and the variability in what individuals think) about the achievement gap and equity in a setting that brands itself as being actively engaged in the work of trying to combat it. Many U.S. school districts aim to develop a workforce that puts equity at the center of their work; these results indicate that applicants do have different ideas about how best to do so. Incorporating some type of application essay on this topic can help discern important differences and improve selection processes. Moreover, essay prompts discussing other educational challenges could be implemented to similar effect.

This study identifies important variation in applicants' beliefs about how to address a ma-

jor educational challenge: the achievement gap. It also reveals that such responses influence both application and hiring behaviors in this district. As in this study, data-mining techniques can unlock novel insights from text-rich administrative data and can inform conversations about educator hiring, diversity, retention, and many other policy-relevant issues. This method summarized complexities about educators' beliefs, values, and preferences that are otherwise costly and challenging to collect across a large number of individuals. Such data could be further linked to student and educator records to examine relationships between applicant attitudes, student outcomes, and educator job performance and tenure—relationships we will address in our setting in future work. Similar data in other settings could be matched to other employee outcomes including performance metrics, satisfaction, and retention. In drawing insights from existing administrative data, this relatively new combination of data and methods has the potential to make novel contributions to general knowledge about human thought and behavior as well as innovative improvements in policy and practice, in education and elsewhere.

APPENDIX

The text that follows includes excerpts from applicant essays that exemplify each of the ten most common essay themes.

Special Services

In order to engage all students in their learning process, it is my role as an educator to 1) establish a consistent, comfortable, all-inclusive, stimulating and trust-enhancing environment inside the classroom; 2) to maintain open and trusting ongoing communication with students, their families and caretakers in the classroom and an open-door policy; 3) maintain a school-wide support system with other staff members, all students, and their communities; 4) modify curricula based upon the IEPs of students with special needs and ongoing data based on students' experiences, cultural and language background, interests, progress and behavior; 5) develop appropriate, ongoing, and viable rewards systems for individual students as well as student teams; 6) use caution in applying research-based interventions to extinguish undesirable behavior; 6) work closely with teaching team for optimal teaching and school-wide environment.

Family and Community Engagement

I believe that it is essential to get the whole community involved in the education of our children. This is one of the main reasons I want to work in [district], because I live here. It is important to reach out to parents, do home visits, get them involved in school activities. If the entire family feels a part of the school setting then it will be easier for all of us to work together to close the achievement gap.

Believe to Overcome

Students will not here anything you are saying unless they believe you actually care about them. This comes from having a heart and spirit of humbleness and sincerity. You cannot fool these kids into believing what you are saying is truth unless they know you care about their well-being. Once you as a teacher have established in your heart that you are in this profession for the kids and not for any other reason then as a teacher your presence will convey and show this to the kids. Next you can present the content matter that you have a passion for. I believe it is the teachers job and goal to inspire these kids. There have been many methods and programs in our schools to try and do this, but one cannot inspired unless one is inspired. This means as a education you must be full of courage and strength exhibiting genuine love to these kids no matter how hard they may be on you as a teacher or how negative their personal situation is at home. As a teacher you must believe in yourself and the kids. Energy and love must flow from your being to help these kids because a lot of them do not have many people who really care about them.

Experience with Cultural-Linguistic Diversity

Being multi-lingual is essential to closing the achievement gap. Many [district] students

speak one language at home and another language at school: English. I have spoken Spanish with parents, translators, students, and with my paraprofessionals. I have even learned some Chinese and Arabic to speak with my students with Speech needs. I think the District should provide condensed History lessons for all teachers of all the various cultures of our students: African-American, Cambodian, Vietnamese, Chinese, Latino, European and Southeast Asian cultures. Since I have travelled throughout Mexico, I have some understanding of the culture. Having studied American History, I have some idea of the African-American experience.

Standards and Assessment

It takes an outstanding educator to meet the needs of students and work toward closing the achievement gap. An outstanding educator is constantly and consistently assessing student learning as well as involving students in assessing their own progress. Progressive educators use the information from ongoing formal and informal assessments to guide the content and methodology of instruction to meet the needs of individual students. Documenting and communicating progress to students, parents, and those who work with the student not only conveys to the student the importance of being a self-motivated learner, but also improves professional practice.

Educators' Responsibilities

Additionally, we need to critically examine the practices that are in place within the educational system, despite our best intentions, the very practices that we use daily work against students from minority ethnicities and low SES backgrounds. Utilizing culturally responsive teaching is at the epicenter of closing the achievement gap. Acknowledging our students' diverse backgrounds gives legitimacy to each student's culture, which in turn creates a clear connection between home and school, which is crucial in promoting student achievement. When this approach to instruction is used, learning becomes more appropriate and effective, because instruction stems from, and responds to the students' strengths and existing knowledge. Due to the fact that non-dominant discourses are often ignored or shamed inside classrooms, it is crucial that the voices of students with non-dominant discourses are given focus and emphasis. The very nature and spotlight of spoken academic language in the classroom, needs to be shifted to fit the voices of the students.

Supportive Classroom

It is truly critical for the students themselves to value education and be self motivated. However, the gap will not be closed by the students or the teachers alone; the gap will ultimately be closed by the joining of the school, the families, and the [district] community. By making education a community effort, students will feel supported and understand the importance of their education. Thus my main role, as an educator, is to facilitate the learning of the students. I truly believe in a positive classroom environment in which the students are free to express themselves and learn with a lowered affective filter. I will do this by promoting an encouraging learning environment and showing students the true life value of education.

Naming Structural Causes

In his book Savage Inequalities the education writer Jonathan Kozol describes how urban public school systems in the 1990s struggled to serve the needs of the poorest families in the country, especially in predominantly African-American and immigrant communities. He cites research indicating that the factors most closely correlated with high achievement in school are family income and education level of parents. Thus inner city schools working to improve achievement face the greatest challenge in the nation. Kozol's own research revealed the stark reality that those same schools are funded by the lowest expenditures per student and equipped with the poorest physical resources. He reports on both the deplorable conditions in many impoverished school districts and the contrasting profusion of resources available in neighboring wealthy districts.

Cross-Subject Strategies

I believe that it is very important for the educator to come up with creative strategies that are specifically tailored to each student's needs. . . . I designed and taught nine week project classes with integrated curriculum. The 6th, 7th, and 8th grade students signed up for the project class that they wanted to be in. (Student choice is very important in motivating the student to achieve.) One example of a project was Mystery Play. For Language Arts, the students read and wrote mystery stories. In Math, they did logic and secret codes. In Social Studies, they studied actual cases. In science, they studied forensics. Each project had a final challenge and in this project the final challenge involved writing, producing, and performing a whodunit mystery play using the elements that they had learned.

Differentiated Instruction

I believe that the number one key component in closing the achievement gap in schools today is adapting instruction to tailor to the specific needs of individual students. . . . I try to create interesting lessons and differentiate instruction so I can reach each of my students. . . . It is a well known fact that not all students learn in the same way. Other than basic learning style differences, some students have learning disabilities or impairments which can significantly alter the way they receive and process information. I make a conscious effort to be very aware of things that may cause students to learn differently, and I use that information to help differentiate my instruction and future lessons. One way that I differentiate my instruction is through the use of centers. I like to set up various learning centers during a lesson to provide a variety of learning experiences that will allow students to learn according to their individual learning styles. The centers can include a variety of direct instruction of material, hands on learning activities, historical documents that students can analyze, and a center where students take the knowledge they have gained and create something.

REFERENCES

Alvidrez, Jennifer, and Rhona S. Weinstein. 1999. "Early Teacher Perceptions and Later Student Academic Achievement." *Journal of Educational Psychology* 91(4): 731.

Baker, Courtney N., Marianne H. Tichovolsky, Janis B. Kupersmidt, Mary Ellen Voegler-Lee, and David H. Arnold. 2015. "Teacher (Mis)Perceptions of Preschoolers' Academic Skills: Predictors and Associations with Longitudinal Outcomes." *Journal of Educational Psychology* 107 (3): 805.

Balter, Dana, and William Duncombe. 2006. "Staffing Classrooms: Do Teacher Hiring Practices Affect Teacher Qualifications." Paper presented at the annual meeting of the American Education Finance Association, Denver, Colo. (March 23–25, 2006).

Bettinger, Eric, Jing Liu, and Susanna Loeb. 2016. "Connections Matter: How Interactive Peers Affect Students in Online College Courses." *Journal of Policy Analysis and Management* 35(4): 932–54.

Bireda, Saba, and Robin Chait. 2011. "Increasing Teacher Diversity." *Center for American Progress* (blog). November 9, 2011. Accessed October 25, 2018. https://www.americanprogress.org/issues/education-k-12/reports/2011/11/09/10636/increasing-teacher-diversity.

Bourdieu, Pierre. 2000. "Cultural Reproduction and Social Reproduction." In *The Structure of Schooling: Readings in the Sociology of Education*, edited by Richard Arum and Irene Beattie, 56–68. Thousand Oaks, Calif.: Pine Forde Press.

Boyd, Donald J., Hamilton Lankford, Susanna Loeb, and James Wyckoff. 2005. "The Draw of Home: How Teachers' Preferences for Proximity Disadvantage Urban Schools." *Journal of Policy Analysis and Management* 24(1): 113–32.

Brown, Catherine, and Ulrich Boser. 2017. "Revisiting the Persistent Teacher Diversity Problem." *Center for American Progress* (blog). September 28, 2017. Accessed October 25, 2018. https://www.americanprogress.org/issues/education-k-12/news/2017/09/28/415203/revisiting-persistent-teacher-diversity-problem.

Buehl, Michelle M., and Jori S. Beck. 2015. "The Relationship Between Teachers' Beliefs and Teachers' Practices." In *International Handbook of Research on Teachers' Beliefs*, edited by Helenrose

Fives and Michele Gregoire Gill. New York: Routledge.

Calarco, Jessica McCrory. 2011. "'I Need Help!' Social Class and Children's Help-Seeking in Elementary School." *American Sociological Review* 76(6): 862–82.

California Department of Education. 2015. "Largest & Smallest Public School Districts." CalEdFacts, September 21. Accessed October 25, 2018. http://www.cde.ca.gov/ds/sd/cb/ceflargesmalldist.asp.

———. 2016a. "Facts About English Learners in California." CalEdFacts, September 29. Accessed October 25, 2018. http://www.cde.ca.gov/ds/sd/cb/cefelfacts.asp.

———. 2016b. "Fingertip Facts on Education in California." CalEdFacts. September 29. Accessed October 25, 2018. http://www.cde.ca.gov/ds/sd/cb/ceffingertipfacts.asp.

Coburn, Cynthia E. 2001. "Collective Sensemaking About Reading: How Teachers Mediate Reading Policy in Their Professional Communities." *Educational Evaluation and Policy Analysis* 23(2): 145–70.

———. 2004. "Beyond Decoupling: Rethinking the Relationship between the Institutional Environment and the Classroom." *Sociology of Education* 77(3): 211–44.

Coburn, Cynthia E., Heather C. Hill, and James P. Spillane. 2016. "Alignment and Accountability in Policy Design and Implementation: The Common Core State Standards and Implementation Research." *Educational Researcher* 45(4): 243–51.

DeArmond, Michael, Betheny Gross, and Dan Goldhaber. 2010. "Is It Better to Be Good or Lucky? Decentralized Teacher Selection in 10 Elementary Schools." *Educational Administration Quarterly* 46(3): 322–62.

Dee, Thomas S. 2004. "Teachers, Race, and Student Achievement in a Randomized Experiment." *Review of Economics and Statistics* 86(1): 195–210.

———. 2005. "A Teacher Like Me: Does Race, Ethnicity, or Gender Matter?" *The American Economic Review* 95(2): 158–65.

Downey, Douglas B., and Shana Pribesh. 2004. "When Race Matters: Teachers' Evaluations of Students' Classroom Behavior." *Sociology of Education* 77(4): 267–82.

Egalite, Anna J., and Brian Kisida. 2018. "The Effects of Teacher Match on Students' Academic Perceptions and Attitudes." *Educational Evaluation and Policy Analysis* 40(1): 59–81.

Egalite, Anna J., Brian Kisida, and Marcus A. Winters. 2015. "Representation in the Classroom: The Effect of Own-Race Teachers on Student Achievement." *Economics of Education Review* 45(1): 44–52.

Ferguson, Ronald F. 2003. "Teachers' Perceptions and Expectations and the Black-White Test Score Gap." *Urban Education* 38(4): 460–507.

Fernandez, Roberto M., and Colette Friedrich. 2011. "Gender Sorting at the Application Interface." *Industrial Relations: A Journal of Economy and Society* 50 (4): 591–609.

Fernandez, Roberto M., and Brian Rubineau. 2019. "Network Recruitment and the Glass Ceiling: Evidence from Two Firms." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5(3): 88–102. DOI: 10.7758/RSF.2019.5.3.05.

Fox, Lindsay. 2016. "Seeing Potential: The Effects of Student–Teacher Demographic Congruence on Teacher Expectations and Recommendations." *AERA Open* 2(1): 1–17. DOI: 10.1177/23328584 15623758.

Gamoran, Adam. 2001. "American Schooling and Educational Inequality: A Forecast for the 21st Century." *Sociology of Education* 74:135–53.

Gentzkow, Matthew, Bryan T. Kelly, and Matt Taddy. 2017. "Text as Data." NBER working paper no. w23276. Cambridge, Mass.: National Bureau of Economic Research.

Gershenson, Seth, Cassandra Hart, Constance Lindsay, and Nicholas W. Papageorge. 2017. "The Long-Run Impacts of Same-Race Teachers." IZA discussion paper no. 10630. Bonn: Institute of Labor Economics.

Golann, Joanne W. 2018. "Conformers, Adaptors, Imitators, and Rejecters: How No-Excuses Teachers' Cultural Toolkits Shape Their Responses to Control." *Sociology of Education* 91(1): 28–45.

Goldhaber, Dan, Cyrus Grout, and Nick Huntington-Klein. 2016. "Screen Twice, Cut Once: Assessing the Predictive Validity of Applicant Selection Tools." *Education Finance and Policy* 12(2): 197–223.

Harris, Douglas N., and Carolyn D. Herrington. 2006. "Accountability, Standards, and the Growing Achievement Gap: Lessons from the Past Half-

Century." *American Journal of Education* 112(2): 209–38.

Hollingsworth, Sandra. 1989. "Prior Beliefs and Cognitive Change in Learning to Teach." *American Educational Research Journal* 26(2): 160–89.

Jacob, Brian, Jonah E. Rockoff, Eric S. Taylor, Benjamin Lindy, and Rachel Rosen. 2016. "Teacher Applicant Hiring and Teacher Performance: Evidence from DC Public Schools." NBER working paper no. 22054. Cambridge, Mass.: National Bureau of Economic Research.

Jockers, Matthew. 2017. Syuzhet: Extracts Sentiment and Sentiment-Derived Plot Arcs from Text. Version 1.0.1. Accessed October 25, 2018. https://cran.r-project.org/web/packages/syuzhet/index.html.

Jones, Dawn S. 2012. "How Do I Handle Those Online Application Essays?" *Education Week—Career Corner* (blog). March 28, 2012. Accessed October 25, 2018. http://blogs.edweek.org/topschooljobs/careers/2012/03/how_do_i_handle_those_online_a.html?cmp=SOC-SHR-FB.

Kelly, Sean, Andrew M. Olney, Patrick Donnelly, Martin Nystrand, and Sidney K. D'Mello. 2018. "Automatically Measuring Question Authenticity in Real-World Classrooms." *Educational Researcher* 47(7): 451–64.

King, John B., Jr., Amy McIntosh, and Jennifer Bell-Ellwanger. 2016. "The State of Racial Diversity in the Educator Workforce." Washington: U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service.

Labaree, David F. 1997. "Public Goods, Private Goods: The American Struggle over Educational Goals." *American Educational Research Journal* 34(1): 39–81.

Lindsay, Constance A., and Cassandra M. D. Hart. 2017. "Teacher Race and School Discipline." *Education Next* 17(1). Accessed October 25, 2018. https://www.educationnext.org/teacher-race-and-school-discipline-suspensions-research/.

Liu, Jing. 2018. "Measuring Beneficial Teacher Practices at Scale: A Novel Application of Text-as-Data Methods." Working paper. Providence, R.I.: Brown University.

Ludwig, Jens. 2018. "Administrative Data and Prediction Policy Problems." Paper presented at the Russell Sage Foundation conference Using Administrative Data for Science and Policy. New York (January 18, 2018).

McFarland, Daniel A., Dan Jurafsky, and Craig Rawlings. 2013. "Making the Connection: Social Bonding in Courtship Situations." *American Journal of Sociology* 118(6): 1596–1649.

McGrady, Patrick B., and John R. Reynolds. 2013. "Racial Mismatch in the Classroom: Beyond Black-White Differences." *Sociology of Education* 86(1): 3–17.

Murphey, David. 2014. "The Academic Achievement of English Language Learners: Data for the U.S. and Each of the States." Publication no. 2014–62. Washington, D.C.: Child Trends.

National Center for Education Statistics. 2004. "Schools and Staffing Survey (SASS)." 2004. Accessed October 25, 2018. http://nces.ed.gov/surveys/sass/tables/state_2004_18.asp.

Okonofua, Jason A., and Jennifer L. Eberhardt. 2015. "Two Strikes: Race and the Disciplining of Young Students." *Psychological Science* 26(5): 617–24.

Okonofua, Jason A., David Paunesku, and Gregory M. Walton. 2016. "Brief Intervention to Encourage Empathic Discipline Cuts Suspension Rates in Half among Adolescents." *Proceedings of the National Academy of Sciences* 113(19): 5221–26.

Opfer, V. Darleen, and David Pedder. 2011. "Conceptualizing Teacher Professional Learning." *Review of Educational Research* 81(3): 376–407.

Pajares, M. Frank. 1992. "Teachers' Beliefs and Educational Research: Cleaning up a Messy Construct." *Review of Educational Research* 62(3): 307–32.

Petersen, Trond, and Thea Togstad. 2006. "Getting the Offer: Sex Discrimination in Hiring." *Research in Social Stratification and Mobility* 24(3): 239–57.

Reardon, Sean F. 2011. "The Widening Academic Achievement Gap Between the Rich and the Poor: New Evidence and Possible Explanations." In *Whither Opportunity: Rising Inequality, Schools, and Children's Life Chances*, edited by Greg J. Duncan and Richard Murnane. New York: Russell Sage Foundation.

Reardon, Sean F., and Kendra Bischoff. 2011. "Income Inequality and Income Segregation." *American Journal of Sociology* 116(4): 1092–153.

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.

Rimm-Kaufman, Sara E., Melissa D. Storm, Brook E. Sawyer, Robert C. Pianta, and Karen M. LaParo. 2006. "The Teacher Belief Q-Sort: A Measure of Teachers' Priorities in Relation to Disciplinary Practices, Teaching Practices, and Beliefs about Children." *Journal of School Psychology* 44(2): 141–65.

Rist, Ray. 1970. "Student Social Class and Teacher Expectations: The Self-Fulfilling Prophecy in Ghetto Education." *Harvard Educational Review* 40(3): 411–51.

Rivera, Lauren A. 2016. *Pedigree: How Elite Students Get Elite Jobs*. Princeton, N.J.: Princeton University Press.

Roberts, Margaret E., Brandon M. Stewart, and Dustin Tingley. 2014. "STM: R Package for Structural Topic Models." *Journal of Statistical Software* 10(2): 1–40. Accessed October 25, 2018. https://cran.r-project.org/web/packages/stm/vignettes/stmVignette.pdf.

Rochmes, Jane. 2018. "The Power of Seeing Potential: Teachers' Beliefs about Overcoming Social Disadvantage and Student Achievement." Stanford, Calif.: Stanford University,

Rockoff, Jonah E., Brian A. Jacob, Thomas J. Kane, and Douglas O. Staiger. 2011. "Can You Recognize an Effective Teacher When You Recruit One?" *Education* 6(1): 43–74.

Saavedra, Juan, Dario Maldonado, Lucrecia Santibanez, and Luis Omar Herrera Prada. 2017. "Premium or Penalty? Labor Market Returns to Novice Public Sector Teachers." NBER working paper no. 24012. Cambridge, Mass.: National Bureau of Economic Research.

Senter, R. J., and Edgar A. Smith. 1967. "Automated Readability Index." AMRL-T-66-220. Cincinnati University, Ohio.

Smagorinsky, Peter, Leslie Susan Cook, Cynthia Moore, Alecia Y. Jackson, and Pamela G. Fry. 2004. "Tensions in Learning to Teach: Accommodation and the Development of a Teaching Identity." *Journal of Teacher Education* 55(1): 8–24.

Spillane, James P. 1999. "External Reform Initiatives and Teachers' Efforts to Reconstruct Their Practice: The Mediating Role of Teachers' Zones of Enactment." *Journal of Curriculum Studies* 31(2): 143–75.

Templin, Mildred C. 1957. *Certain Language Skills in Children: Their Development and Interrelationships*. Minneapolis: University of Minnesota Press.

Villegas, Ana Maria, and Jacqueline Jordan Irvine. 2010. "Diversifying the Teaching Force: An Examination of Major Arguments." *Urban Review* 42(3): 175–92.

Wilkins, Jesse L. M. 2008. "The Relationship among Elementary Teachers' Content Knowledge, Attitudes, Beliefs, and Practices." *Journal of Mathematics Teacher Education* 11(2): 139–64.