The (Mis)measure of Schools: How Data Affect Stakeholder Knowledge and Perceptions of Quality

JACK SCHNEIDER

College of the Holy Cross

REBECCA JACOBSEN

Michigan State University

RACHEL S. WHITE

Michigan State University

HUNTER GEHLBACH

University of California at Santa Barbara

Purpose/Objective: *Under the reauthorized Every Student Succeeds Act (ESSA), states and districts retain greater discretion over the measures included in school quality report cards. Moreover, ESSA now requires states to expand their measurement efforts to address factors like school climate. This shift toward more comprehensive measures of school quality provides an opportunity for states and districts to think intentionally about a basic question: What specific information should schools collect and report to their communities?*

Setting: *This study took place in the community surrounding a small, highly diverse urban school district.*

Population/Participants: *Forty-five local residents representing a range of demographic backgrounds participated in a modified deliberative poll with an experimental treatment.*

Intervention/Program/Practice: *We randomly assigned participants into two conditions. In the first, participants accessed the state web portal, which houses all publicly available educational data about districts in the state. In the second condition, participants accessed a customized portal that contained a wider array of school performance information collected by the research team.*

Research Design: *This mixed-methods study used a modified deliberative polling format, in conjunction with a randomized controlled field trial.*

Teachers College Record Volume 120, 050304, May 2018, 40 pages

0161-4681

Data Collection and Analysis: *Participants in both conditions completed a battery of survey items that were analyzed through multiple regressions.*

Findings/Results: *When users of a more holistic and comprehensive data system evaluated unfamiliar schools, they not only valued the information more highly but also expressed more confidence in the quality of the schools.*

Conclusions/Recommendations: *We doubt that more comprehensive information will inevitably lead to higher ratings of school quality. However, it appears—both from prior research, from theory, and from this project—that deeper familiarity with a school often fosters more positive perceptions. This may be because those unfamiliar with particular schools rely on a limited range of data, which fail to adequately capture the full range of performance variables, particularly in the case of urban schools. We encourage future exploration of this topic, which may have implications for school choice, parental engagement, and accountability policy.*

Over the past two decades, the amount of publicly available educational data has exploded. Due primarily to No Child Left Behind (NCLB) and its successor, the Every Student Succeeds Act (ESSA), anyone with an Internet connection can access a state-run data system housing reams of information about districts and schools.

One of the chief aims in developing these systems has been to inform the public. With more information about school quality, it is presumed, parents will become more active in making choices, and communities will exert stronger pressure for accountability. In keeping with this belief, policymakers have expanded public access to school performance data (e.g., Duncan, 2010). And though use of these systems differs across demographic groups, it does appear that educational data do shape stakeholder behavior (Hastings & Weinstein, 2008).

If these information systems are designed to instruct behavior, it seems appropriate to ask how well they inform. Certainly users can learn a great deal from examining the data collected and made available by the state. But what kind of picture do they get of a school? Given the strong orientation of these systems toward standardized test results, it may be that data answer only some questions about school performance. And if that is the case—if the information is partial—these systems may produce biased perceptions of school quality.

Some evidence suggests that state data systems, despite their potential value, have produced a troubling side-effect: undermining public confidence in public education. Americans have long expressed more positive views toward the schools they know well—the schools attended by their own children—as compared with schools in general. But ratings of unfamiliar schools dipped to a new nadir during the NCLB era (Rhodes, 2015). In 2002, the year NCLB was signed into law, 60% of respondents in

the annual *Phi Delta Kappan*/Gallup poll gave the nation's public schools a "C" or a "D" grade (Rose & Gallup, 2002). Thirteen years later, 69% gave the schools a "C" or "D" (Bushaw & Calderon, 2015). Of course, these more negative responses may reflect a clearer sense of reality, or real declines in quality. Yet parents have continued to rate their own children's schools quite positively: The 72% of respondents who gave their children's schools an "A" or a "B" in 2015, for instance, mirrors the 71% who did so in 2002. Such discrepancies present a puzzle. Why do parents view unfamiliar schools so much more pessimistically than they view their own, familiar schools? What information is shaping their views?

If current data systems inform only partially, and if they foster unreasonably negative perceptions, we might question the sufficiency of what those systems include. Current changes in ESSA require state education agencies to incorporate at least one other indicator of school quality or student success—above and beyond students' test scores—in their public reporting. They suggest a variety of measures that could meet this requirement, including student engagement, educator engagement, student access to and completion of advanced coursework, postsecondary readiness, school climate, and safety. The law also requires that parents be included in the development and implementation of new accountability systems, which may further expand measurement systems. A majority of Delaware parents, for instance, expressed strong support for including social-emotional learning, civic attendance, and surveys of parents and students in the state's accountability system (Delaware Department of Education, 2014). Similarly, roughly 90% of California parents want to hold schools accountable for ensuring that children improve their social and emotional skills and become good citizens (PACE/USC Rossier Poll, 2016). By contrast, only 68% of Californians felt that schools should be held accountable for improving students' scores on standardized achievement tests (PACE/USC Rossier Poll, 2016).

To date, however, states have yet to include many of these additional factors valued by the American public (Downey, von Hippel, & Hughes, 2008; Mintrop & Sunderman, 2009; Rothstein, Jacobsen, & Wilder, 2008). Instead, state data systems report chiefly on student standardized test scores, which not only offer a relatively narrow picture of school quality but also tend to be strongly influenced by student background variables. Consequently, they may mislead stakeholders about school quality—for example, portraying schools with large percentages of low-income and minority students as weaker than they are (Davis-Kean, 2005; Reardon, 2011).

One way to test this "differential data" hypothesis would be to randomly assign community members to different types of educational data for the

purpose of evaluating schools. This is exactly the approach we took for a small, diverse urban school district. We wondered: Might a broader and more comprehensive set of data help stakeholders answer more detailed questions about school performance? And, in doing so, might participants see areas of strength currently rendered invisible by existing reporting systems, thereby raising their overall appraisal of school quality?

This article details results from a randomized experiment, in which we used a modified deliberative polling experience to test how parents and community members would respond to a broader array of school performance data. Comparing this group of participants against a control group that relied on the state's webpage for information, we found that the new data system allowed stakeholders to weigh in on a broader range of questions about school quality and to express greater confidence in their knowledge. Additionally, the broader array of data appeared to improve perceptions of unfamiliar schools—producing overall scores that matched those issued by familiar raters.

BACKGROUND

Generally speaking, actors within organizations possess better information about organizational performance than do those on the outside (Arrow, 1969). This discrepancy may pose few problems if information is easily acquired or if the outsiders do not need information about the organization. But when those with a vested interest in organizational performance cannot easily acquire relevant information, they can lose much of their capacity for making rational decisions, as well as their ability to monitor their agents and representatives.

This information discrepancy may be particularly acute in education. Aims in education are multiple, making organizational effectiveness hard to distill (e.g., Eisner, 2001). Given the breadth of educational aims, some values are easier to measure than others (Figlio & Loeb, 2011), and strong performance in one area does not necessarily indicate equally strong performance in another (e.g., Rumberger & Palardy, 2005). Additionally, communication about performance is hindered by the fact that many schooling aims tend to be clustered into abstract concepts (e.g., Jacob & Lefgren, 2007) or described in different ways by different people (e.g., Maxwell & Thomas, 1991).

This informational divide has direct implications for the ability of parents to select schools for their children. Generally, student assignment policies mean that most parents engage in school choice only indirectly—by considering schools when choosing a home. Still, parents do appear to seek out information that will help them structure their decisions.

Research, for instance, indicates that school choices change when parents are provided with performance data (Hastings & Weinstein, 2008; Rich & Jennings, 2015). Yet research also suggests that parents lack sufficient information to make educated choices (Data Quality Campaign, 2016). Moreover, many parents know little about their local school beyond their child's performance, creating challenges for decision-making. Consequently, many parents rely on their social networks for information about schools—information that is of mixed quality and that is inequitably distributed among parents (Hastings, Van Weelden, & Weinstein, 2007; M. Schneider, Teske, Roch, & Marschall, 1997; M. Schneider, Teske, Marshall, & Roch, 1998). This lack of information hinders not only parents' ability to assist their children but also school accountability more broadly (Data Quality Campaign, 2016; Jacobsen & Saultz, 2016).

These information discrepancies also affect public oversight of the schools. Theoretically, communities hold schools accountable for results by exerting pressure on civic and political leaders (Hirschman, 1970; Rhodes, 2015). And laypeople maintain significant power in shaping school budgets and organizing community resources (Epstein, 1995). To succeed in these roles, however, community members need to know how schools are performing on a range of relevant metrics. Though current state data systems provide a great deal of information to the public, they tend to include only a subset of what parents and community members value (Figlio & Loeb, 2011; Rothstein et al., 2008). Consequently, the public's use of data can be difficult to predict and often seems unrelated to the purpose of strengthening school performance (e.g., Goldring & Rowley, 2006; Harris & Larsen, 2015; Henig, 1994).

Finally, the information available to school "outsiders" can shape perceptions about organizational functionality, impacting public support for a public good. Research indicates that satisfaction is an important predictor of the public's willingness to support schools financially (Figlio & Kenny, 2009; Simonsen & Robbins, 2003) and to remain engaged in democratic participation (Lyons & Lowery, 1986; Mintrom, 2001). Insofar as that is true, then, it is important that data accurately reflect reality, particularly given that lower perceptions of performance can erode public confidence and foster feelings of detachment (Jacobsen, Saultz, & Snyder, 2013; Rhodes, 2015; Wichowsky & Moynihan, 2008). In such cases, feedback can lead to a "vicious chain of low trust," wherein declining resources produce lower perceptions of performance, which then further erode trust (Holzer & Zhang, 2004, p. 238).

Educational data systems, it seems, have the power to shape parental choices, community engagement, and public support by equalizing what "insiders" and "outsiders" know about schools. Current

systems, however, appear to present incomplete information about schools. According to Figlio and Loeb (2011), "school accountability systems generally do not cover even the full set of valued academic outcomes, instead often focusing solely on reading and mathematics performance" (p. 387). In equal part, though, distortion occurs because available measures of academic performance tend to correlate with demographic characteristics, especially at the school level (Sirin, 2005). This is a matter of particular concern in urban districts, which serve large populations of students whose background variables tend to predict lower standardized test scores (Davis-Kean, 2005; Reardon, 2011), even if performance on other valued school outcomes is strong (e.g., Rumberger & Palardy, 2005). Given these weaknesses, current data systems appear to fall short in their potential to inform the public and may do some degree of harm in the process.

Our project seeks to explore the effect of more comprehensive school performance data on the public understandings of educational quality. Would a broader set of performance data give the public more valuable information than the existing state data system? Would they rate schools differently as a consequence? Would any of this differ based on familiarity with a school?

METHODS

To understand how school quality information might affect public knowledge and perceptions of local schools, our experiment took the form of a modified deliberative poll. Deliberative polling usually entails taking a representative sample of citizens, providing them with balanced, comprehensive information on a subject, and encouraging reflection and discussion. This polling format is meant to correct a common complaint about many public opinion polls—that respondents, often ill informed, essentially pick an option at random to satisfy the pollster asking the question. The goal of a deliberative poll, then, is to uncover what public opinion would be if people had time, background knowledge, and opportunity for deliberation (Fishkin, 2009). Deliberative polling has shown strong internal and external validity and today represents "the gold standard of attempts to sample what a considered public opinion might be on issues of political importance" (Mansbridge, 2010, p. 55). For our purposes, it also provides an analog to how friends and neighbors learn about schools by exchanging information through various social networks. The model, in short, is ideal for addressing how more robust information might affect views of schools.

In our experiment, the traditional deliberative polling structure was modified slightly to accommodate our research questions, the project's resources, and participants' time constraints. Our poll took place over one afternoon, as opposed to multiple days, and participants were exposed to only one set of data, depending on whether they had been assigned to experimental or control group rather than to competing data sets and to presentations from experts. Although the precise impact of the modifications made to the deliberative poll—namely, the shortened length—on the strength of the study is unknown, we suspect that they have minimal implications for interpreting our findings. For one, the "deliberation" that this model seeks to promote occurs in the "learning, thinking and talking" that occurs during the poll (Fishkin & Luskin, 2005, p. 288). While Fishkin and Luskin (2005) suggested a deliberative poll "typically last[s] a weekend" (p. 288), the "learning, thinking, and talking" that occur between community members in the real world last for a variety of time periods. Furthermore, other researchers have conducted both one-day and multiple-day deliberative polls, with little evidence that length of time is a key factor in changing opinions (Andersen & Hansen, 2007; Eggins, Reynolds, Oakes, & Mavor, 2007; Hall, Wilson, & Newman, 2011).

PARTICIPANTS

The poll was conducted in one relatively small urban school district (approximately 5,000 students) located in New England. We recruited participants by posting information about the study on city websites, social media outlets, and school district media outlets. Community liaisons in the school district facilitated the recruitment of participants from underrepresented communities. Interested parties emailed the researchers their responses to a short demographic background survey. A total of 90 people—a mix of parents and nonparents—completed this initial survey.

In selecting participants for inclusion in the experiment, the research team employed a random, stratified sampling approach with the goal of selecting 50 individuals from the pool of applicants. For the stratification process, we divided potential participants into subgroups by race/ethnicity, gender, age, income, and child in school, first working to match the racial demography of our sample to that of the city. Next, we included all men, as the pool was skewed toward females by a roughly 2-to-1 ratio. From the remaining female volunteers, we sorted by income category and randomly selected participants until all four income categories had roughly equal numbers. We then checked the

number of participants with children in the city's public schools and found an imbalance that was remedied by replacing four public school parents with demographically similar individuals without children in the schools. Because of the modest sample size and constraints of the initial pool of volunteers, the final sample is not perfectly representative of the larger community. However, as Table 1 indicates, the sample does reflect the larger community across multiple important demographic characteristics.

Table 1. Research Participant Demographics and City Demographics

	Control Group	Treatment Group	All Research Participants	Citywide
Total	**23**	**22**	**45**	-
Race/Ethnicity				
White	17 (74%)	15 (68%)	32 (71%)	*73.9%*
African American	2 (9%)	2 (9%)	4 (9%)	*6.8%*
Hispanic	3 (13%)	2 (9%)	5 (11%)	*10.6%*
Asian	1 (4%)	2 (9%)	3 (7%)	*8.7%*
Native American	0	0	0	*0.3%*
Pacific Islander	0	0	0	*0.0%*
Other	0	1 (5%)	1 (2%)	*6.7%*
Language spoken at home				
English	19 (83%)	18 (82%)	37 (82%)	*68.8%*
Language other than English	4 (17%)	4 (18%)	8 (18%)	*31.2%*
Gender				
Male	9 (39%)	8 (36%)	17 (38%)	*49.1%*
Female	14 (61%)	14 (64%)	28 (62%)	*50.9%*
Highest level of school completed*				
Did not complete high school	0	0	0	*11.0%*
High school graduate	2 (9%)	3 (14%)	5 (11%)	*20.0%*
Some college	1 (4%)	2 (9%)	3 (7%)	*9.7%*
Associate's	1 (4%)	0	1 (2%)	*3.7%*
Bachelor's degree	4 (17%)	9 (41%)	13 (29%)	*28.6%*
Graduate degree	15 (65%)	8 (36%)	23 (51%)	*26.9%*

	Control Group	Treatment Group	All Research Participants	Citywide
Annual household income				
Less than $24,999	1 (4%)	1 (5%)	2 (4%)	*18.9%*
$25,000–49,999	5 (22%)	5 (23%)	10 (22%)	*18.1%*
$50,000–74,999	4 (17%)	4 (18%)	8 (18%)	*17.2%*
$75,000–124,999**	5 (22%)	5 (23%)	10 (22%)	*23.1%*
$125,000–199,999**	4 (17%)	5 (23%)	9 (20%)	*16.5%*
Greater than $200,000	3 (13%)	1 (5%)	4 (9%)	*6.2%*
Age				
10–19	1 (4%)	1 (5%)	2 (4%)	*7.2%*
20–29	4 (17%)	3 (14%)	7 (16%)	*20.8%*
30–39	6 (26%)	4 (18%)	10 (22%)	*21.1%*
40–49	5 (22%)	7 (32%)	12 (27%)	*9.5%*
50–59	5 (22%)	7 (32%)	12 (27%)	*9.1%*
60–69	1 (4%)	0	1 (2%)	*6.4%*
70–79	1 (4%)	0	1 (2%)	*3.8%*

* Citywide U.S. Census data refer solely to education level of population 25 years and older.

** Citywide U.S. Census income bands are $75,000–99,999, $100,000–149,999, and $150,000–199,999. The $75,000–124,999 and $125,000–199,999 bands were estimated by splitting the $100,000–149,999 band.

Forty-three of 50 confirmed participants arrived on the day of the poll along with two day-of-event arrivals, bringing the total sample size to 45. All participants who completed the 3-hour polling process, which took place in the spring of 2015, received $100 for their participation.

EXPERIMENTAL CONDITIONS

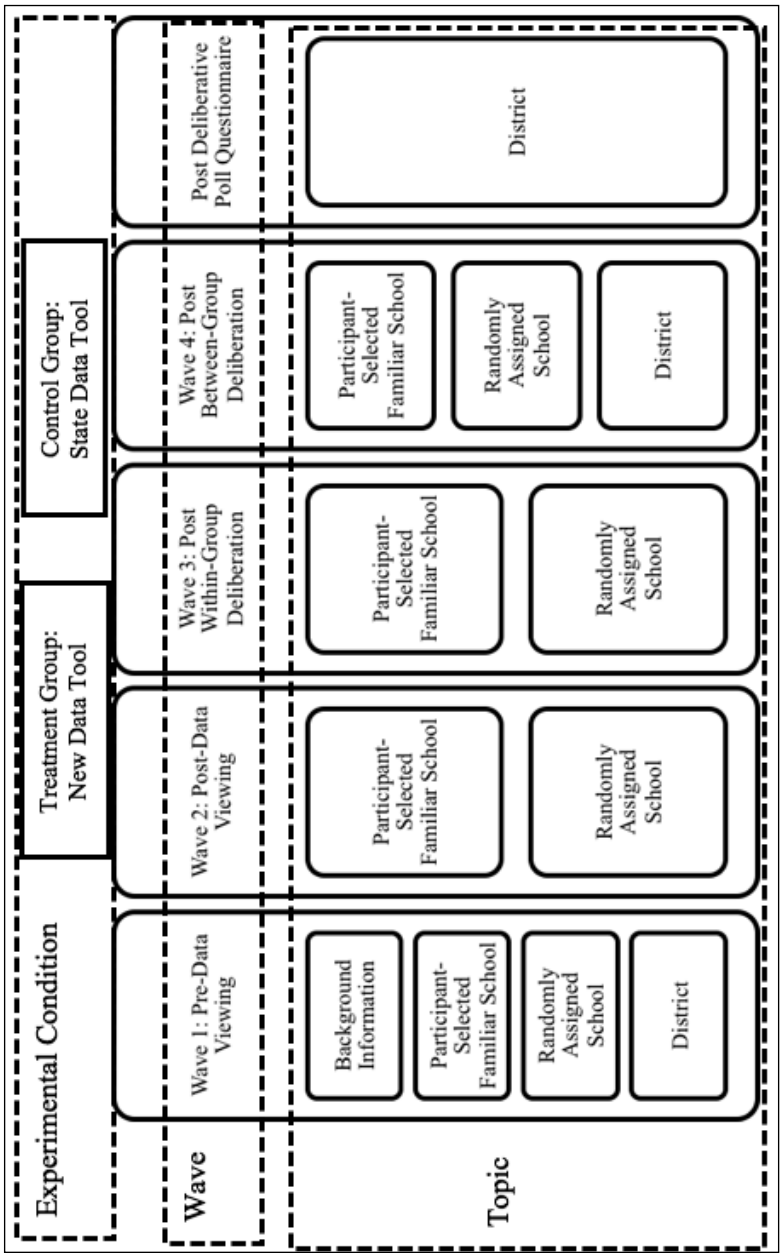
After completing the aforementioned stratification process, we randomly assigned participants within strata to one of two groups: a control group, which would view the state's education data system, and a treatment group, which would view a newly created data tool designed to convey a richer array of relevant school data. Participants selected one school in the district that was most familiar to them to review and report on. After selecting the "familiar school," a computer program randomly selected a second school for participants to review and report on. For both the "familiar" school and the randomly assigned school, participants indicated their familiarity with the school using a 5-point Likert scale ranging from *not familiar at all* to *extremely familiar*.

The control group viewed the data available on the state's website—a site that included both districtwide and school-specific information. Student data (e.g., demographic composition, attendance rates, class size), teacher data (e.g., demographic composition), assessment data (e.g., state assessment results, including percent of students at each achievement level, student growth), and accountability data (e.g., progress toward reducing proficiency gaps by subgroup) were all included in the state's web-based data system. At the school level, benchmarking data were provided for each category relative to the district as a whole and the entire state. These data are typical of many school report cards currently disseminated by state departments of education.

The treatment group viewed data from a newly created digital tool, which was organized around five conceptual school quality categories: Teachers and the Teaching Environment, School Culture, Resources, Indicators of Academic Achievement, and Character and Well-being Outcomes (see Appendix A). These five categories were developed in response to polling on what Americans want their schools to do (e.g., Phi Delta Kappan/Gallup, 2015; Rothstein & Jacobsen, 2006), as well as in response to a review of research relevant to those expressed values (J. Schneider, 2017). The organization of the framework—including categories and subcategories—was then refined through a series of surveys and focus groups with community members.

In terms of navigating the web tool, users could click on any of the five major categories to view relevant subcategories. After clicking the School Culture tab, for instance, users would see data on Safety, Relationships, and Academic Orientation. Clicking a subcategory would take users down another level, to even more detailed information. A click on the Safety tab, for instance, would reveal more specific data on Student Physical Safety and on Bullying and Trust. Data for the tool were drawn from four sources: district administrative records, state-run standardized testing, a student perception survey administered to all students in Grades 4–8, and a teacher perception survey completed by the district's full-time teachers. The surveys were designed by the research team to gather information aligned with the various categories and subcategories (J. Schneider, 2017).

Figure 1. Polling and data viewing across waves



DATA

We recorded four "waves" of participants' perceptions through online Qualtrics surveys: (1) before they viewed any data, (2) after they viewed data in isolation, (3) after they discussed the data in a small group of participants within condition, and (4) after they discussed the data with a mixed group of participants from both treatment and control conditions. As Figure 1 illustrates, participants responded to the same sets of questions about the familiar and randomly assigned schools in each wave of questioning.

Each wave of the survey included school-level "perceived knowledge" questions related to poll participants' perceptions of school climate, effectiveness of teaching, and overall impressions of school quality (see Appendix B for a complete list of questions). Because one of the goals of the experiment was to understand whether either set of data contributed to the building of new knowledge, we asked respondents how accurately they believed they could identify areas in which a particular school needed to improve. And to better understand the relationship between data and future behavior, we asked respondents about their intended actions based on their perceptions of the schools. As shown in Figure 1, the survey at waves 1 and 4 also asked respondents to assess the school district's performance, using adapted versions of the questions those described above. Finally, participants completed a series of demographic questions.

At the conclusion of the polling event, the research team asked participants to complete a follow-up response. Upon exiting the polling location, participants were provided with a self-addressed stamped envelope, as well as a questionnaire that included three question prompts about: (1) what the district is doing well, (2) what recommendations participants would make for improving the schools, and (3) any additional ideas participants might have. Participants were asked to complete and return the questionnaire to the research team within two weeks. We hoped to see whether the quantity and/or quality of participants' responses varied by experimental condition.

DELIBERATIVE PROCEDURES

Participants began the polling session by completing the initial survey prior to viewing any data. After viewing data in isolation, participants then completed the second survey—a procedure intended to determine how new data, on their own, might shape stakeholder knowledge and perception.

Next, participants met in small groups with others who had viewed the same set of data—a procedure designed to allow them to share knowledge, as they might in a real-world setting. Participants began by sharing which schools they viewed and were asked clarifying questions by the other members of their group. The research team answered any process-related questions that groups posed; we did not, however, interpret the data for participants, even when groups disagreed or explicitly asked for such assistance. We then asked participants to discuss the following questions: (1) What were the strengths and weaknesses of each school you viewed? (2) What were the strengths and weaknesses of the district? (3) How did you come to those conclusions? While these questions provided a starting point for the small-group discussions, most groups expanded on them, discussing other issues related to their interests and personal prior knowledge of schools. At the end of this discussion, participants completed their third survey.

After a short break, we placed participants into mixed groups—including members from both control and treatment conditions—for a second deliberative opportunity. Participants again discussed the three questions from their first deliberations. In addition, we asked participants to describe the data they viewed and to discuss what they had learned from these data. The purpose of mixing groups was to see if engagement with either set of data might affect those who had not actually looked at it. In other words, was there a spillover effect? After completing a fourth survey, participants were paid and given the questionnaire with an addressed stamped envelope.

HYPOTHESES

Congruent with recent best practices for experimental studies (Simmons, Nelson, & Simonsohn, 2011), the research team preregistered hypotheses using *Open Science Framework* (see Appendix B for the statement of transparency).

The four hypotheses that follow were informed by the literature discussed in the Background section of this article. Especially worth noting, however, is hypothesis 2, which was informed by research on the relationship between test scores and demography. In the urban district where this research took place, levels of academic proficiency—as measured by standardized test scores—are somewhat lower than state averages at all grade levels. This led us to believe that state data would present a generally negative view of the schools—something not likely to be the case in all districts and which will be explored further in the Discussion section.

Hypothesis 1: As compared with the control group, participants who interacted with the new, more comprehensive data will report valuing the information they received more highly.

Hypothesis 2: As compared with the control group, participants who interacted with the new, more comprehensive data will report higher overall ratings of individual school quality, and of the school district, at the second, third, and fourth time points.

Hypothesis 3: As compared with the control group, participants who interacted with the new, more comprehensive data will manifest greater changes in their opinions as a consequence of the two deliberations.

Hypothesis 4: As compared with the control group, participants who interacted with the new, more comprehensive data will write more in follow-up letters included in the study, expressing broader definitions of school quality.

ANALYSIS

In addition to descriptive statistics and cross-tabs, we conducted analyses of covariance and ordinary least squares (OLS) regressions to statistically examine the relationship between the treatment (i.e., viewing the new, more comprehensive data) and any changes in perception of school quality or valuing of the data. Specific statistical analyses for hypotheses 1, 2, and 3 are listed in Table 2. To ensure the integrity of our findings, these analytic decisions were made prior to any examinations of data and are described in detail in the statement of transparency.

To examine hypothesis 4, the research team measured the length of the postpoll questionnaire responses described earlier and coded those responses for analysis. Specifically, we used a baseline a priori coding scheme, informed by the aforementioned school framework, which we then refined to reflect emergent themes and ideas that had not been captured by the a priori codes. Using this revised scheme, we coded written responses using the constant comparative method (Patton, 2002). The process was both iterative and theory-driven, and it reflected inductive and deductive analysis (Strauss & Corbin, 1998).

Table 2. Statistical Analyses of Stated Hypotheses

Hypotheses	Method of Analyses	Dependent Variables	Main Independent Variable	Controls/ Covariates
Hypothesis 1: As compared with the control group, participants who interacted with the new, more comprehensive data will report valuing the information they received more highly.	OLS regression	Wave 2 amount of information learned from data Wave 2 usefulness of data Wave 3 amount of information learned Wave 3 usefulness of data	Treatment	Familiarity with the school
Hypothesis 2a: As compared with the control group, participants who interacted with the new, more comprehensive data will report higher overall ratings of individual school quality at the second, third, and fourth time points.	OLS regression (wave 2) and fixed effects OLS regression (wave 3 & 4)	Wave 2 randomly assigned school quality ratings Wave 2 familiar school quality ratings Wave 3 randomly assigned school quality ratings Wave 3 familiar school quality ratings Wave 4 randomly assigned school quality ratings Wave 4 familiar school quality ratings	Treatment	Familiarity with the school
Hypothesis 2b: As compared with the control group, participants who interacted with the new, more comprehensive data will report higher overall ratings of school district quality at the fourth time points.	OLS regression	Wave 4 school district quality ratings	Treatment	Average familiarity with the respondents' familiar and randomly assigned school
Hypothesis 3: As compared with the control group, participants who interacted with the new, more comprehensive data will manifest greater changes in their opinions as a consequence of the two deliberations.	Analysis of covariance (ANCOVA)	Wave 4 opinion change ratings of randomly assigned school Wave 4 opinion change ratings of familiar school	Treatment	Familiarity with the school

FINDINGS

Our analysis provided insight into variations that emerged among our participants' perceptions of schools when provided with new, more comprehensive data that rely less heavily on standardized test scores. As evidenced in the next section, users of the new, more comprehensive data system valued this information more highly and became more positive about the quality of schools. Moreover, we found spillover effects: When viewers of the new data deliberated with users of the state data, perceptions of school quality increased for state data users, suggesting that vicarious exposure to this more comprehensive data may have impacted their views. Trends were particularly salient when respondents reported on schools they were previously unfamiliar with.

IMPACT ON INFORMATION VALUE

The first hypothesis was that users would value the new information more highly than the information available on the state website—largely test score data. To examine this, the research team compared self-reported views, examining differences between the treatment and control groups in wave 2 (after the initial viewing of the data) and wave 3 (after within-group deliberation). Across three "information value" questions, participants in the treatment group—those viewing the new, more comprehensive data—consistently reported higher information value (see Table 3).

Table 3. Average Response to Questions Related to Impact of Data on Information Value, by Group

Survey Question	Control Group (State Data)		Treatment Group (New Data)		Treatment–Control Group Difference	
	Wave 2	Wave 3	Wave 2	Wave 3	Wave 2	Wave 3
"How much did you learn from this information about the two schools that was new to you?"	3.0	3.2	3.6	3.5	0.6	0.3
"How confident are you in how much you know about these two schools?"	2.5	2.8	2.9	3.1	0.4	0.3
"How useful was this information in allowing you to form an opinion of these schools?"	2.7	3.2	3.4	3.6	0.7	0.4

The OLS regression analyses provide insight into the statistical significance of these findings. As shown in Table 4, the effect of the treatment after the initial viewing of the data (wave 2) on the amount learned and usefulness of the data was positive and significant. In wave 2, no significant relationship existed between respondents' familiarity with a school and either the amount of information learned from the data or the usefulness of the data. Fixed-effects OLS regressions, taking into account the discussion groups that respondents were in, revealed that the effect of the treatment on the amount learned and usefulness of the data was, again, positive and significant in wave 3.

Table 4. OLS and FE Regression: Relationship Between Value of Learning Experience Variables and Treatment: Unstandardized β and (*SE*)

	Amount Learned From Information That Is New		**Usefulness of Information in Forming an Opinion About Schools**		**Information Value Composite**	
Independent Variable	Wave 2 OLS	Wave 3 FE	Wave 2 OLS	Wave 3 FE	Wave 2 OLS	Wave 3 FE
Treatment	0.886***	0.470**	0.798***	0.615***	0.692***	0.543***
	(0.298)	(0.189)	(0.268)	(0.194)	(0.235)	(0.172)
School familiarity	-0.009	0.027	-0.094	-0.002	-0.049	0.013
	(0.129)	(0.082)	(0.116)	(0.084)	(0.102)	(0.074)

* $p < 0.1$. ** $p < 0.05$. *** $p < 0.01$.

An exploratory analysis provides an additional way to gauge the value of the information to participants across the two groups. Throughout the survey, respondents had the option to select "I don't know" when rating schools. For both the treatment and control groups, the majority of such responses occurred in wave 1—before respondents viewed any data. We examined the extent to which these "I don't know" responses persisted after viewing data, comparing treatment and control groups. The baseline rates at wave 1 were very similar for randomly assigned schools (67% for control, 69% for treatment) and for familiar schools (24% for control, 23% for treatment).

As shown in Table 5, "I don't know" responses decreased substantially more among those viewing the new data tool than among those viewing state data. Among users of the new data tool, "I don't know" responses decreased 80% to 100% for all questions, regardless of whether the school was familiar or randomly assigned to the participant.

Table 5. "I Don't Know" Responses as a Percent of Total Responses by Question, Wave, and Treatment Group

	Question Topic					
	Health of School Climate *n* (%)	**Teaching Effectiveness** *n* (%)	**Student Preparedness for Future** *n* (%)	**Willingness to Recommend School to a Friend** *n* (%)	**Overall Impression of School Quality** *n* (%)	**Ability to Identify Weaknesses** *n* (%)
Random School						
Control (*N* = 24)						
Wave 1	16 (66.7%)	19 (79.2%)	19 (79.2%)	14 (58.3%)	18 (75.0%)	10 (41.7%)
Wave 2	10 (41.7%)	3 (12.5%)	10 (41.7%)	6 (25.0%)	5 (20.8%)	5 (20.8%)
Wave 3	6 (25.0%)	4 (16.7%)	9 (37.5%)	6 (25.0%)	5 (20.8%)	6 (25.0%)
Wave 4	4 (16.7%)	5 (20.8%)	12 (50.0%)	7 (29.2%)	5 (20.8%)	6 (25.0%)
Change	-75.0%	-73.7%	-36.8%	-50.0%	-72.2%	-40.0%
Treatment (*N* = 22)						
Wave 1	17 (77.3%)	16 (72.7%)	17 (77.3%)	13 (59.1%)	15 (68.2%)	13 (59.1%)
Wave 2	1 (4.6%)	1 (4.6%)	5 (22.7%)	1 (4.6%)	1 (4.6%)	3 (13.6%)
Wave 3	1 (4.6%)	1 (4.6%)	3 (13.6%)	0 (0.0%)	1 (4.6%)	1 (4.6%)
Wave 4	1 (4.6%)	1 (4.6%)	2 (9.1%)	0 (0.0%)	1 (4.6%)	1 (4.6%)
Change	-94.1%	-93.8%	-88.2%	-100.0%	-93.3%	-92.3%
Familiar School						
Control (*N* = 24)						
Wave 1	6 (25.0%)	5 (20.8%)	10 (41.7%)	2 (8.3%)	4 (16.7%)	7 (29.2%)
Wave 2	4 (16.7%)	0 (0.0%)	4 (16.7%)	2 (8.3%)	0 (0.0%)	2 (8.3%)
Wave 3	4 (16.7%)	0 (0.0%)	5 (20.8%)	0 (0.0%)	1 (4.2%)	3 (12.%)

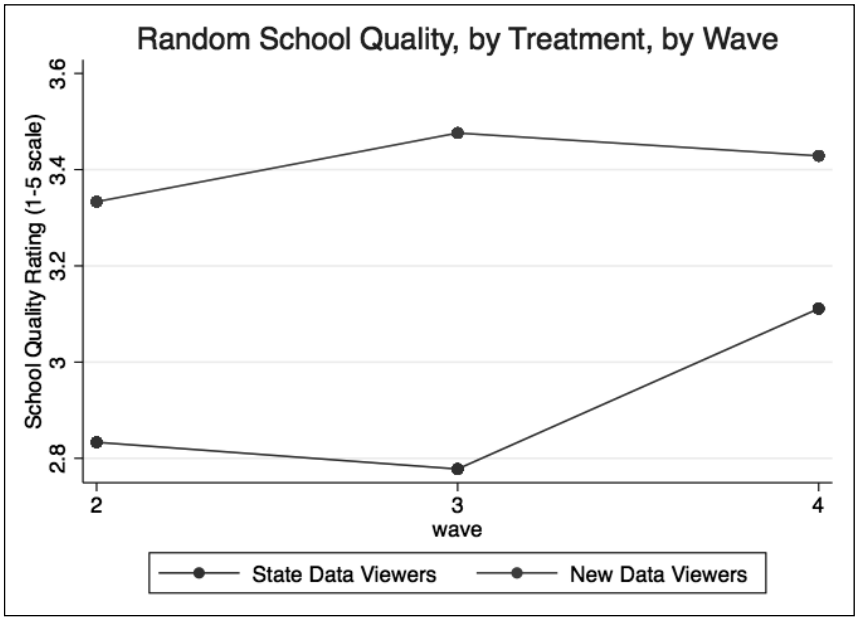
	Question Topic					
	Health of School Climate *n* (%)	**Teaching Effectiveness** *n* (%)	**Student Preparedness for Future** *n* (%)	**Willingness to Recommend School to a Friend** *n* (%)	**Overall Impression of School Quality** *n* (%)	**Ability to Identify Weaknesses** *n* (%)
Wave 4	4 (16.7%)	2 (8.3%)	7 (29.2%)	1 (4.2%)	1 (4.2%)	3 (12.%)
Change	-33.3%	-60.0%	-30.0%	-50.0%	-75.0%	-57.1%
Treatment (N = 22)						
Wave 1	3 (27.3%)	7 (31.8%)	8 (36.4%)	3 (13.6%)	4 (18.2%)	5 (22.7%)
Wave 2	0 (0.0%)	0 (0.0%)	3 (13.6%)	0 (0.0%)	0 (0.0%)	1 (4.6%)
Wave 3	0 (0.0%)	0 (0.0%)	2 (9.1%)	0 (0.0%)	0 (0.0%)	1 (4.6%)
Wave 4	0 (0.0%)	0 (0.0%)	3 (13.6%)	0 (0.0%)	0 (0.0%)	1 (4.6%)
Change	-100.0%	-100.0%	-62.5%	-100.0%	-100.0%	-80.0%

In sum, it appears that users of the new, broader set of data not only valued this information more highly—indicating that they learned more from it and had more confidence in their own knowledge—but also expressed more confidence in their knowledge by selecting the "I don't know" option less frequently than those relying on state-provided data.

IMPACT ON PERCEPTION OF SCHOOL AND DISTRICT QUALITY

The research team also hypothesized that, given the demography of the district in question, treatment participants viewing the new comprehensive data would express more positive views of school and district quality than those expressed by control participants viewing the state data. We expected this because of the strong correlation between standardized test scores and the demographic background of students. Because the district we examined has several schools with large populations of lower income students and non-native English speakers, users looking primarily at test score data might issue lower ratings of school quality for these schools. But because other measures of school quality are less tightly correlated with demographics, we expected that participants who viewed these data would see areas of strength not revealed by test scores alone.

Figure 2. Randomly Assigned School Quality Ratings by Treatment, by Wave



We found positive evidence for this hypothesis, but only with regard to the schools that were unfamiliar to participants. After viewing data for unfamiliar, randomly assigned schools, respondents in the treatment group expressed more positive views of performance than those in the control group (treatment = 3.3 vs. control = 2.9). This gap widened even further in wave 3, after participants discussed the data during their first deliberation, with participants in the treatment group growing more positive about the performance of their randomly assigned school (treatment = 3.5 vs. control = 2.8). Table 5 and Figure 2 also suggest that, after new data viewers (treatment group) talked with state data viewers (control group) in wave 4, the state data viewers' school quality ratings increased (treatment = 3.5 vs. control = 3.1). This may indicate that the effects of the new data system travel beyond those who engage directly with it.

Interestingly, as shown in Figure 2, opinions about school quality for familiar schools appeared to be consistent for both treatment and control groups across all four time points (treatment: wave 1 = 3.6, wave 2 = 3.4, wave 3 = 3.4, wave 4 = 3.3; control wave 1 = 3.8, wave 2 = 3.4, wave 3 = 3.5,

wave 4 = 3.7). We found no significant differences in the ratings issued by treatment and control groups to their familiar schools (wave 1, $t = 0.572$, $p = 0.571$; wave 2, $t = 0.229$, $p = 0.820$; wave 3, $t = 0.586$, $p = 0.561$; wave 4, $t = 1.213$, $p = 0.232$).

Analysis of opinions about randomly assigned schools is complicated by the overwhelming number of "I don't know" responses issued in wave 1. Though not surprising, as participants were mostly unfamiliar with these schools, this trend rendered it impossible to make any inferences from wave 1. That said, interesting patterns did emerge across treatment and control groups across waves 2, 3, and 4. As shown in Figure 2, treatment participants had higher perceptions of school quality than did control participants. And, as shown in Table 6, the effect of the treatment is statistically significant in both waves 2 and 3. Although significant differences disappear by wave 4—after mixed-group deliberation—this shift is not due to a decline in perception among treatment participants. Instead, as Figure 2 illustrates, control participants become more positive in their opinions about their randomly assigned schools after talking in small groups with treatment participants.

Table 6. OLS and FE Regression: Relationship Between School Quality and Treatment, Familiar & Random Schools: Unstandardized β and (*SE*)

	School Quality – Randomly Assigned School			**School Quality – Familiar School**		
	Wave 2 OLS	**Wave 3 FE**	**Wave 4 FE**	**Wave 2 OLS**	**Wave 3 FE**	**Wave 4 FE**
Treatment	0.500*	0.698**	0.317	-0.071	-0.182	-0.364
	(0.304)	(0.335)	(0.329)	(0.311)	(0.311)	(0.300)

* $p < 0.1$. ** $p < 0.05$. *** $p < 0.01$.

We also conducted OLS regression to examine the relationship between the treatment and respondents' ratings of overall school district quality at the final time point of the deliberative poll. Holding constant respondents' opinions of district quality at the beginning of the poll, the effects of the treatment are not statistically significant (t = -0.050, p = 0.958). However, respondents' previous perceptions of the quality of the district is a significant predictor of their perception of the quality of the district at the end of the deliberative poll ($b = 0.76$, $p < 0.01$).

In sum, a broader set of performance data produced more positive ratings for unfamiliar schools. Interestingly, the higher scores that the treatment group gave to randomly assigned schools mirrored the scores issued to the familiar schools. Moreover, it appears that the broader set of school

performance data may have had spillover effects. After cross-treatment deliberation (wave 4), users of the state data system rated the quality of their randomly assigned schools more highly. With regard to familiar schools, we found little evidence of any change in perspective among both the treatment and control groups. It may be that performance data, however comprehensive, only reaffirms what people already know in some general way about their familiar schools. Or, it may be that existing impressions are more difficult to change. In either case, our data are congruent with our second hypothesis, but only for randomly assigned schools.

IMPACT ON PERCEPTION VIA DELIBERATION

Our third hypothesis posited that participants interacting with the more comprehensive data would manifest greater changes in their opinions as a consequence of deliberation. Recall that participants had two opportunities to deliberate about school performance and the data itself.

Contrary to the hypothesized impact, we found little to no influence from the first deliberation, in which participants spoke with others who had viewed the same data. This was true among the treatment group (familiar school $t = 0.000$, $p = 1.000$; random school $t = -1.453$, $p = 0.163$), as well as the control group (familiar school $t = -0.568$, $p = 0.576$; random school $t = 1.382$, $p = 0.189$). It seems that talking with others after viewing the same data sources did little to change performance perceptions among our participants.

Things changed a bit in the second deliberation, however, when participants from the treatment and control groups were mixed together and encouraged to share details about the data they viewed, as well as about the conclusions they drew.

Congruent with our other findings, it appears that one's familiarity with the school is a main driver of whether the new, more comprehensive data will have an impact. Inasmuch as that is the case, the wave 3-to-wave 4 cross-treatment deliberations did not affect ratings issued to familiar schools (treatment $t = 0.371$, $p = 0.715$; control $t = -1.000$, $p = 0.329$). It also seems that the deliberation did not change the opinions of those in the treatment condition who were rating randomly assigned schools.

But cross-treatment deliberation did appear to impact the ratings of the randomly assigned schools for those in the control group. After speaking with members of the treatment group, control group participants expressed slightly higher impressions of school performance for their randomly assigned school ($t = -1.775$, $p = 0.096$) despite not having viewed the data themselves.

IMPACT ON BREADTH OF "SCHOOL QUALITY" DEFINITIONS

Finally, we hypothesized that participants in the treatment group would express not only more positive impressions of school quality (as examined earlier with the survey data) but also a broader conceptualization of school quality.

A total of 46% of all participants returned responses to the follow-up questions that were given to them at the end of the deliberative poll. Roughly equal numbers of participants in the control group and treatment group returned responses (control = 11 of 24 vs. treatment = 10 of 22). And, contrary to our hypothesis, we found few differences between treatment and control groups in the length of the follow-up letters or the conceptualization of school quality presented in the letters.

We did, however, find some small but consistent differences in the responses, which seemed to reflect the nature of the data presented to them in the intervention. For example, those in the control group were more likely to mention subgroups of students and frequently cited standardized tests, sometimes even lamenting the emphasis on testing. The treatment group, on the other hand, often referenced measures that were only available through the new data tool.

Given the limited number of responses, any conclusions should be interpreted cautiously. That said, we believe that these findings suggest a fruitful avenue for future research into the longitudinal impact of data.

DISCUSSION

For over a decade, education leaders have pursued policies aimed at increasing the amount of data available to the public—data that can be used to judge the quality of the public schools. In theory, providing information will enable higher levels of public engagement and oversight among both parents and concerned citizens. Most available data comes from standardized tests—a relatively narrow range of information that may misrepresent the quality of particular schools. Thus, although the impact of these data systems is not entirely clear, it seems that any potential to empower and engage stakeholders has not been fully realized.

In our experiment, we attempted to uncover how more comprehensive information might impact public views of schools. This is a matter of policy significance, and one at the heart of an enduring mystery—why do Americans rate their local schools so positively while they deplore the state of public schools nationally? As federal law opens the door to new forms of measurement, the matter is also one of increasing policy relevance, and one that leaders in many states are already considering. Our

experiment, though modest in scale, seems to shed some light on the issue, and it may even offer some direction to policy leaders. Specifically, it appears to suggest that if we want to strengthen educational information systems, we must address not only the *amount* of data available but also the *types* of data available.

EMPOWERING STAKEHOLDERS

Our results suggest that providing more comprehensive performance data can help parents and community members learn more about a school's strengths and weaknesses, particularly in the case of unfamiliar schools. Specifically, those with little familiarity with a school were more confident in their knowledge when using the new tool and were better able to weigh in on a wider range of questions. Such results may impact the ability of parents to make informed school choices and empower communities to more effectively advocate for their schools.

But raters of unfamiliar schools were not the only ones who appeared to benefit from a more comprehensive set of data. Although familiar raters working with the new data did not generally change their overall impressions of school quality, they did express greater confidence in their knowledge and less frequently selected "I don't know" when asked direct questions about school performance. Thus, although those familiar with a school may already understand its strengths and weaknesses in a holistic sense, a more comprehensive set of data may better empower them as advocates—giving them specific, consistent, and quantifiable information to supplement their more general qualitative understandings.

CLOSING THE PERCEPTION GAP

Americans consistently issue much higher ratings to the schools they are most familiar with (e.g., Phi Delta Kappan, 2015)—a persistent enigma in education polling. One possible explanation for this is that stakeholders may be influenced by what might be termed a "home team bias," ignoring data to cling to positive impressions. Research from psychology, for instance, supports this idea that people develop an affinity for those things they are more familiar with (Zajonc, 2001). At the same time, however, the public has demonstrated a generally accurate perception of how children in local schools are performing (e.g., West, 2014). An alternative explanation is that the higher ratings given to familiar schools may reflect a fuller account of performance. In other words, raters of familiar schools may take other information into account, along with test scores, thereby arriving at more balanced assessments. As others have documented, parents often refer to "the feel" of a building when

describing performance (Mandinach & Miskell, 2017)—including factors like school safety, the supportiveness of the learning environment, student engagement levels, and opportunities to be creative and engage in exploration. Until now, only those familiar with a school would have access to such information.

In our study, users of the more comprehensive data issued significantly higher ratings to unfamiliar schools than did users of the state data system. Their scores, which mirrored those issued by familiar raters, suggest that a broader range of data may help address the perception gap between those who are familiar with a school and those who are not. Of course, such gaps may not exist everywhere. Specifically, perception gaps may exist only in districts with lower than average test scores, like the one in which this study was conducted. It may also be true, at least in the case of some schools, that low test scores are reflective of larger, systemic problems. In that case, additional data would reaffirm impressions generated by standardized test scores. Nevertheless, a large number of schools likely suffer from perceptions that do not align with their true quality. In those cases, more comprehensive data might make a significant difference.

IMPROVING WORD-OF-MOUTH

Word-of-mouth is historically one of the leading ways that parents and community members obtain information about school quality. Yet it is unclear whether word-of-mouth can serve as an accurate and reliable source of knowledge. It might be possible, for instance, that simplified messages will have an impact via word-of-mouth, even if they are inaccurate. As discussed earlier, however, that appears not to have been the case in this experiment. After engaging in cross-talk discussion with users of the new data system, participants working with state data had significantly higher perceptions of their randomly assigned schools. Additionally, as the research team observed, these discussions did not revert to simplistic assertions; rather, conversations were generally robust in nature and tended to incorporate a wide range of data.

This is a promising finding worth exploring further because it seems to indicate that new information about school quality, even if not consumed directly, can influence public opinion. Though more robust data alone would not uniformly transform word-of-mouth into a reliable source of information about schools, such data might expand the base of evidence circulating in conversations among the public.

LIMITATIONS

Despite our efforts to cultivate a representative sample, our participants ultimately consist of those willing to spend their Saturdays reviewing school performance data. It is impossible to know for certain how this self-selected sample differs from average citizens within the school district. It does seem likely that our participants are more interested in the city's schools, and it is possible that such interest is fueled by a high level of either skepticism or support. Still, the nature of this experiment makes many of the imperfections in the representativeness of the sample relatively inconsequential. Additionally, we found no clear pattern of bias in our sample. So, although it remains unknown whether comprehensive data would have an equally large impact on less interested residents, it is not obvious that the impact would be markedly different.

It is also worth noting that our experiment was rather small in scale. This experiment produced some rich data. However, it also drew on a limited number of participants ($n = 45$). Insofar as that is the case, we are cautious not to draw strong causal claims.

CONCLUSION

In the age of accountability, states and school districts have poured enormous resources into the creation and dissemination of data on school quality. A tremendous amount of information is now available to the public. Still, questions remain about how parents and local community members use this information, as well as about what the impacts of that use are.

The new revision to the Elementary and Secondary Education Act—the Every Student Succeeds Act—will likely prompt states and districts to revisit their information systems. And we see great potential in this revision process. Certainly it is possible to go through the motions, merely adding a new data point and continuing on with business as usual. Yet there is also an opening to create more comprehensive systems that better inform stakeholders—empowering them to make better decisions and to engage in more effective advocacy. Additionally, such systems might lay the groundwork for policies that even further expand the powers of parents and community members—from intradistrict choice models to systems of co-governance.

As our experiment suggests, more comprehensive data systems might also improve public perceptions of unfamiliar schools, at least with regard to those with lower than average standardized test scores. Given that most parents already rate their children's schools highly, this may seem a matter of relatively small importance because those most intimately involved in a school—families sending their children there—already understand

the school's quality in some general fashion. We must recall, however, that many families rely on data—whether by accessing a state data system, reading about outcomes in the newspaper, or hearing about results via word-of-mouth—when making high-stakes decisions about where to live and where to send their children to school. Biased measures of school quality, then, may exacerbate segregation patterns by steering well-resourced and quality-conscious parents away from perfectly good schools, and, in doing so, they may enact a self-fulfilling prophecy by concentrating inequality. Moreover, public schools rely on the support of all citizens, not just those with children. As our experiment suggests, more comprehensive systems may both empower and strengthen commitment to public schools by revealing areas of strength not discernible from test scores alone.

Of course, more information will not lead inexorably to more positive perceptions of all unfamiliar schools. In the case of schools that have prioritized test scores over other kinds of outcomes and processes, for instance, more robust data might actually *depress* perceptions of school quality. Seeing that a school is succeeding in one dimension but not in many others might cause parents and community members to reevaluate it. Yet here, too, the creation of more robust systems might accomplish a great deal—by restoring balance to a school's mission.

Educational data systems hold great potential for engaging public stakeholders and empowering them to act in ways that strengthen schools. But to realize that potential, these systems must first be informative. To achieve that, policy makers must work to incorporate a broader range of measures into the data offered to the public. Specifically, they must build systems that align with the public's vision of a good school and not merely with a single metric. They must measure what matters, and they must measure with care.

REFERENCES

Andersen, V. N., & Hansen, K. M. (2007). How deliberation makes better citizens: The Danish Deliberative Poll on the euro. *European Journal of Political Research*, *46*(4), 531–556.

Arrow, K. J. (1969). The organization of economic activity: Issues pertinent to the choice of market versus nonmarket allocation. *The Analysis and Evaluation of Public Expenditure: The PPB System*, *1*, 59–73.

Bushaw, W. J., & Calderon, V. J. (2015). *The 47th annual PDK/Gallup poll of the public's attitudes toward the public schools*. Bloomington, IN: PDK International.

Data Quality Campaign. (2016). *How data empowers parents*. Retrieved from http://dataqualitycampaign.org/resource/data-empowers-parents/

Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment. *Journal of Family Psychology, 19*(2), 294–304.

Delaware Department of Education. (2014). *2014 Delaware School Accountability Community Survey*. Retrieved from http://dedoe.schoolwires.net/site/Default.aspx?PageType=3&DomainID=38&PageID=106&ViewID=047e6be3-6d87-4130-8424-d8e4e9ed6c2a&FlexDataID=12504

Downey, D. B., von Hippel, P. T., & Hughes, M. (2008). Are "failing" schools really failing? Using seasonal comparison to evaluate school effectiveness. *Sociology of Education, 81*(3), 242–270.

Duncan, A. (2010, August 25). Secretary Arne Duncan's remarks at the Statehouse Convention Center in Little Rock, Arkansas. Retrieved from http://www.ed.gov/news/speeches/secretary-arne-duncans-remarks-statehouse-convention-centerlittle-rock-arkansas

Eggins, R. A., Reynolds, K. J., Oakes, P. J., & Mavor, K. I. (2007). Citizen participation in a deliberative poll: Factors predicting attitude change and political engagement. *Australian Journal of Psychology*, *59*(2), 94–100.

Eisner, E. W. (2001). What does it mean to say a school is doing well? *Phi Delta Kappan*, *82*(5), 367–372.

Epstein, J. L. (1995). School/family/community partnerships. *Phi Delta Kappan*, *76*(9), 701–712.

Figlio, D., & Kenny, L. (2009). Public sector performance measurement and stakeholder support. *Journal of Public Economics*, *93*(9–10), 1069–1077.

Figlio, D. N., & Loeb, S. (2011). School accountability. In E. A. Hanushek, S. J. Machin, & L. Woessmann (Eds.), *Handbooks in economics: Economics of education* (Vol. 3, pp. 383–421). Amsterdam, the Netherlands: Elsevier.

Fishkin, J. (2009). *When the people speak: Deliberative democracy and public consultation*. New York, NY: Oxford University Press.

Fishkin, J. S., & Luskin, R. C. (2005). Experimenting with a democratic ideal: Deliberative polling and public opinion. *Acta Politica, 40*(3), 284–298.

Goldring, E., & Rowley, K. J. (2006, April). *Parent preferences and parent choices: The public-private decision about school choice*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Hall, T. E., Wilson, P., & Newman, J. (2011). Evaluating the short-and long-term effects of a modified deliberative poll on Idahoans' attitudes and civic engagement related to energy options. *Journal of Public Deliberation*, *7*(1).

Harris, D. N. & Larsen, M. F. (2015). *What schools do families want (and why)? New Orleans families and their school choices before and after Katrina* [Policy brief]. New Orleans, LA: Education Research Alliance for New Orleans.

Hastings, J. S., Van Weelden, R., & Weinstein, J. M. (2007). Preferences, information, and parental choice behavior in public school choice. *NBER Working Paper, 12995*.

Hastings, J. S., & Weinstein, J. M. (2008). Information, school choice, and academic achievement: Evidence from two experiments. *Quarterly Journal of Economics, 123*(4), 1373–1414.

Henig, J. R. (1994). *Rethinking school choice: Limits of the market metaphor*. Princeton, NJ: Princeton University Press.

Hirschman, A. O. (1970). *Exit, voice, and loyalty: Responses to decline in firms, organizations, and states*. Cambridge, MA: Harvard University Press.

Holme, J. J. (2002). Buying homes, buying schools: School choice and the social construction of school quality. *Harvard Educational Review, 72*(2), 177–206.

Holzer, M., & Zhang, M. (2004). Trust, performance, and the pressures for productivity in the public sector. In M. Holder & S. H. Lee (Eds.), *The public productivity handbook* (2nd ed., pp. 215–229). New York, NY: Marcel Dekker.

Jacob, B. A., & Lefgren, L. (2007). What do parents value in education? An empirical investigation of parents' revealed preferences for teachers. *Quarterly Journal of Economics, 122*(4), 1603–1637.

Jacobsen, R., Saultz, A., & Snyder, J. W. (2013). When accountability strategies collide: Do policy changes that raise accountability standards also erode public satisfaction? *Educational Policy, 27*(2), 360–389.

Jacobsen, R., & Saultz, A. (2016). Will performance management restore citizens' faith in public education? *Public Performance & Management Review, 39*(2), 476–497.

Lyons, W. E., & Lowery, D. (1986). The organization and political space and citizen responses to dissatisfaction in urban communities: An integrative model. *Journal of Politics, 48*(2), 321–346.

Mandinach, E. B., & Miskell, R. C. (2017). *Focus groups to support the development and utility of EdWise for parental stakeholders: Findings from the parental focus groups*. San Francisco, CA: WestEd.

Mansbridge, J. (2010). Deliberative polling as the gold standard. *The Good Society, 19*(1), 55–62.

Maxwell, T. W., & Thomas, A. R. (1991). School climate and school culture. *Journal of Educational Administration, 29*(2), 72–82.

Mintrom, M. (2001). Educational governance and democratic practice. *Educational Policy, 15*(5), 615–642.

Mintrop, H., & Sunderman, G.L. (2009). Predictable failure of federal sanctions-driven accountability for school improvement—And why we may retain it anyway. *Educational Researcher, 38*(5), 353–364.

Phi Delta Kappan/Gallup. (2015, September). *The 47th Annual PDK/Gallup Poll of the public's attitudes toward the public schools*. Bloomington, IN: PDK International.

Policy Analysis for California Education (PACE) and University of Southern California (USC) Rossier School of Education. (2016). *Fifth annual PACE/USC Rossier Poll*. Tulchin Research and Moore Information [Distributor]. Retrieved from http://www.edpolicyinca.org/polls

Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In G. J Duncan & R. J. Murnane (Eds.), *Whither opportunity?* (pp. 91–116). New York, NY: Russell Sage Foundation.

Rhodes, J. H. (2015). Learning citizenship? How state education reforms affect parents' political attitudes and behavior. *Political Behavior, 37*(1), 181–220.

Rich, P. M., & Jennings, J. L. (2015). Choice, information, and constrained options: School transfers in a stratified educational system. *American Sociological Review, 80*(5), 1069–1098.

Rose, L. C., & Gallup, A. M. (2002). *The 34th annual PDK/Gallup Poll of the public's attitudes toward the public schools*. Bloomington, IN: PDK International.

Rothstein, R., & Jacobsen, R. (2006). The goals of education. *Phi Delta Kappan, 88*(4), 264–272.

Rothstein, R., Jacobsen, R., & Wilder, T. (2008). *Grading education: Getting accountability right*. New York, NY: Teachers College Press.

Rumberger, R. W., & Palardy, G. J. (2005). Test scores, dropout rates, and transfer rates as alternative indicators of high school performance. *American Educational Research Journal, 42*(1), 3–42.

Schneider, J. (2017). *Beyond test scores: A better way to measure school quality*. Cambridge, MA: Harvard University Press.

Schneider, M., Teske, P., Marshall, M., & Roch, C. (1998). Shopping for schools: In the land of the blind, the one-eyed parent may be enough. *American Journal of Political Science, 42*(3), 769–793.

Schneider, M., Teske, P., Roch, C., & Marschall, M. (1997). Networks to nowhere: Segregation and stratification in networks of information about schools. *American Journal of Political Science, 41*(4), 1201–1223.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359–1366.

Simonsen, B., & Robbins, M. D. (2003). Reasonableness, satisfaction, and willingness to pay property taxes. *Urban Affairs Review, 38*(6), 831–854.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417–453.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Procedures and techniques for developing grounded theory*. Thousand Oaks, CA: Sage.

West, M. (2014, October 23). *Why do Americans rate their local public schools so favorably?* The Brown Center Chalkboard Series. Washington, DC: Brookings Institution.

Wichowsky, A., & Moynihan, D. (2008). Measuring how administration shapes citizenship: A policy feedback perspective on performance management. *Public Administration Review, 68*, 908–920.

Zajonc, R. B. (2001). Mere exposure: A gateway to the subliminal. *Current Directions in Psychological Science, 10*(6), 224–228.

APPENDIX A

New Data Set Measurement Information

Main Category	Subcategory	Measure	Method of Measurement
Teachers & the Teaching Environment	1A: Knowledge and Skills of Teachers	Professional qualifications	Administrative data and teacher survey
		Effective practices	Student survey
		Teacher temperament	Student survey
		Teacher turnover	Administrative data
	1B: Teaching Environment	Support for teaching development and growth	Administrative data and teacher survey
		Effective leadership	Teacher survey
School Culture	Safety	Student physical safety	Student survey
		Bullying/trust	Teacher survey
	Relationships	Sense of belonging	Student survey
		Student/teacher relationships	Student survey
	Academic Orientation	Attendance and graduation	Administrative data
		Academic press	Student survey
Resources	Facilities and Personnel	Physical spaces and materials	Administrative data and teacher survey
		Content specialists and support staff	Administrative data and teacher survey
	Curricular Resources	Curricular strength and variety	Teacher survey
		Class size	Administrative data and teacher survey
	Community Support	Family/school relationships	Teacher survey
		Community involvement and external partnerships	Teacher survey

Main Category	Subcategory	Measure	Method of Measurement
Indicators of Academic Learning	Performance	Test score growth	Administrative data
		Portfolio/alternative assessments	Teacher survey
	Student Commitment to Learning	Engagement in school	Student survey
		Value of learning	Student survey
	Critical Thinking	Problem solving emphasis	Teacher survey
		Problem-solving skills	N/A
	College and Career Readiness	College-going	N/A
		College performance	N/A
Character and Well-Being Outcomes	Civic Engagement	Understanding others	Student survey
		Appreciation for diversity	Student survey
	Work Ethic	Perseverance and determination	Student survey
		Growth mindset	N/A
	Artistic and Creative Traits	Participation in arts and literature	Teacher survey
		Creativity	N/A
	Health	Social and emotional health	Administrative data and student survey
		Physical health	Administrative data and teacher survey

APPENDIX B

Statement of Transparency

STUDY BACKGROUND

Information on School Performance

The 2001 No Child Left Behind (NCLB) Act requires states and local education agencies to publicly disseminate school performance data and information, making school report cards ubiquitous. The dissemination of data is part of a larger strategy to improve performance by holding schools accountable (Moynihan, 2008; Spillane, 2012). Today, parents and interested citizens can access vast quantities of data and information about school quality.

Performance data is thought to "help citizens judge the value that government creates for them" (Osborne & Plastrik, 2000, p., 247). According to the theory of action, once armed with data and information, interested parties will be empowered to select the best school and/or demand change from their elected representatives or their local school administrators (Moynihan, 2008). Believing in the value of performance information, policy makers have rapidly expanded the availability of education data available to parents (Feuer, 2008; McDonnell, 2008).

This Study

Most existing state data systems focus narrowly on student academic performance in literacy and mathematics. This study examines how citizens (both parents and nonparents) respond to different types of school performance data. Additionally, because data and information use is not an activity typically conducted in isolation, we examine how opinions change when participants engage in deliberative discussions about the data and information.

Toward this end, we developed a new system to present a wide array of data on a particular school district and tested it against the state's website. Specifically, our participants were randomly assigned to interact with the new system or the existing system. At specified times throughout the session, they also interacted with each other. The goal of the study was to ascertain how opinions developed differently between these two groups as a result of the types of data that they had access to.

This statement of transparency was written after our data were collected but before any data were viewed. This timing allows us to report on and be transparent about any irregularities that emerged during the data collection and make sensible decisions about data exclusions but still preregister our hypotheses.

METHODOLOGY

To test the usefulness of the new data system, we designed an experiment in the form of a representative poll. Forty-five participants were randomly divided between two high school computer labs—one in which participants were given access to the state of Massachusetts website reporting educational outcomes, and the other in which participants were provided with a web portal designed by our research team. Both were given an online survey to complete as they viewed the data.

In selecting participants, we pursued a random stratified sampling approach to select 50 participants from a pool of 90. After dividing potential participants into subgroups—race/ethnicity, gender, age, income, child in public school—we first worked to match the racial demography of the city by randomly selecting participants from the relatively small number of non-White subgroups. After doing so, we included all available men, because our pool was skewed female by a roughly 2-to-1 ratio. From the remaining pool of volunteers, we sorted by income category and randomly selected participants until all four income categories had roughly equal numbers. We then checked the number of participants with children in the public schools and found an imbalance that we remedied by replacing four parents with demographically similar nonparents. This created demographic matching across the groups, though creating matched pairs across all five criteria was impossible given the pool of potential participants. A total of 43 of 50 confirmed participants arrived on the day of the poll, with 2 day-of-event arrivals bringing the total number to 45.

The procedures unfolded as follows: Participants explored data on their own, discussed the data in small groups within their experimental condition, and then engaged in small-group discussions that included members from both lab A and lab B. Because these nine final groups were created by randomly selecting identification numbers on the day of the event, one group was not heterogeneous with regard to which data were explored, and four groups had ratios of 4-1. When they first arrived, after each of these stages, and after the final discussion, participants completed surveys to assess their opinions about a pair of schools.

Participant Activities on Polling Day

Activity:	Approximate time
Survey 1	10:25–10:40
Data Viewing	10:40–11:00
Survey 2	11:00–11:10
Within Experimental Condition Small-Group Discussion	11:10–11:30
Survey 3	11:30–11:40
Heterogeneous Group Discussion	11:40–12:10
Survey 4 and Demographic Survey	12:10–12:20
Sign-out	12:20–12:30

At the end of the event, participants signed out, were given $100 stipends, and were asked to complete and mail back some feedback to the school district, functioning as a behavioral outcome for the experiment (that is, one of our dependent variables of interest was whether people would write additional feedback and mail the letter back to the district). The letter, accompanied by a self-addressed stamped envelope, asked participants to list five things the district is doing well and five recommendations they would make for improving the schools, as well as to list any additional thoughts about the district as a whole. Letters and envelopes were labeled with unique identifiers.

Two irregularities are worth noting throughout these procedures. First, in completing surveys, several participants started and then restarted their work, having errantly navigated through the survey or forgotten their places. In these cases, they were directed to create new entries that would then be hand-sorted. Additionally, one participant, at the end of the study, walked into his nonassigned computer lab and began to explore the new data. Although this behavior could not impact his survey responses, it could affect the behavioral outcome. Because we were able to intervene quickly and ask him to wait until a later date, we retained him in the sample for all analyses.

List of Variables Collected in the Present Study

Self-report measures:	Number of items
How familiar are you with _____?	1
Overall *school* rating (asked for 2 schools on 4 occasions)	**5**
If you were in charge of improving _____, how accurately could you identify the top three areas in need of improvement? (asked for two schools on 4 occasions)	2
Overall *district* rating (asked for the district on 2 occasions)	**5**
If you were in charge of improving the city's public schools, how accurately could you identify the top three areas in need of improvement? (asked on 2 occasions)	1
In what way, if at all, has your opinion of _____ changed? (asked for two schools on 3 occasions)	**1**
Information value (asked on 2 occasions)	4
Behavioral measure	
Do participants return the letter? (yes/no)	1
Number of items responded to in letter	11
Word count of letter	
Background information	
How long have you been a resident of this city?	1
How much do you feel you know about the city's public schools?	1
How comfortable are you interacting with data?	1
How much research have you done on the city's public schools?	1
Do you have a child enrolled in school?	1
In what year were you born?	1
How would you describe your race/ethnicity?	1
What language do you speak at home?	1
What is your gender?	1
What is the highest level of school you have completed?	1
What is your approximate annual household income?	1

*Note: **Bolded** variables will be described in the article's Measures section and will be used to analyze the focal a priori hypotheses for this study. The nonbolded variables may be used for exploratory analyses.

PRIMARY HYPOTHESES

We will test the following hypotheses, which illuminate the differences between the value of the new, multifaceted data presentation as compared with the types of data the public can typically access:

Hypothesis 1: Value of the learning experience

As compared with the control group, participants who interacted with the multifaceted data will report valuing the information they received more highly.

Hypothesis 2: Understanding of school/district quality

a. As compared with the control group, participants who interacted with the multifaceted data will report higher overall ratings of individual school quality at the second, third, and fourth time points.

b. As compared with the control group, participants who interacted with the multifaceted data will report higher overall ratings of the school district at the final time point.

Hypothesis 3: Attitude change

As compared with the control group, participants who interacted with the multifaceted data will manifest greater changes in their opinions as a consequence of the first two discussions.

Hypothesis 4: Investment in school system

As compared with the control group, participants who interacted with the multifaceted data will write more in those letters, indicating broader definitions of school quality.

ANALYTIC DETAILS

Exclusion Criteria

We will not exclude any participants. For the respondents who skipped ahead in the survey administration, we had them return to the survey and complete the same set of items after they had participated in the discussions. We will exclude the data from their original responses on the final segment and instead use their responses from after they had participated in the discussions.

Analysis

For the first hypothesis, we will regress our treatment variable onto the information-value composite. We will run two such regressions: one for the first time when participants are asked about the school information that they just used, and one for the time when participants have just finished their initial discussion about the schools. The background question regarding knowledge of the schools will be included as a covariate. In the second regression, we will use fixed effects to account for which discussion group they were in.

For the second hypothesis, we will (a) examine the effect of the treatment on the school rating composite. Because these ratings are provided across four time points, we will use repeated measures analysis of covariance (ANCOVA) to test for differences between the treatment and control groups. We will run two such ANCOVAs: one for the school participants are most familiar with, and a second for the school they are randomly assigned to report on. Their self-reported familiarity with the school will be included as a covariate for each ANCOVA. We will also (b) regress the treatment variable onto the district rating composite at the final time point. Time point #1 will provide a baseline estimate of participants' opinions, but we do not expect significant differences here.

For the third hypothesis, we will examine the effect of the treatment on how much participants felt that their opinion changed. Because these ratings are provided across three time points, we will use repeated measures ANCOVA to test for differences between the treatment and control groups. We will run two such ANCOVAs: one for the school participants are most familiar with, and a second for the school they are randomly assigned to report on. Their self-reported familiarity with the school will be included as a covariate for each ANCOVA.

For the fourth hypothesis, we will regress the treatment on the number of words written by participants. The background question regarding knowledge of the schools will be included as a covariate.

For the sake of clarity in communicating our findings, we will include graphs and the associated 95% confidence intervals for each of the hypotheses.

We will register this statement on June 16, 2015, and will not look at our data prior to the completion of that process.

Signed on behalf of all co-authors,

Author

REFERENCES

Feuer, M. J. (2008). Future directions for educational accountability: Notes for a political economy of measurement. *The Future of Test-Based Educational Accountability*, 293–306.

McDonnell, L. M. (2008). The politics of educational accountability: Can the clock be turned back? *The Future of Test-Based Educational Accountability*, 47–68.

Moynihan, D. P. (2008). *The dynamics of performance management: Constructing information and reform*. Washington, DC: Georgetown University Press.

Osborne, D. E., & Plastrik, P. (2000). *The reinventor's fieldbook: Tools for transforming your government* (p. 42). San Francisco, CA: Jossey-Bass.

Spillane, J. P. (2012). Data in practice: Conceptualizing the data-based decision-making phenomena. *American Journal of Education*, *118*(2), 113–141.

JACK SCHNEIDER is Assistant Professor of Education at the College of the Holy Cross and Director of Research for the Massachusetts Consortium for Innovative Education Assessment. His latest book is *Beyond Test Scores: A Better Way to Measure School Quality* (Harvard University Press in 2017).

REBECCA JACOBSEN is an associate professor in the College of Education at Michigan State University and associate director of the Education Policy Center. Her research examines how policies shape both opportunities for and barriers to engagement with the public education system. She has written extensively about the politics of accountability policies and how performance reporting shapes public trust in and support for public education. Her research has been published in *Public Opinion Quarterly, American Education Research Journal, American Journal of Education*, and *Journal of Public Administration Research and Theory.*

RACHEL S. WHITE is a doctoral candidate at Michigan State University. Her research focuses on issues of power, politics, and democratic accountability in educational policy making, as well as on the degree to which stakeholder voices are incorporated in the crafting and implementation of policy. Her research has been published in *Educational Evaluation and Policy Analysis* and *Educational Policy Analysis Archives.*

HUNTER GEHLBACH is an associate professor at UC-Santa Barbara's Gevirtz Graduate School of Education and the director of research at Panorama Education. An educational psychologist by training and a social psychologist at heart, his interests lie in improving the social side of schools, questionnaire design, and (recently) environmental education. His recent field experiment, "Creating Birds of Similar Feathers: Leveraging Similarity to Improve Teacher–Student Relationships and Academic Achievement," was published in the *Journal of Educational Psychology* and covered by NPR and *The Atlantic.*