

Identifying Good Teachers: Expert Versus Ordinary Knowledge

David Casalaspi, *Michigan State University*

Ethan Hutt, *University of Marland-College Park*

Jack Schneider, *College of Holy Cross*

Abstract While much has been written about the effects of standardized testing on student achievement, less work has addressed how parents take up this information. Drawing on a survey of 286 parents in a diverse urban school district, this research illuminates three aspects of parental response to test score information: 1) how parents relate various teacher traits to quality teaching; 2) how parents know if their child has a good teacher; and 3) how parents think teachers should be evaluated. Results show that test score data are perceived as both imperfect and incomplete with regard to measuring teacher quality, and that parents often rely more on “ordinary” forms of knowledge. This raises questions about the value of test data as an informational spur to reform.

Keywords Accountability; Parents; Teachers; Testing; Value-added

David Casalaspi, Ethan Hutt, & Jack Schneider (2018). Identifying Good Teachers: Expert Versus Ordinary Knowledge. *International Journal of Education Policy & Leadership* 13(4). URL: <http://journals.sfu.ca/ijepl/index.php/ijepl/article/view/744> doi: 10.22230/ijepl.2018v13n4a744

IJEPL is a joint publication of PDK International, the Faculty of Education at Simon Fraser University, the College of Education and Human Development at George Mason University, and the University of Delaware. By virtue of their appearance in this open access journal, articles are free to use, with proper attribution, in educational and other non-commercial settings 90 days after initial publication. Copyright for articles published in IJEPL is retained by the authors. More information is available on the IJEPL website: <http://www.ijepl.org>



Parents don't get solid, helpful information on their children's schools and teachers. Instead they rely on legend, intuition and chance.
(Duncan, 2010, p. 8)

IJEPL 13(4) 2018

Casalaspi, Hutt,
& Schneider

Identifying Good
Teachers: Expert
Versus Ordinary
Knowledge

Introduction

For at least two decades, U.S. federal education policy has made it a goal to provide parents and members of the public with objective information about the performance of schools, teachers, and students. The reform theory behind the drive to collect and disseminate this data, which often take the form of student standardized test scores, is rooted in the twin assumptions that the public is mostly in the dark about the performance of its schools and that public ignorance is a serious policy problem. As the theory of action goes, without objective information about school and teacher performance, parents will be unable to evaluate performance effectively and demand change when change is needed—something that threatens not only the goal of universal educational excellence (Schneider, 2011) but also the cherished ideal of local control (Fuhrman & Elmore, 1990). As both President Bush and Secretary of Education Rod Paige put the issue while stumping for the No Child Left Behind Act in 2001, test score data would equip parents to be “the great enforcer” of educational excellence, furnishing parents with the information they need to confront unresponsive and ineffective educational bureaucracies (United States Senate, 2001, p. 27). “There’s nothing like getting a mother fired up when she sees the fact that her school may not be performing quite like she thought it was going to be,” President Bush (2001, p. 6) once said. “Parents, armed with data are the best forces of accountability in education.”

The belief that test score data are critical in improving schools has today become so pervasive that recent efforts to reduce its flow have been met at times with outrage and defiance. Former Secretary of Education Arne Duncan (2013, p. 1) once rejected a request from the State of California to suspend testing by framing the issue in moral terms: “A request ... to not measure the achievement of millions of students this year is not something we could approve in good conscience.” Civil rights groups too have decried efforts to decrease the amount of testing by arguing that “we cannot fix what we cannot measure” (Leadership Conference on Civil and Human Rights, 2015). And in response to calls to reduce testing loads on districts and students, Russ Whitehurst, former director of the Institute for Education Sciences (IES), countered that what our country needs is *more* test score information, not less (Lindquist & Whitehurst, 2014).

Acting on this theory about the actionable value of test score information, editors at major newspapers have also taken it upon themselves to publish educational data gathered through Freedom of Information Act (FOIA) requests. The *Los Angeles Times* (2010), for instance, hired an economist to calculate teacher value-added scores based on student achievement data and then published the results—articulating a desire to inform parents and the public in the process. “Until now,” the editors wrote, “parents have had no objective information about the effectiveness of their child’s teacher.” Even a Court of Appeals (2014, p. 11) judge who objected to the release

of the teacher value-added information accepted as “common sense” the fact that this information is invaluable to parents, adding, “Parents will naturally have a strong desire to get their children into classes with the highest scoring teachers.”

While many policymakers on both the left and the right take it as self-evident that test score data represents an informational advance over what was previously available, it is not clear whether or not parents share this view. What will standardized test data teach parents that they do not already know? How useful do parents find this data in comparison to information from other, less technically sophisticated sources? The answers to these questions are important not only from a policy effectiveness standpoint but also because the information provided by test scores can have real impacts on the students, teachers, schools, and communities being assessed (e.g., Jacobsen, Saultz, & Snyder, 2013). Moreover, if the intended audience for this information does not find it any more useful than sources they already have in their possession, it may be necessary to revise current accountability regimes so that they can produce more comprehensive and authoritative information.

This study advances this kind of assessment by providing results from a survey of 286 parents that captured parental perceptions of test score data. The findings are situated within theories of knowledge use and conceptions of expert and ordinary knowledge specifically.

Historical and theoretical background

In recent years, a tremendous amount of research has studied standardized testing, focusing most frequently on issues of effectiveness, comprehensiveness, and validity. A principal concern here has been determining whether high-stakes testing policies by themselves lead to improvements in student achievement (e.g., Carnoy & Loeb, 2002; Gandara & Randall, 2015; Grissmer, Beekman, Ober, 2014; Hanushek & Raymond, 2005; Jacob, 2005; Judson, 2012; Nichols, Glass, & Berliner, 2006, 2012; Supovitz, 2009). Additional work has examined the effects of high-stakes testing on other outcomes, such as deeper student learning (e.g., Darling-Hammond & Snyder, 2015; Darling-Hammond, Wilhoit & Pittenger, 2014; Koretz, 2008; Marion & Leather, 2015), equity (e.g., Booher-Jennings, 2005; Kim & Sunderman, 2005; Nese, Tindal, Stevens, & Elliott, 2015), curricular breadth (e.g., Misco, Patterson, & Doppen, 2011), teacher anxiety (e.g., Amrein-Beardsley, 2009; Berryhill, Linney, & Fromewick, 2009), and teacher retention (e.g., Winters & Cowen, 2013), among others.

The purpose of this article is not to revisit these studies or weigh in on these debates. Rather, it aims to highlight a different dimension of standardized tests: their value in informing parents about teacher effectiveness. Indeed, less work has studied parent responses to standardized test score data, and the most salient work to date has been a series of school report card studies conducted by Rebecca Jacobsen and her colleagues. In one of those studies, Jacobsen, Andrew Saultz, and Jeffrey Snyder (2013) examined school report card changes in New York City and found that when school accountability ratings declined after the implementation of higher standards, parent satisfaction with schools declined, threatening an important base of political support for the local education system. In a later study, Jacobsen, Snyder, and Saultz (2014) further found that the *format* of school report cards influenced parent per-

ceptions of school performance, with parents interpreting report cards for identical schools differently based on the formatting (see also Hastings & Weinstein, 2008).

Given how limited work in this area has been overall, it remains an important question whether or not parents find the information from test score reports useful and actionable, particularly with regard to identifying effective teachers. It is also not clearly understood how parents value test score data in relation to information derived from other less formal sources—such as everyday interactions or personal experiences—and there has been scant effort to link such analyses to broader theoretical frameworks about knowledge use.

A brief history of teacher evaluation practices

To fully consider the potential knowledge value of standardized testing and teacher evaluation scores to parents, it is important to understand the long history of teacher evaluation, and recognize that new standardized test score data are newcomers in an existing information ecosystem.

Indeed, teacher evaluation has been a central concern in American public education since the creation of common schools, although because the technology available for this task was ill-equipped to measure teacher quality in a standardized way, parents, administrators, and community members often relied on the more informal tools at their disposal—namely, direct experience and in-person interactions (Sedlak, 1989). In the time of nineteenth-century education reformer Horace Mann, for instance, it was not uncommon for parents to sit in classrooms and observe teachers directly, and in one high-profile incident, a group of parents even commandeered local classrooms, administered a pop quiz to the students, and then condemned the educators after the students flunked (Reports of the Annual Visiting Committees, 1845). In the early and mid-twentieth century, explosive face-to-face interactions became less frequent as the school system grew and bureaucratized, but the creation of parent-teacher associations (PTAs), parent-teacher conferences, and modern school board meetings helped preserve some degree of direct interaction with school officials.

Nevertheless, school managers and government agents imbued with the teachings of Taylorism ultimately viewed the informal information of parents to be vague, subjective, and idiosyncratic, and they sought to generate more technical, objective information about school and teacher quality (Reese, 2013). Consequently, school overseers produced a vast amount of statistical information about American schooling in the early and mid-twentieth century, but due to technological constraints, much of this information was too descriptive to offer much use from an evaluation or reform standpoint (Warren, 1974).

Despite a flood of statistical information, then, there remained a healthy buffer between these statistics and the realities of schools as experienced by communities. This gap was reinforced in the design of later educational policies from the 1980s onward. The ambitious effort to create a national report card—the National Assessment of Educational Progress—was intentionally executed in such a way as to forestall the possibility of disaggregating the national statistics (Hazlett, 1974). Even the No Child Left Behind Act (NCLB) was initially designed to thwart the narrow parsing of educational statistics by prohibiting the tracking of individual students and teachers and

mandating that data be aggregated by grade level—a move backed in several states by laws that expressly forbade the linking of individual student- and teacher-level data (Duncan, 2009). Thus, for the better part of a century, parents and community leaders did not rely on government-produced statistics to inform them about teacher quality.

The dramatic developments in teacher evaluation that have taken place over the last decade—in which the tools for analyzing student achievement data have become increasingly sophisticated and powerful in guiding policy decision-making—should be viewed against this historical background. This effort accelerated with the passage of the NCLB, and it has reached its zenith with the creation of value-added measures (VAMs), which attempt to capture the achievement gains that an individual teacher is responsible for over the course of a year. Unlike the status measures of achievement (e.g., proficiency levels) that were mandated under the NCLB, value-added measures are designed to measure student growth over the course of the year, essentially controlling for achievement influences outside of the school or individual classroom. Many scholars have long believed that VAMs hold promise for evaluation purposes (e.g., Harris 2009, 2011; Ladd, 2001), and though the statistical reliability and validity of these estimates are still contested (Ballou & Springer, 2015; Brown, 2008; Haertel, 2013; Harris, 2009), some scholars believe they are nevertheless sound enough for making high-stakes decisions about teacher effectiveness (Aaronson, Barrow, & Sander, 2007; Chetty, Friedman, & Rockoff, 2011; Jacob & Lefgren, 2005).

Amid the excitement surrounding VAMs, a number of districts have jumped headlong into the use of VAMs for gauging teacher effectiveness, including New York City, Chicago, and Washington, D.C. (Headden, 2011). Moreover, spurred by Race to the Top (RTTT) and the Obama administration's NCLB waiver policies, thirty of the fifty states have begun linking teacher evaluations with student test score data (Polikoff, McEachin, Wrabel, & Duque, 2014). These technical notions of teacher quality also found judicial expression in the 2014 court case *Vergara v. California*, in which expert witnesses used value-added data to claim that California policies allowed ineffective teachers to remain in the classroom and that these teachers disproportionately served poor and minority students, depriving children of their constitutional right to a quality education.

These developments clearly suggest that, at least in elite circles, quantified “expert” knowledge has made real inroads in displacing the “ordinary” knowledge historically possessed by those close to schools. Yet despite the promise of VAMs as more precise and less corruptible measures of teacher effects on student test scores, fundamental questions still remain about the reception of this information by local audiences. It is not surprising to find a preference for quantitative data among those whose roles put them in control of, but at a considerable distance from, individual schools. But, as noted above, it is not yet clear if test score data provide the kind of particularistic information that is of interest to local parents.

Expert versus ordinary knowledge

Considering the differences between what scholars have termed expert knowledge and ordinary knowledge can enhance an understanding of how contemporary test score data can inform parents. To date, there is a sizeable literature on the divergent

reception of certain official forms of information, and this study provides an interesting opportunity to explore two theoretical dimensions: the potential disconnect between expert knowledge and ordinary knowledge; and the value of quantitative, seemingly objective forms of information when compared with more subjective forms of information.

Generally speaking, three books dominate the literature on knowledge use. The first is Charles Lindblom and David Cohen's (1979) *Usable Knowledge*, which lays out the differences between scientific expertise (called Professional Social Inquiry, or PSI) and "ordinary knowledge." Professional Social Inquiry encompasses academic social science as well as other technical methods of investigation. Ordinary knowledge, on the other hand, is defined as "knowledge that does not owe its origin, testing, degree of verification, truth status or currency to distinctive PSI professional techniques but rather to common sense, casual empiricism or thoughtful speculation and reflection" (p. 12). While policy experts often disparage ordinary knowledge as mere folklore or superstition, it is often quite valuable as a source of truth. Ordinary knowledge tells us that when demand increases, prices increase; that cities have higher homicide rates than the suburbs; that children watch a lot of television; and that when it rains, the road behind your house is prone to flooding. These are true pieces of knowledge that can guide everyday social action, and Lindblom and Cohen (1979) argue that practitioners of PSI should more openly recognize their limited ability to reshape and displace ordinary knowledge. Often, when PSI is most authoritative in changing social behavior, it is because the findings of PSI confirm existing ordinary knowledge (as when VAMs confirm the popular belief that Mr. X is a bad math teacher). When divorced from ordinary knowledge, though, PSI is often ignored, especially if its results are inconclusive or contested.

The second book is Frank Fischer's (2000) *Citizens, Experts, and the Environment*, which examines expert and citizen interactions vis-à-vis environmental policy. Fischer begins by acknowledging that modern society is socially and technically complex—something that leads observers to wonder if ordinary citizens can intelligently deliberate on complex policy issues. Using debates around environmental policy as a case study, Fischer argues that ordinary citizens *can* not only participate but that their participation is vital in the search for policy solutions because it brings forth new knowledge and ideas that can address gaps in expert knowledge. He therefore rejects the prevailing notion that ordinary citizens are confused or uninformed and suggests that their knowledge is an important complement to the work of experts.

Throughout his book, Fischer (2000) develops the concept of "local knowledge." Like Lindblom and Cohen's (1979) conception of ordinary knowledge, local knowledge is understood as "knowledge about a local context or setting" that owes its power not to distinctive professional methods of inquiry, but to "casual empiricism, thoughtful reflection, and common sense" (p. 146). According to Fischer (2000), local knowledge differs from expert knowledge in that it not only accounts for local context but it also possesses "cultural rationality." In contrast to technical rationality, which relies on expert judgments and emphasizes positivist notions of logical consistency and generalizability, cultural rationality is "geared to—or at least gives equal weight to—personal and familiar experiences rather than depersonalized technical

calculations” (p. 132). It centers on the opinions of social peers and is the “rationality of the social lifeworld” (p. 133).

The third book, James C. Scott’s (1998) *Seeing Like a State*, adds an additional dimension to the literature on expert versus ordinary knowledge by more explicitly theorizing about the state’s impetus to produce certain kinds of expert knowledge. Specifically, Scott notes that the goal of the modern state is to make complex social domains more comprehensible, or “legible,” so that they can be more readily governed and organized for efficiency. Scott points out that government-led efforts to improve society are often imbued with a “high-modernist ideology” characterized by uncritical self-confidence in scientific progress and the rational design of the social order (p. 5). But despite the hopes of scientific and governmental elites, efforts at social engineering often end in failure and produce a myriad of unintended consequences. To explain why, Scott blames the fact that these efforts often exclude the ordinary knowledge embedded in local practice, and instead rely too heavily on scientific knowledge and decontextualized quantitative data. This is a mistake because ordinary knowledge is practically superior to scientific knowledge due to the fact that it is both nestled in the local context and held by those who have a passionate interest in the desired social outcome.

In general, Scott is most interested in instances in which the state’s quest for legibility blinds it to the subtle complexity of local political ecologies, but what is perhaps most striking about testing policies in the American education context is the augmented justification of the use of official knowledge to empower civil society. While testing policies such as NCLB, RTTT, and the Every Student Succeeds Act (ESSA) have embraced a variety of sanctions-based policy tools to improve school and teacher quality, a central mechanism in the theory of action leading from testing to school improvement is the promotion of public engagement with education through the public reporting of test score data intended to ostensibly replace incomplete local knowledge. In this way, American test-based accountability policies rely not only on a top-down approach (i.e., governmental sanctions) but also on a bottom-up approach (i.e., parents using the information to demand change from their schools or otherwise exit the system). As such, the government’s need to improve education is intimately intertwined with the provision and uptake of parental school-quality information, and parents are both the beneficiaries and the agents of the government’s test-based education reform program.

While expert knowledge does not always take quantitative form, it frequently does in the cases described above and in contemporary society—a fact that adds an additional layer of complexity to the tension between expert and ordinary knowledge. Historian Theodore Porter (1996, p. ix) argues that quantification is best understood as a “technology of distance”—that is, numbers have the ability to travel long distances while remaining intact and also function as form of public communication given their rule-bound character. The universality of numbers, Porter argues, suggests that they have transcended personal judgment—something that allows us to put our “trust” in them even when they come to us across great distances and from sources directly unknown to us. Indeed, numbers can have broad political and social appeal even though quantification is just one of many rhetorical strategies subject

to contestation and counter-narratives (e.g., Carruthers & Espeland, 1991; Espeland & Stevens, 2008; Power, 2004; Stone, 2011).

The crucial point in all of this literature is that the advent of new kinds of data does not instantaneously rewrite the social world. Rather, these new forms of knowledge exist and compete for acceptance alongside the old forms, allowing for the uneven penetration of expert, quantitative indicators in different circles. The undeniable penetration of test score data in elite education policy circles, then, does not foretell a parallel acceptance among parents who may already have access to different and competing forms of information at the local level. The extent of this parental acceptance remains an open empirical question—one that is examined here.

Method

Survey

This article explores how parents define good teaching, how they acquire information about teachers, and how they feel teachers should be evaluated. To understand how parents think about these things, a descriptive survey of parents in a highly diverse, urban school district in the American Northeast was conducted. According to Jelke Bethlehem (2009) descriptive surveys are best used to describe a phenomenon, explore interesting dilemmas or tensions, and act as a spur for further research.

The school district under study had been producing data reports required by the state and federal governments for years, but it did not believe that such reports accurately captured school quality or were of much informational value to parents. This is both because such reports are constructed from a relatively narrow set of measures, and because many of the measures correlate strongly with family socioeconomic status. Insofar as that was the case, the district was curious what kinds of information would be valuable to parents and the public, with the intention of eventually producing its own (more robust) report cards that would more systematically and comprehensively guide decision-making about key policy issues such as teacher evaluation and financial resource allocations.

This survey asked a total of 33 questions, most of which aimed to capture respondents' opinions of and relationship to their child's public school. The heart of the survey, however, was a set of three questions that dealt with teacher quality concerns and the value of standardized test score data as an informational source. These questions, and the response options for them, were developed over multiple weeks, drawing on outside literature and piloting through parent focus groups to ensure clarity and comprehensiveness.

1. "Please rank the importance of the following teacher characteristics."
(1 = most important, 6 = least important)
 - a. Ability to Relate to Students
 - b. High Level of Enthusiasm
 - c. Ability to Maintain Order
 - d. Cultural Sensitivity
 - e. Ability to Produce High Test Scores
 - f. Racial/Ethnic Background

2. "Which of the following tells parents if their children have good teachers?" (yes/no)
 - a. Children Tell Them
 - b. Parent Interactions with Teachers
 - c. Parent Interactions with Other Parents
 - d. Test Scores
 - e. Impossible to Know
3. "Which of the following should be used to evaluate teachers?" (yes/no)
 - a. Principal Observations
 - b. Parent Feedback
 - c. Student Feedback
 - d. Observations by a District Official
 - e. Test Scores

The sample

Participants for the survey were recruited widely through several mechanisms. Recruitment letters were sent home via student backpacks in grades K–8 in English and Spanish to all families in the district. The survey itself was electronically administered via SurveyMonkey in both languages, and a link was provided in the recruitment letters. The link was also posted on online discussion boards and to a schools-oriented listserv. In order to reach K–8 families without internet access, printed copies of the survey were distributed to all schools through a family-outreach program. The authors also partnered with English as a second language (ESL) classes being taught to parents. Parents were given six weeks to complete the survey, and the average parent completed the survey in nine minutes.

The district is a relatively small one, with roughly 3,300 students enrolled in grades K–8. According to a district representative, there were 2,579 separate families in the district with students enrolled in grades K–8 at the time of the survey. From this pool of 2,579 potential participants, 394 responses were received, of which 286 provided complete information—a complete response rate of 11.1 percent. While lower than hoped, this response rate is comparable to other published education studies employing web-based surveys (e.g., Bridich, 2016; Dodson, 2015; Laitsch, 2002; Milanowski, Longwell-Grice, Saffold, Jones, Schomisch, & Odden, 2009; Misco, Patterson, & Doppen, 2011), and it is acceptable for the type of descriptive, non-causal, exploratory analyses conducted here provided that either 1) the sample is relatively reflective of the population as a whole, or 2) significant differences in response patterns do not emerge between groups and bias the results (De Vaus, 2014). Demographic information about the survey respondents and the district as a whole can be found in Table 1.

As Table 1 shows, respondents represent a fairly diverse group, albeit one with relatively high levels of education and income. With regard to race, 63.4 percent of respondents identified as White, 13.6 percent identified as Hispanic, and 14.3 percent identified as another race. The sample also contains a large number of individuals from all major income brackets, although it does overrepresent those in the two high-

est income brackets. The sample is also well educated, with 15.5 percent of respondents lacking a college degree and 63.8 percent holding at least a Bachelor's degree.

Table 1. Demographic information of survey respondents and the district as a whole

Identifying Good
Teachers: Expert
Versus Ordinary
Knowledge

	Sample	District as a Whole
Gender of parent	Number (Percent)	Percent
Male	78 (27.3%)	49.1%
Female	196 (68.5%)	50.9%
Parental education level		
Less than high school	8 (2.3%)	12.2%
High school degree	17 (5.9%)	20.9%
Some college	21 (7.3%)	11.0%
Associate's degree	20 (7.0%)	3.6%
Bachelor's degree	70 (24.5%)	26.4%
Graduate degree	141 (49.3%)	26.0%
Family income		
\$0 – \$49,999	40 (14.0%)	37.0%
\$50,00 – \$99,999	69 (24.1%)	31.0%
\$100,000 – \$149,999	73 (25.5%)	18.7%
\$150,000+	75 (26.2%)	13.4%
Race		
White	183 (63.4%)	69.4%
Hispanic	39 (13.6%)	11.1%
Other (total)	41 (14.3%)	19.5%
Black	2 (0.7%)	6.3%
Asian	19 (6.6%)	9.6%
Native American	4 (1.4%)	0.1%
Other	16 (5.6%)	3.5%
Child's grade level		
Grade K–2	123 (43.0%)	38.1%
Grade 3–8	162 (56.6%)	61.9%

Notes: Total $n = 286$. Percentages do not add up to 100 due to item nonresponse.

Overall, the sample is mostly reflective of the district as a whole, with the exception of the overrepresentation of female respondents, respondents with high levels of educational attainment, and respondents making more than \$100,000 per year. With regard to gender, no statistically significant differences in response patterns between male and female respondents were found on any survey items, so the risk of sampling bias here is minimal. Moreover, given that political proponents of accountability have long touted the ability of test score data to empower *mothers*, this may even be advantageous for the research (McGuinn, 2006). With regard to income, those making \$100,000+ annually comprise 51.7 percent of the sample even though this group only makes up 32.1 percent of the population of the district as a whole. However, because all income groups are well represented in the study and statistically significant differences emerged in the response patterns of high- and low-income

individuals only rarely, bias from this overrepresentation is not overly concerning. Additionally, as discussed in the limitations section of this article, the overrepresentation of high-income individuals produces some risk of understating the conclusions, but there is much less risk that it would lead to the more fatal mistake of overstating the conclusions. Thus it is reasonable to proceed with caution.

Data analysis

After the survey results were compiled, STATA statistical software was used to generate basic descriptive analyses of the data—namely tables of frequencies, means, and medians. A second level of analysis was then conducted by disaggregating the data by race, income, education, and student grade-level in order to add nuance to the main findings. One-way ANOVAs and two-tailed *t*-tests were run to identify any statistically significant differences in response patterns across subgroups. The majority of between-group differences were not statistically significant and the subgroup sample sizes are limited in some cases (such as Hispanics, $n = 39$), so the subgroup analyses should be read merely as adding descriptive nuance to the main findings.

Findings

The survey results provide information on three areas related to teacher quality and evaluation: 1) how parents relate various teacher traits to quality teaching; 2) how parents know if their child has a good teacher; and 3) how parents think teachers should be evaluated. Each of these will be examined in turn.

Teacher traits

First, parents were asked to rank the importance of six different teacher characteristics on a scale of 1 to 6 (with 1 being the highest): ability to produce high test scores; sensitivity to student racial and cultural backgrounds; ability to maintain order and discipline; a high level of energy and enthusiasm; ability to relate to students; and the racial/ethnic/language background of the teacher.

As Table 2 shows, parents overwhelmingly chose “the ability to relate to students” and “a high level of enthusiasm” as the two most important characteristics of good teachers. Both of those criteria received median rankings of 2 on the 1–6 scale, with 70.0 percent of respondents ranking relatability in their top-two concerns and 63.5 percent of respondents ranking enthusiasm in their top-two concerns. The ability to produce high test scores, on the other hand, was deemed much less important, with a median ranking of 5 and only 18.6 percent of respondents ranking that characteristic in their top-two concerns. The ability to produce high test scores was also trumped in importance by the ability to maintain order (median 3, 30.7% in top two) and sensitivity to student backgrounds (median 4, 16.7% in top two).

Similar results were found when examining which characteristics respondents ranked as *least important* (see Table 2). Less than 10 percent of parents assigned relatability or enthusiasm a 5 or 6 ranking, whereas a majority (56.8%) did so for the ability to produce high test scores. Of all the characteristics listed, only “race, eth-

nicity, or language background” was deemed less important than test scores on the survey, with 84.9 percent of respondents assigning it a 5 or 6 ranking.

Table 2. Parent rankings of teacher characteristics

Teacher characteristics (all respondents)												
Item	Top 2 (%)			Bottom 2 (%)			Median					
Ability to relate to students (Relatability)	70.0			8.8			2					
High level of enthusiasm (Enthusiasm)	63.5			8.8			2					
Ability to maintain order (Order)	30.7			10.3			3					
Cultural sensitivity (Sensitivity)	16.7			27.3			4					
Racial/Language background (Racial Background)	6.0			84.9			6					
Ability to produce high test scores (Test Scores)	18.6			56.8			5					
Teacher characteristics (by race)												
Item	Top 2 (%)			Bottom 2 (%)			Median					
	White	Hispanic	Other	White	Hispanic	Other	White	Hispanic	Other			
Relatability	70.4	60.0	69.2	5.6	20.0	15.4	2	2	2			
Enthusiasm	62.6	53.9	71.8	7.8	23.1	5.1	2	2	2			
Order	33.0	25.0	28.2	10.1	8.3	10.3	3	4	3			
Sensitivity	15.1	25.9	20.5	29.0	25.9	23.1	4	3	3			
Racial background	2.8	30.8	0.0	88.8	57.7	89.7	6	5	6			
Test scores	16.2	46.2	10.3	58.7	42.3	56.4	5	3	5			
Teacher characteristics (by income bracket)												
Item	Top 2 (%)				Bottom 2 (%)				Median 2			
	1	2	3	4	1	2	3	4	1	2	3	4
Relatability	63.6	71.9	68.5	75.0	12.1	4.7	6.8	5.6	2	2	2	1
Enthusiasm	51.6	60.9	67.1	66.7	9.7	9.4	13.7	3.8	2	2	2	2
Order	32.3	34.4	28.8	30.6	6.4	10.9	9.6	9.7	3	3	3	3
Sensitivity	21.2	14.1	15.1	13.9	30.3	39.1	17.8	27.8	4	4	3	4
Racial background	12.9	3.1	4.1	1.4	48.4	84.4	87.7	95.8	5	6	6	6
Test scores	42.4	15.6	16.4	12.5	45.4	51.6	64.4	58.3	4	5	5	5

Notes: Income Bracket 1: \$0–\$49,999; Income Bracket 2: \$50,000–\$99,999; Income Bracket 3: \$100,000–\$149,999; Income Bracket 4: \$150,000+

These response patterns generally held up when the data were disaggregated by demographic characteristics. Although support for test scores diminished as income increased, no differences were statistically significant. There were also no discernable differences across education, income, gender, or student grade-level. However, there were two subgroup differences worth mentioning when it came to Hispanic parents. First, Hispanic parents were more evenly split about the importance of test scores: 46.2 percent of Hispanic parents placed test scores in their top-two concerns, and a

Identifying Good
Teachers: Expert
Versus Ordinary
Knowledge

similar percentage (42.3%) placed test scores in their bottom-two concerns. Second, Hispanic parents ranked racial background of the teacher much more highly than non-Hispanic parents. Although only 2.8 percent of white parents ranked racial background as a top-two concern, 30.8 percent of Hispanic parents did so.

How parents know

Next, parents were asked which sources of information help them ascertain teacher quality (question 2 above). As Table 3 shows, three sources of information earned overwhelmingly positive responses: interactions with one's own children (93.1%); interactions with teachers (94.4%); and interactions with other parents or community members (85.3%). Meanwhile, fewer than half of respondents (49.4%) believed that test scores were useful in identifying good teachers.

Table 3. Parent responses about information

Teacher quality information sources (all respondents)								
Item	Yes (%)			No (%)				
Children tell them	93.1			6.9				
Interactions with teachers	94.4			5.6				
Interactions with parents	85.3			14.7				
Test scores	49.4			50.6				
Impossible to know	9.8			90.2				
Teacher quality information sources (by race)								
Item	Yes (%)			No (%)				
	White	Hispanic	Other	White	Hispanic	Other		
Children tell them	93.2	88.9	97.6	6.8	11.1	2.4		
Interactions with teachers	96.1	90.0	95.0	3.9	10.0	5.0		
Interactions with parents	86.7	75.0	85.0	13.3	25.0	15.0		
Test scores	46.3	51.7	54.3	53.7	48.3	45.7		
Impossible to know	6.7	25.0 ^b	8.8	93.3	75.0 ^b	91.2		
Teacher quality information sources (by income)								
Item	1	2	3	4	1	2	3	4
Children tell	89.7	89.2	95.7	96.0	10.3	10.8	4.3	4.0
Interactions with teachers	94.1	92.5	93.2	97.3	5.9	7.5	6.8	2.7
Interactions with parents	87.5	80.6	88.6	90.4	12.5	19.4	11.4	9.6
Test scores	57.8	53.4	39.1	46.5	42.2	46.6	60.9	53.5
Impossible to know	25.0	8.9 ^a	6.9 ^b	3.1 ^c	75.0	91.1 ^a	93.1 ^b	96.9 ^c

Notes: Income bracket 1: \$0–\$49,999; Income bracket 2: \$50,000–\$99,999; Income bracket 3: \$100,000–\$149,999; Income bracket 4: \$150,000+. ^a = $p < 0.10$, ^b = $p < 0.05$, ^c = $p < 0.01$

These results held up across income, education, gender, and student grade-level differences. However, a few differences did once again emerge in the response patterns between White parents and Hispanic parents. First, White parents and

Hispanic parents demonstrated differences with regard to the perceived usefulness of parental interactions with both teachers and parents. White parents showed more faith in interactions with teachers than Hispanic parents (96.1% and 90.0%, respectively). Similarly, White parents also showed more faith in parental interactions with other parents than Hispanics (86.7 and 75.0%, respectively). Because both of these variables are based on communication and social interaction, it seems likely that these differences reflect the language barriers facing Hispanics in the district. One additional factor—and one that demonstrated statistically significant differences between White parents and Hispanic parents—was the frequency with which they believed it was impossible for them, given the tools available, to know. While only 6.7 percent of White parents felt that it was impossible for them to tell if a child had a good teacher, nearly four times as many Hispanic parents (25.0%) felt this way.

As part of the survey, we also asked parents how often they engaged in various school-related activities, such as visiting their child's school, talking to teachers, or talking with other parents. Contrary to expectations, parents who visited their child's school, talked with teachers, or talked with parents more frequently were no more or less likely to value any of the information sources asked in this question. This suggests that even those who do not actively interact with their child's school, or discuss it with others, nevertheless believe such interactions are a valuable source of information about teacher quality.

How teachers should be evaluated

Finally, parents were asked how they believe teachers should be evaluated (question 3 above). In general, the data reveal that parents appreciate the value of several different tools in the teacher-evaluation process. Every tool garnered the support of over 50 percent of respondents, and four of the five tools garnered the support of over 74 percent of respondents. Overall, parents seemed to strongly favor principal observations (93.4%), with parent feedback (88.2%), student feedback (88.0%), and reviews by district officials (74.9%) following. Test scores, however, demonstrated the least support with 57.1 percent of respondents supporting their use in evaluation. Significantly, a sizeable minority of parents (42.9%) felt that test scores should *not* be used to evaluate teachers. These results were similar across all racial groups, although differences did emerge across income groups with regard to principal observations and parent feedback.

Table 4. Parent responses about evaluations

Teacher evaluation mechanisms (all respondents)		
Item	Yes (%)	No (%)
Principal observations	93.4	6.6
Parent feedback	88.2	11.8
Student feedback	88.0	12
Official observations	74.9	25.1
Test scores	57.1	42.9

Table 4. (continued)

Teacher quality information sources (by race)								
Item	Yes (%)			No (%)				
	White	Hispanic	Other	White	Hispanic	Other		
Principal observations	95.6	87.1	90.2	4.4	12.9	9.8		
Parent feedback	88.5	87.1	87.2	11.5	12.9	12.8		
Student feedback	88.6	90.6	84.6	11.4	9.4	15.4		
Official observations	79.4	70.0	61.5	20.6	30.0	38.5		
Test scores	52.0	60.0	66.7	48.0	40.0	33.3		
Parent responses for teacher quality information sources (by income)								
Item	1	2	3	4	1	2	3	4
Principal observations	82.9	92.8	95.9 ^a	97.3 ^b	17.1	7.2	4.1 ^a	2.7 ^b
Parent feedback	94.1 ^a	76.9	91.5 ^b	91.7 ^b	5.9 ^a	23.1	8.5 ^b	8.3 ^b
Student feedback	94.1	86.2	91.5	87.8	5.9	13.8	8.5	12.2
Official observations	81.8	67.7	76.8	77.0	18.2	32.3	23.2	23.0
Test scores	63.2	50.0	50.0	59.7	36.8	50.0	50.0	40.3

Notes: Income bracket 1: \$0–\$49,999; Income bracket 2: \$50,000–\$99,999; Income bracket 3: \$100,000–\$149,999; Income bracket 4: \$150,000+. ^a = p < 0.10, ^b = p < 0.05, ^c = p < 0.01

i replaced * with
superscript a, **
with b and ***with c

Discussion and implications

The results of this study suggest that parents believe effective teachers cannot always, or even primarily, be effectively identified through methodologies relying on student test scores. This is reflected both in their rankings of teacher characteristics and in their opinions about how teachers should be evaluated. Furthermore, it appears that parents believe that they and their children have something valuable to say about teacher quality—something reflected not only in their opinions about how teachers should be evaluated, but also in their articulation of how they themselves are *already* evaluating teachers. Indeed, parents appear to believe they can tell through their day-to-day interactions with teachers, parents, and children something essential about teacher quality. The results of this survey thus raise important questions about the usefulness of standardized test score data as an informational spur to reform, as well as its relative worth when compared with mechanisms relying on ordinary knowledge. In particular, the survey challenges two assumptions underlying test-based accountability regimes: the prevailing ignorance of parents when it comes to teacher performance, and the indispensability of quantitative measurement in social problem-solving.

First, at the core of accountability logic is the belief that educational stakeholders do not have enough reliable, high-quality information about school or teacher performance and require a remedy provided by test score information. This belief is not unique to the field of education, and Scott (1998) has noted that “What is perhaps most striking about [technocratic] schemes [in public policy] ... is how little confidence they repose in the skills, intelligence, and experience of ordinary people”

(p. 346). Similar to Scott's observation, these data suggest that policymakers might perhaps be too cynical in their estimations of parental and community knowledge. Just because the essence of quality teaching may be unintelligible to policymakers far removed from the schoolhouse does not necessarily mean that it is incomprehensible to the people who directly or indirectly engage with the teachers in question. This study suggests that parents, to the degree necessary, believe they are able to obtain information on teacher strengths and weaknesses without advanced analytics, learning instead through their daily experiences and interactions with children, neighbors, and school officials. While these information-gathering processes may not be technically advanced, they do, for all practical purposes, appear to work satisfactorily for the majority of people who depend on them.

A second accountability assumption that this study casts into doubt is the idea that we cannot fix what we cannot measure. This assumption has animated decades of education policy and research, much of which has been consumed by the aspiration of one day creating a comprehensive teacher quality production function capable of identifying which variables best produce particular student outcomes (Hodas, 1993). However, which variables affect student outcomes has been a topic debated at least since the Coleman Report, and the debate will probably never be entirely settled (Coleman et al, 1966). Illustrating the breadth of these variables, parents in this survey seemed to care deeply about teachers' affective traits and temperaments, even though these concerns are extremely difficult to measure and empirical evidence on their importance is mixed (e.g., Berliner & Tikunoff, 1976; Darling-Hammond, 2000; Dobbie, 2011; Duckworth, Quinn, & Seligman, 2009; Ryans, 1960).

There is good reason to believe that this persistent empirical uncertainty will not necessarily preclude educational improvement, however, for history shows that some problems can in fact be solved without understanding them—often by simply drawing on ordinary knowledge (Lindblom & Cohen, 1979). To cite one example of this, doctors during the Civil War learned through experience that cleaning their blades before beginning a new surgery was one way to prevent infection, even though they had no understanding of the principles of bacteriology that undergirded this knowledge. In this same way, the parents in this survey believed that it is possible for them to see and know something essential about teacher quality even without any additional improvements to accountability systems. This belief is not merely an illusion on the part of parents; it has been corroborated in research by other scholars. Caroline Hoxby (2001), for instance, has written that parents are usually able to identify high value-added schools even if they have never encountered official information about those schools' test scores. Edward Fiske & Helen Ladd (2000) and Martin West (2014) have also advanced similar arguments. Consequently, it appears that communal interaction, the vicissitudes of daily experience, and ordinary trial and error may already provide local people with valuable information to assess and improve their schools organically.

Policy implications

This study has a number of policy implications for states and districts as they consider reforming their accountability systems. Since the enactment of the ESSA and

the abolition of NCLB waivers, states and districts have been granted some leeway in the redesign and improvement of their accountability systems. While the ESSA retains the federal requirement for annual testing in reading and mathematics, it also allows states to significantly scale back the role of test scores in evaluating teacher performance. Between 2009 and 2015, a majority of states overhauled their teacher evaluation systems and linked them with standardized test-score performance—often as a condition for securing NCLB waivers. But in the wake of the ESSA, states and districts can now scrap or substantially revise their teacher evaluation systems to make them more reflective of stakeholder concerns. In this environment, this study yields three recommendations for policymakers and researchers aiming to improve teacher evaluation policies in the future: 1) try to link state-produced expert knowledge with ordinary knowledge whenever possible and better frame expert knowledge in ways to facilitate public uptake; 2) consult parents at all stages in the design of accountability systems; and 3) take into consideration the diverse concerns of different constituencies.

First, this study suggests that while test score reports have some informational value, they are not dispositive for parents seeking information about teacher quality, and they therefore need to be augmented with other forms of knowledge about the totality of teacher-quality concerns if they are to have the desired effect in spurring educational change. As this survey suggests, these forms of ordinary knowledge can include personal interactions with parents or school officials, daily experiences with a school, or feedback from students, among others. Without a link to ordinary knowledge, advanced metrics of teacher quality are likely to lack comprehensiveness and be greeted with skepticism, indifference, or even resistance. When state-produced expert knowledge is seen as tangential to local concerns, parents will continue to lean on their social networks and daily experiences for information, perhaps dismissing standardized test scores in the process (Goldring & Rowley, 2006). Thus, policymakers should not assume that the introduction of new official forms of information will overwrite existing ordinary knowledge. Rather, they should try to harness the value of ordinary knowledge and frame new state-created information in ways to facilitate uptake from the public.

By more directly linking state-produced expert knowledge with ordinary knowledge, accountability systems might also begin to yield information that is more “authoritative” as a spur to reform. According to Lindblom and Cohen (1979), authoritative information is that which the recipient feels compelled to act upon, and expert knowledge can achieve authoritativeness either independently or dependently. Independently authoritative information is that for which scientific conclusiveness and consensus has been established. In other words, the truth-value of the information is no longer in question, and action cannot reasonably be postponed. Such authoritativeness is rarely realized, however, since scientific consensus tends to be elusive—especially when the questions at hand involve thorny moral and ethical dilemmas such as those characterizing the social sciences. As a result, expert knowledge is most likely to achieve authoritativeness when it is *dependently authoritative*—that is, when it confirms what is already known or suspected through ordinary and experiential knowledge. In this respect, student test scores are likeliest to

be acted upon when they confirm what is already known or suspected by parents, community members, and school officials on the ground. According to this survey, the hope that expert knowledge can ever entirely dislodge and replace ordinary knowledge is perhaps a bit utopian, and rather than endlessly refining and perfecting the metrics of teacher quality, research should be conducted instead into what social, political, and psychological conditions might best facilitate the revision of ordinary knowledge by expert knowledge.

Second, it is recommended that parents be consulted as often as possible in the design of accountability systems, including the creation of the indicators, the formatting of data reports, and the channels through which the reports will be disseminated. Such efforts would provide opportunities for ordinary knowledge to come into contact with expert knowledge, enhance authoritative status, and increase the probability that parents will take up the new information and act on it. As Fischer (2000) argues, citizen involvement is critical in solving social problems:

While no evidence suggests that the general citizenry can altogether reject the experts and go it alone in a complex society, the citizenry is more intelligent than many politicians and opinion researchers suggest ... Although citizens need experts, the experts – especially policy experts – themselves need citizen assistance much more than their professional ideologies have acknowledged. (pp. 34–35)

Furthermore, citizen involvement can have the added benefit of ensuring that policy enactments have enough democratic legitimacy to persist and be accepted by the people they intend to help (Montpetit, 2007).

In any dialogue with citizens, experts should therefore aspire to assist citizens in efforts to examine their own interests and make their own decisions—not, as is sometimes the case, to provide technical information designed to bring political discussions to an end. One of the most important things policy elites can do is to find ways to “interconnect and coordinate the different but inherently interdependent discourses of citizens and experts” (Fischer, 2000, p. 45). They should ask what data and information parents already have, assess and acknowledge its value, and act to augment or perfect the information ecosystem accordingly. When discourses among experts and citizens are disconnected and different forms of information do not come into contact in the informational ecosystem, scientifically grounded social improvement plans are destined to underwhelm.

Third, while existing accountability regimes take a one-size-fits-all approach in terms of the data published, this study reveals that parents of different racial, language, and economic backgrounds can weight their concerns differently when it comes to teacher quality and evaluation. Modified accountability systems should therefore use an array of indicators that map the distinct concerns of different constituencies so that all groups can receive information on the dimensions of teacher quality they care most about. Additionally, more research should be conducted to probe the diverse assortment of concerns that manifest themselves across demographic groups. While much has been written about the variety of educational goals in America at the macro-level (e.g., Rothstein, Jacobsen, & Wilder, 2008), less work has looked at the differences in response patterns at lower levels of social organiza-

tion. To date, the most relevant body of work in this area has examined parent decision-making in school choice contexts—research that has found that parents have a variety of school-selection criteria, including academic achievement, school climate, pedagogical styles, location, and racial composition, and that the weights attached to these concerns tend to vary depending on the race and income levels of the parents (e.g., Bast & Walberg, 2003; Hanushek, Kain, Rivkin, & Branch, 2007; Hastings, Kane, & Staiger, 2005; Henig, 1990; Lankford & Wyckoff, 2000; Schneider & Buckley, 2002; Teske & Schneider, 2001; Witte, 2000). Little research, though, has parceled out these concerns in the context of accountability systems.

Limitations

There are a few caveats that should be emphasized with regard to the interpretation of these results so that the claims are not stretched beyond reasonable bounds. First, as previously mentioned, this survey was intended to be descriptive in nature. Descriptive surveys are designed merely to highlight interesting patterns or tensions in a particular policy area, and as a result, this survey is not intended to make a definitive statement one way or the other on the benefits or utility of test-based evaluation policies the way a causal survey might. Instead, it is meant to suggest some lines of inquiry that it is hopeful future researchers will explore.

A second limitation has to do with the representativeness of the sample. As noted earlier, because of the descriptive, non-causal nature of the survey, there was less concern about the overall representativeness of the sample than there would have been if it had been a definitive policy evaluation. This sample is relatively representative of the district as a whole, although it over-represents females, high-income individuals, and highly educated individuals. However, the analyses revealed no significant differences in response patterns between male and female respondents or between individuals with different levels of educational attainment. Statistically significant differences emerged on only a few items with respect to income groups, but these findings were consistent with earlier studies that have suggested that wealthier parents are more likely to seek out publicly available testing data (e.g., Hastings, Kane, & Staiger, 2005, 2009; Reay & Ball, 1998). As a result, this sample could perhaps *overstate* the utility of test score data—something that would only reinforce the main conclusions, not repudiate them.

Finally, in this study parents were asked to reflect upon the school experiences of their own child. Given this fact, parents were likely to answer the questions in light of the highly personal experiences they have had with their child, their child's teachers, and other community members associated with their child's school. If parents had been asked these same questions about schools with which they were unfamiliar and to which they had no personal ties, the results may have been different. Moreover, it is not at all self-evident that ordinary knowledge devoid of personal experience is unproblematic as a tool for guiding school-quality inquiry. In fact, when evaluating unfamiliar schools, ordinary knowledge could be less reliable than other sources of information. What this research hopes to convey, though, is the importance of reincorporating ordinary knowledge into the informational ecosystem so that expert and ordinary knowledge work together to improve schooling in America.

Conclusion

In conclusion, this survey suggests that existing metrics of student achievement and teacher quality are not always deemed useful by parents, in part because they do not address many of the characteristics that parents believe are important for teachers to possess. This does not mean that test score data have no role to play in discussions about teacher quality. Indeed, many of the respondents in the survey felt that test score data do have some use in identifying effective teachers and evaluating their performance. Moreover, the value of test score data may turn out to be quite important in shattering inaccurate myths or filling information deficits vis-à-vis the subset of schools with which parents have no direct, personal experiences. However, the data also suggest that existing test score data may need to be revised or supplemented by other measures if they are to be useful and authoritative in broadly spurring change. Expert knowledge does not simply overwrite ordinary knowledge, and how these two types of knowledge interact (and under what conditions) is something that future researchers and policymakers should consider more comprehensively.

References

- Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, 25(1), 95–135. URL: <http://www.jstor.org/stable/10.1086/508733> [May 11, 2015].
- Amrein-Beardsley, A. (2009). The unintended, pernicious consequences of “staying the course” on the United States’ No Child Left Behind policy. *International Journal of Education Policy and Leadership*, 4(6). URL: <http://journals.sfu.ca/ijepl/index.php/ijepl/article/view/199>.
- Ballou, D., & Springer, M.G. (2015). Using student test scores to measure teacher performance some problems in the design and implementation of evaluation systems. *Educational Researcher*, 44(2), 77–86.
- Bast, J.L., & Walberg, H.J. (2004). Can parents choose the best schools for their children? *Economics of Education Review*, 23(4), 431–440.
- Berliner, D.C., & Tikunoff, W.J. (1976). The California beginning teacher evaluation study: Overview of the ethnographic study. *Journal of Teacher Education*, 27(1), 24–30.
- Berryhill, J., Linney, J., & Fromewick, J. (2009). The effects of educational accountability on teachers: Are policies too stress provoking for their own good? *International Journal of Education Policy and Leadership*, 4(5). URL: <http://journals.sfu.ca/ijepl/index.php/ijepl/article/view/99> [May 11, 2015].
- Bethlehem, J. (2009). *Applied survey methods: A statistical perspective*. Hoboken, NJ: John Wiley & Sons, Inc.
- Booher-Jennings, J. (2005). Below the bubble: “Educational triage” and the Texas accountability system. *American Educational Research Journal*, 42(2), 231–268.
- Bridich, S. (2016). The invisible schism: Teachers’ and administrators’ differing perceptions of education reforms. *Education Policy Analysis Archives*, 24(87).
- Brown, K. (2008). Testing the testing: Validity of a state growth model. *International Journal of Education Policy and Leadership*, 3(6). URL: <http://journals.sfu.ca/ijepl/index.php/ijepl/article/view/106> [May 11, 2015].
- Bush, George W. (2001, April 11). Remarks at Concord Middle School in Concord, North Carolina. George W. Bush Public Papers. URL: <http://www.archives.gov/federal-register/publications/presidential-papers.html> [May 11, 2015].
- Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational evaluation and policy analysis*, 24(4), 305–331.
- Carruthers, B.G., & Espeland, W.N. (1991). Accounting for rationality: Double-entry bookkeeping and the rhetoric of economic rationality. *American Journal of Sociology*, 97(1), 31–69.

- Chetty, R., Friedman, J.N., and Rockoff, J.E. (2011). *The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood*. (No. w17699). Cambridge, MA: National Bureau of Economic Research.
- Coleman, J.S., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F., & York, R. (1966). The Coleman Report. *Equality of Educational Opportunity*. [Office of Education (OE-38001 and supp.)] Washington, DC: US Department of Health, Education & Welfare.
- Court of Appeals, Second district, Division 8, California. 2014. *Los Angeles Unified School District v. The Superior Court of Los Angeles County*. Case decided July 23, 2014. URL: <http://caselaw.findlaw.com/ca-court-of-appeal/1673671.html> [May 11, 2015].
- Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives*, 8(1), 1–44.
- Darling-Hammond, L., & Snyder, J. (2015). Meaningful learning in a new paradigm for accountability: An introduction. *Education Policy Analysis Archives*, 23(7), 1–9.
- Darling-Hammond, L., Wilhoit, G., & Pittenger, L. (2014). Accountability for college and career readiness: Developing a new paradigm. *Education Policy Analysis Archives*, 22(86), 1–38.
- De Vaus, D. (2014). *Surveys in social research* (6th ed.). London, UK: Routledge.
- Dobbie, W. (2011). *Teacher characteristics and student achievement: Evidence from Teach For America*. [Unpublished manuscript]. Cambridge, MA: Harvard University.
- Dodson, R. (2015). What makes them the best? An analysis of the relationship between state education quality and principal preparation practices. *International Journal of Education Policy and Leadership*, 10(7), 1–21. URL: <http://journals.sfu.ca/ijepl/index.php/ijepl/article/view/634> [December 1, 2015].
- Duckworth, A.L., Quinn, P.D., & Seligman, M.E. (2009). Positive predictors of teacher effectiveness. *The Journal of Positive Psychology*, 4(6), 540–547.
- Duncan, A. (2009, June 8). *Robust data gives us a roadmap to reform*. Address at the Fourth Annual Institute of Education Science Research Conference. URL: <http://www2.ed.gov/news/speeches/2009/06/06082009.html> [May 11, 2015].
- Duncan, A. (2010, August 25). *Secretary Arne Duncan's remarks at the statehouse convention center in Little Rock, Arkansas*. URL: <http://www.ed.gov/news/speeches/secretary-arne-duncans-remarks-statehouse-convention-center-little-rock-arkansas> [May 11, 2015].
- Duncan, A. (2013, September 9). *Statement by U.S. Secretary of Education Arne Duncan on California bill to exempt millions of students from state assessments*. [Press release]. URL: <http://www.ed.gov/news/press-releases/statement-us-secretary-education-arne-duncan-california-bill-exempt-millions-stu> [May 11, 2015].
- Espeland, W.N., & Stevens, M.L. (2008). A sociology of quantification. *European Journal of Sociology*, 49(3), 401–436.
- Fischer, F. (2000). *Citizens, experts and the environment*. Durham, NC: Duke University Press.
- Fiske, E.B., & Ladd, H.F. (2000). *When schools compete: A cautionary tale*. Washington, DC: Brookings Institution Press.
- Fuhrman, S.H., & Elmore, R.F. (1990). Understanding local control in the wake of state education reform. *Educational Evaluation and Policy Analysis*, 12(1), 82–96.
- Gandara, F., & Randall, J. (2015). Investigating the relationship between school-level accountability practices and science achievement. *Education Policy Analysis Archives*, 23(112), 1–25.
- Goldring, E., & Rowley, K.J. (2006). *Parent preferences and parent choices: The public private decision about school choice*. Paper Presented at the Annual Meeting of the American Educational Research Association.
- Grissmer, D., Beekman, J., & Ober, D. (2014). Focusing on short-term achievement gains fails to produce long-term gains. *Education Policy Analysis Archives*, 22(5), 1–36.
- Haertel, E.H. (2013). *Reliability and validity of inferences about teachers based on student test scores*. The 14th William H. Angoff Memorial Lecture Presented at the National Press Club, Washington, DC, on March 22, 2013. URL: <https://www.ets.org/Media/Research/pdf/PICANG14.pdf> [May 11, 2015].
- Hanushek, E.A., & Raymond, M.E. (2005). Does school accountability lead to improved student performance? *Journal of policy analysis and management*, 24(2), 297–327.

- Hanushek, E.A., Kain, J.F., Rivkin, S.G., & Branch, G.F. (2007). Charter school quality and parental decision making with school choice. *Journal of public economics*, 91(5), 823–848.
- Harris, D.N. (2009). Would accountability based on teacher value added be smart policy? An examination of the statistical properties and policy alternatives. *Education Finance and Policy*, 4(4), 319–350.
- Harris, D.N. (2011). *Value-added measures in education: What every educator needs to know*. Cambridge, MA: Harvard Education Press.
- Hastings, J.S., & Weinstein, J.M. (2008). *Information, school choice, and academic achievement: Evidence from two experiments* (No. w13623). Cambridge, MA: National Bureau of Economic Research.
- Hastings, J.S., Kane, T.J., & Staiger, D.O. (2005). *Parental preferences and school competition: Evidence from a public school choice program* (No. w11805). Cambridge, MA: National Bureau of Economic Research.
- Hastings, J.S., Kane, T.J., & Staiger, D.O. (2009). *Heterogeneous preferences and the efficacy of public school choice* (No. w2145). Cambridge, MA: National Bureau of Economic Research.
- Hazlett, J.A. (1974). *A history of the National Assessment of Educational Progress, 1963–1973: A look at some conflicting ideas and issues in contemporary American education*. [Doctoral dissertation]. Lawrence, KS: University of Kansas.
- Headden, S. (2011). *Inside IMPACT: D.C.'s model teacher evaluation system*. Washington, DC: Education Sector. URL: [http://www.nnstoy.org/download/evaluation/Impact Report Release.pdf](http://www.nnstoy.org/download/evaluation/Impact%20Report%20Release.pdf) [May 11, 2015].
- Henderson, M.B., Peterson, P.E., & West, M.R. (2015). The 2015 EdNext poll on school reform. *Education Next*, 16(1).
- Henig, J. (1990). Choice in public schools: An analysis of transfer requests in magnet schools. *Social Science Quarterly*, 71, 69–82.
- Hodas, S. (1993). Is water an input to a fish? Problems with the production-function model in education. *Education Policy Analysis Archives*, 1(12), 1–7.
- Hoxby, C.M. (2001). If families matter most. In T.M. Moe (Ed.), *A primer on America's schools* (pp. 89–125). Stanford, CA: Hoover Institution Press.
- Jacob, B.A. (2005). Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools. *Journal of Public Economics*, 89(5), 761–796.
- Jacob, B.A., & Lefgren, L. (2005). *Principals as agents: Subjective performance measurement in education* (No. 11463). Cambridge, MA. URL: <http://www.nber.org/papers/11463> [May 11, 2015].
- 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., Snyder, J.W., & Saultz, A. (2014). Informing or shaping public opinion? The influence of school accountability data format on public perceptions of school quality. *American Journal of Education*, 121(1), 1–27.
- Judson, E. (2012). When science counts as much as reading and mathematics: An examination of differing state accountability policies. *Education Policy Analysis Archives*, 20(26), 1–21.
- Kim, J.S., & Sunderman, G.L. (2005). Measuring academic proficiency under the No Child Left Behind Act: Implications for educational equity. *Educational Researcher*, 34(8), 3–13.
- Koretz, D. (2008). The pending reauthorization of NCLB. In G. Sunderman (Ed.), *Holding NCLB accountable: Achieving accountability, equity, & school reform* (pp. 9–26). Thousand Oaks, CA: Corwin Press.
- Ladd, H. F. (2001). School-Based Educational Accountability Systems: The Promise and the Pitfalls. *National Tax Journal*, 54(2), 385–400.
- Laitsch, D. (2002). K-12 voucher programs and education policy. *Education Policy Analysis Archives*, 10(48), 1–39.
- Lankford, H., & Wyckoff, J. (2006). The effect of school choice and residential location on the racial segregation of students. In T. Gronberg & D. Jansen (Eds.), *Improving school accountability* (pp.185–239). London, UK: Emerald Group Publishing.

- Leadership Conference on Civil and Human Rights. (2015). *We oppose anti-testing efforts: Participation in assessments critical for expanding educational opportunity for all students*. [Press release]. URL: <http://www.civilrights.org/press/2015/anti-testing-efforts.html?referrer=http://anhpe.org/2015/05/24/we-cannot-fix-what-we-cannot-measure-civil-rights-groups-come-out-against-opt-out-movement/> [May 11, 2015].
- Lindblom, C. (1959). The science of “muddling” through. *Public Administration Review*, 19(2), 79–88.
- Lindblom, C.E., & Cohen, D.K. (1979). *Usable knowledge: Social science and social problem solving*. New Haven, CT: Yale University Press.
- Lindquist, K. & Whitehurst, R. (2014). Test more, not less [Report]. Washington, DC: Brookings. URL: <https://www.brookings.edu/research/test-more-not-less/>
- Los Angeles Times. (2010). *FAQ & About – Los Angeles teacher ratings*. URL: <http://projects.latimes.com/value-added/faq/> [May 11, 2015].
- Marion, S., & Leather, P. (2015). Assessment and accountability to support meaningful learning. *Education Policy Analysis Archives*, 23(9), 1–19.
- McGuinn, P. (2006). *No Child Left Behind and the transformation of federal education policy, 1965–2005*. Lawrence, KS: University Press of Kansas.
- Milanowski, A., Longwell-Grice, H., Saffold, F., Jones, J., Schomisch, K., & Odden, A. (2009). Recruiting new teachers to urban school districts: What incentives will work? *International Journal of Education Policy and Leadership*, 4(8), 1–13. URL: from <http://journals.sfu.ca/ijepl/index.php/ijepl/article/view/132> [May 11, 2015].
- Misco, T., Patterson, N., & Doppen, F. (2011). Policy in the way of practice: How assessment legislation is affecting social studies curriculum and instruction in Ohio. *International Journal of Education Policy and Leadership*, 6(7), 1–13. URL: <http://journals.sfu.ca/ijepl/index.php/ijepl/article/view/303/112> [May 11, 2015].
- Montpetit, É. (2007). Policy design for legitimacy: Expert knowledge, citizens, time and inclusion in the United Kingdom’s biotechnology sector. *Public Administration*, 86(1), 259–277.
- Nese, J.F.T., Tindal, G., Stevens, J.J., & Elliott, S.N. (2015). The influence of multiple administrations of a state achievement test on passing rates for student groups. *Education Policy Analysis Archives*, 23(70), 1–25.
- Nichols, S., Glass, G., & Berliner, D. (2006). High-stakes testing and student achievement: Does accountability pressure increase student learning? *Education Policy Analysis Archives*, 14(1), 1–175.
- Nichols, S., Glass, G., & Berliner, D. (2012). High-stakes testing and student achievement: Updated analyses with NAEP data. *Education Policy Analysis Archives*, 20(20), 1–35.
- Polikoff, M.S., McEachin, A.J., Wrabel, S.L., & Duque, M. (2014). The waive of the future? School accountability in the waiver era. *Educational Researcher*, 43(1), 45–54.
- Porter, T.M. (1996). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton, NJ: Princeton University Press.
- Power, M. (2004). Counting, control and calculation: Reflections on measuring and management. *Human relations*, 57(6), 765–783.
- Reay, D., & Ball, S.J. (1998). Making their minds up: Family dynamics of school choice. *British Educational Research Journal*, 24(4), 431–448.
- Reese, WJ. (2013). *Testing wars in the public schools: A forgotten history*. Cambridge, MA: Harvard University Press.
- Reports of the annual visiting committees of the public schools of Boston*. (1845). Boston, MA: City Printer. URL: <https://babel.hathitrust.org/cgi/pt?id=uiug.30112046172869;view=1up;seq=4> [May 11, 2015].
- Rothstein, R., Jacobsen, R., & Wilder, T. (2008). *Grading education: Getting accountability right*. Washington, DC: Economic Policy Institute.
- Ryans, D. (1960). Characteristics of teachers, their description, comparison, and appraisal: A research study. Washington, DC: American Council on Education.
- Schneider, M., & Buckley, J. (2002). What do parents want from schools? Evidence from the Internet. *Educational evaluation and policy analysis*, 24(2), 133–144.
- Scott, J.C. (1998). *Seeing like a state: How certain schemes to improve the human condition have failed*. New Haven, CT: Yale University Press.

- Sedlak, M.W. (1989). Let us go and buy a school master: Historical perspectives on the hiring of teachers in the United States, 1750-1980. In D.R. Warren (Ed.), *American teachers: Histories of a profession at work* (pp. 273–277). New York, NY: Macmillan.
- Schneider, J. (2011). *Excellence for all: How a new breed of reformers is transforming America's public schools*. Nashville: Vanderbilt University Press.
- Stone, D. (2011). *Policy paradox: The art of political decision-making*. New York, NY: W.W. Norton & Company.
- Supovitz, J. (2009). Can high stakes testing leverage educational improvement? Prospects from the last decade of testing and accountability reform. *Journal of Educational Change*, 10(2–3), 211–227.
- Teske, P., & Schneider, M. (2001). What research can tell policymakers about school choice. *Journal of Policy Analysis and Management*, 20(4), 609–631.
- United States Senate. (2001, January 10). *Hearing of the Committee on Health, Education, Labor and Pensions of the United States Senate, One Hundred Seventh Congress, First Session on 'Roderick Paige Of Texas, To Be Secretary of Education.'* Washington, DC: United States.
- Warren, D.R. (1974). *To enforce education: A history of the founding years of the United States Office of Education*. Detroit, MI: Wayne State University Press.
- West, M.R. (2014, October 23). Why do Americans rate their local public schools so favorably? *The Brown Center Chalkboard*. Washington, DC: Brookings Institution.
- Winters, M.A., & Cowen, J.M. (2013). Who would stay, who would be dismissed? An empirical consideration of value-added teacher retention policies. *Educational Researcher*, 42(6), 330–337.
- Witte, J. (2000). *The market approach to education: An analysis of America's first voucher program*. Princeton, NJ: Princeton University Press.