Stéphane Lavertu
The Ohio State University

We All Need Help: "Big Data" and the Mismeasure of Public Administration

Stéphane Lavertu is associate professor in the John Glenn College of Public Affairs, The Ohio State University. Much of his research examines how politics affects the design and operation of public agencies. He is particularly interested in how politics affects the administration of public education.

E-mail: lavertu.1@osu.edu

Abstract: *Rapid advances in our ability to collect, analyze, and disseminate information are transforming public administration. This "big data" revolution presents opportunities for improving the management of public programs, but it also entails some risks. In addition to potentially magnifying well-known problems with public sector performance management—particularly the problem of goal displacement—the widespread dissemination of administrative data and performance information increasingly enables external political actors to peer into and evaluate the administration of public programs. The latter trend is consequential because external actors may have little sense of the validity of performance metrics and little understanding of the policy priorities they capture. The author illustrates these potential problems using recent research on U.S. primary and secondary education and suggests that public administration scholars could help improve governance in the data-rich future by informing the development and dissemination of organizational report cards that better capture the value that public agencies deliver.*

Practitioner Points

- Although it is widely acknowledged that flawed performance measurement abounds and poses a problem for public administration, the notion that "big data" can make matters worse is seldom addressed.
- The performance measurement that greater data availability enables can lead to misperceptions about organizational performance and flawed decision making, particularly among actors external to public agencies (e.g., policy makers and citizens).
- Public administration scholars could help make the best of our data-rich future by playing a more active role in researching and designing performance metrics for external audiences so that performance information promotes more valid inferences about the value of organizations administering public programs.

The Brownlow Committee famously declared that "[t]he President needs help." In particular, the committee's report asserted that the president needed staff to "assist him in obtaining quickly and without delay all pertinent information possessed by any of the executive departments so as to guide him in making his decisions" (PCAM 1937, 5). The committee considered this collection and synthesis of administrative information critical for improving the efficiency and accountability of American government. Importantly, advances in our ability to collect and analyze administrative data facilitated the expanded role of such analysis in the management of the federal executive branch, as illustrated by the implementation of the Planning, Programming, and Budgeting System in the 1960s and, later, the implementation of performance-based "New Public Management" reforms (Reschenthaler and Thompson 1996).

Advances in information technology and analytical techniques have also facilitated the proliferation of policy research well beyond the federal executive branch (Radin 2000). They increasingly enable political actors external to public agencies—such as legislators, judges, administrators in other agencies, researchers in think tanks and advocacy organizations, and citizens—to observe and evaluate administrative behavior at all levels of government. These external actors have demonstrated sustained demand for information on administrative performance (Gormley and Weimer 1999; Hood, Dixon, and Beeston 2008; Van de Walle and Roberts 2008), and technological advances—that is, advances in our ability to collect, analyze, and disseminate information—increasingly enable fulfillment of that demand.[1] The widespread belief that these technological advances will continue to accelerate suggests that external political actors' use of performance metrics to assess the value of organizations administering public programs also will continue to increase significantly.[2]

Commentators have emphasized the opportunities that "big data" and "data analytics" present for policy

analysis and management (see, e.g., Cook 2014; George, Hass, and Pentland 2014; McAfee and Brynjolfsson 2012; Pirog 2014) as well as for enhancing transparency and accountability (e.g., Fung 2013), but serious risks come with the expansion of organizational performance measurement that technological innovation permits. The performance management literature has demonstrated that the incentives tied to performance metrics may not motivate the intended administrative behavior (see, e.g., Dixit 2002; Heinrich and Marschke 2010). Additionally, research has documented how political motives can lead to the construction and dissemination of organizational performance measures that promote incorrect inferences about the value of public programs (see, e.g., Moynihan 2008; Radin 2006). Because technology facilitates the construction, dissemination, and use of organizational performance measures, technological advances could very well exacerbate administrative problems associated with flawed performance measures and incentives.

But there are additional risks associated with the "big data" revolution that receive less attention. First, more data and improved analytical techniques do not necessarily attenuate the problems inherent in measuring organizational performance.[3] In particular, the problem of goal displacement—when administrators direct more effort toward tasks associated with measured outcomes—can worsen as the availability of data expands. More and better data enable one to reduce the proportion of unmeasured outcome dimensions and to increase the validity of performance metrics, for example, by helping to identify statistically the contributions of public agencies to societal outcomes. But more accurate and precise measurement of some outcome dimensions could worsen the problem of goal displacement by increasing *differences* in the precision of performance measures for a given organization or program (see, e.g., Holmstrom and Milgrom 1991).

Second, the massive expansion of data availability, often in the name of transparency and accountability, puts performance measures in the hands of relatively uninformed individuals who, compared with administrators, may have little sense of what quantities performance measures actually capture. In particular, external audiences may be more likely to misinterpret or place undue weight on performance metrics when making policy decisions.[4] Public administration scholars have long pointed out that presidents, agency executives, and Congress need help to improve their analysis and use of administrative data (e.g., Dror 1967; PCAM 1937; Radin 2006), but rapid technological advances in data collection and analysis—and the corresponding increases in the availability of data in the name of transparency and accountability—are quickly increasing the extent to which all political actors in and out of government need such analytical help.

This article illustrates these points using U.S. primary and secondary education as a case. This case is particularly illuminating because scholars have employed rigorous research methods to identify the impact of organizational performance measurement on the thoughts and behaviors of both external political actors and internal actors administering school districts, as well as to identify the significant

> Public administration scholars have a role to play in promoting performance measurement that enhances deliberation about the value of organizations administering public programs.

impact of these behaviors on administrative decisions and outcomes. The case of education also illustrates the growing role of technical experts in prioritizing public goals through performance measurement. After reviewing some of this research, I suggest that public administration scholars have a role to play in promoting performance measurement that enhances deliberation about the value of organizations administering public programs. In particular, I suggest that we are in a position to assist the information processing and decision making of external political actors by developing and disseminating guidance for designing organizational report cards that better capture the value that public agencies deliver.

The purpose of this article is to draw attention to a public problem by illustrating it using rigorous empirical research and to stoke debate about what public administration's role should be in addressing it. The wave of enthusiasm for "big data" in public administration is the most recent manifestation of continued efforts to base public decision making on measured quantities and to enhance transparency and accountability. Public administration scholars are quite familiar with some of the risks associated with such initiatives. Yet we devote far more energy to using existing performance metrics as dependent variables than to improving them.[5] This article illustrates some of the risks of current trends in performance measurement—emphasizing a couple that have received little attention by public management scholars—in the hope of making them more salient in discussions about "big data," and it encourages public administration scholars to engage political actors in the hope that they can help improve the quality and communication of government performance information.

In the spirit of the Brownlow Committee decades ago—and consistent with Lasswell's (1951) "policy scientist of democracy" and Weimer and Vining's (2011) policy analyst promoting the good society—public administration scholars could dedicate themselves to ensuring that the policy-relevant information that external political actors increasingly receive through performance metrics yields valid inferences about the value that public agencies deliver. They can do this by influencing the design of performance metrics and perhaps by creating their own. Working to address the risks I outline in this article would not only help fulfill a professional obligation but also help reassert public administration's "historical sense of purpose" and distinctiveness as a design science (Barzelay and Thompson 2010, S295).

Big Data and the Risk of Goal Displacement

Researchers have documented how performance management systems can fail to incentivize productive administrative behaviors if they do not capture accurately and precisely an organization's (or an individual's) contribution to all of the administrative outcomes it is tasked with pursuing (Heinrich and Marschke 2010). Two often-noted concerns are that failure to measure performance with respect to some outcome dimensions leads to goal displacement and that failure to accurately and precisely identify organizational contributions to societal outcomes can weaken administrative incentives to engage in value-adding behavior. Technological advances lower

the cost of measurement and thus, in principle, could improve performance measurement by improving goal coverage as well as the accuracy and precision with which administrative outcomes are captured.

However, if technological advances lead to imbalances in the precision of performance metrics—for example, if they enable more precise measurement of organizational value added on one dimension without improving measurement precision on other dimensions—then technological advances could incentivize even greater focus on certain goals at the expense of others. In other words, the imbalance in precision with which dimensions of performance are measured may motivate goal displacement in organizations that administer public programs (Holmstrom and Milgrom 1991). Recent developments in U.S. primary and secondary education illustrate how technological advances that improve the accuracy and precision of performance measures can have just such an effect.

Quite recently, the performance of U.S. public schools was evaluated primarily by local stakeholders, many of whom were in close contact with those schools. But local school district governance changed dramatically near the turn of the twenty-first century with a flurry of reforms imposing school and district accountability systems. For example, consistent with accountability schemes already in place in many states (Dee and Jacob 2011), the federal No Child Left Behind (NCLB) Act of 2001 required states to disseminate school- and district-level "report cards" based primarily on a handful of indicators of academic quality—particularly proficiency rates on standardized mathematics and reading examinations, attendance rates, and graduation rates—in order to determine whether schools were making "adequate yearly progress" (AYP) toward promoting student proficiency in tested subjects (Manna 2011).[6] States employed some of their own performance measures to account for additional dimensions of school quality, but report cards primarily captured the subset of academic outcomes that can be measured using low-cost, machine-graded standardized examinations in reading and mathematics. Unsurprisingly, there was a noticeable shift in school resources toward these measured outcomes (Chakrabarti 2014; Reback, Rockoff, and Schwartz 2014).

The wide dissemination of these school- and district-level report cards just prior to every school year—a development made possible in large part by technological improvements in data analysis and electronic communication—significantly increased the performance information available to actors not directly involved in the administration of districts and schools. But the AYP metric used in the NCLB accountability system generally failed to capture many outcomes of interest such as higher-order-thinking math and reading skills, achievement in other subject areas, and a variety of other public goals related to student health and safety, civil rights, and citizenship. Additionally, the AYP performance metric largely captured absolute achievement levels as opposed to how much knowledge schools and districts imparted during a given school year. Consequently, AYP was highly correlated with school and district poverty levels (Kogan, Lavertu, and Peskowitz 2015b), and schools were most likely to fail to reach the AYP performance benchmark because of low proficiency rates among their economically disadvantaged students (Davidson et al. 2013).

Many schools in impoverished neighborhoods were deemed deficient because few of their students could meet the established proficiency targets, even if those schools imparted more knowledge during the school year than high-performing schools in high-income neighborhoods (Downey, Von Hippel, and Hughes 2008; Kogan, Lavertu, and Peskowitz 2015b). Recognizing that educational improvements should be rewarded even if proficiency goals are not met, some states began to employ state-of-the art technology in data collection and analysis to introduce "value-added" measures of school and district performance. Value-added measures improve the identification of school and district performance in tested subjects because they take account of each student's prior academic achievement when estimating the impact of schools and districts on test performance. There remain some significant weaknesses in these measures' ability to identify the contribution of schools and districts to student test outcomes, but these value-added measures represent a marked improvement over the use of absolute achievement levels in terms of measuring school and district contributions to student learning.

Policy makers have displayed enthusiasm for value-added measures based on student performance in math and reading, and econometricians have spent considerable time developing and evaluating techniques for identifying the value added by districts, schools, principals, and, in particular, teachers (see Chetty, Friedman, and Rockoff 2014a; Dieterle et al. 2015; Grissom, Kalogrides, and Loeb 2015; Rothstein 2014). The use of these advanced statistical methods, which can be employed because of major investments in data collection and the massive expansion of computing power, enable external political actors to peer into and seek to affect the management of school districts. For example, lists ranking teachers based on value-added metrics have been made publicly available,[7] and federal and state policy makers and administrators have sought to use these metrics to manage school district staff from afar—primarily by encouraging or requiring the use of value-added metrics in teacher and principal evaluations (see, e.g., NCTQ 2012). Indeed, charter schools in a number of states are required to shut down if their students perform poorly according to these value-added metrics (Carlson and Lavertu 2015).

Thus, improvements in our ability to measure organizational contributions to student learning in mathematics and reading have increased the importance of those subjects in school district administration. Value-added estimates represent a substantial improvement over proficiency rates when it comes to measuring school and district contributions to student achievement in math and reading, but this improvement seems to have fueled even greater focus on a subset of organizational goals in the administration of schools. Additionally, the wide dissemination of these value-added measures has increased the likelihood that external stakeholders—who are often less likely than internal actors to appreciate their limitations—will use them in their decision making. There is some evidence that these developments might motivate improvements in educational quality on some important dimensions,[8] but it is also

> Value-added estimates represent a substantial improvement over proficiency rates when it comes to measuring school and district contributions to student achievement in math and reading.

clear that they are leading to the reallocation of resources toward a smaller subset of goals that schools are tasked with pursuing.

Big Data and External Political Actors

Local control over primary and secondary education is an American political tradition, but state and federal policy makers have centralized school district governance significantly over the past century. In particular, the enactment of NCLB marked perhaps the biggest leap in federal involvement since passage of the Elementary and Secondary Education Act in 1965 (Manna 2006, 2011; McGuinn 2006). Much of this expansion of the federal role is attributable to NCLB's attempt to hold districts accountable for their educational outcomes, which was made possible by technological advances that lowered the cost of collecting, analyzing, and disseminating information on student achievement. Performance-based accountability systems also facilitated the centralization of governance at the state level, in part because NCLB required states to design and administer these systems. Because of NCLB and similar state-level reforms, political actors at all levels of government, as well as actors outside of government, have more information with which to judge the quality of school district administration.

The wide dissemination of school performance information is consistent with a general societal desire to improve American democracy through increased transparency. Proponents of the widespread dissemination of administrative data tend to emphasize the potential benefits for empowering citizens (e.g., Fung 2013). Indeed, lawmakers deliberating over the reauthorization of the Elementary and Secondary Education Act regularly mention the empowerment of parents, advocacy groups, and policy makers as a key benefit of the law.[9] But there is at least one significant problem with providing external political actors with measures of administrative performance: they are less likely than internal administrative actors to understand how to interpret those measures and less likely to know how much weight to assign to them for the purpose of governing school districts.

Political actors external to school districts—such as district residents, advocacy groups, governors and presidents, and state and federal lawmakers—are in a position to direct the administration of school districts. But few among them truly understood how the federal "adequate yearly progress" metric was constructed. The metric—which simply indicated whether a school or district "met" the evolving AYP performance benchmark—was based on a relatively complicated aggregation process. Some external actors knew that in order to receive the "AYP met" designation, a school's students needed to exceed a minimum proficiency rate in both math and reading and that the proficiency rate requirements increased over time. But fewer people understood that students in each of 10 federally defined demographic subgroups—according to race and ethnicity, disability status, economic status, and so on—also needed to attain those minimum average proficiency rates. Most importantly, public discourse surrounding school and district report cards makes clear that few external actors understood that districts were seldom in a position to improve their performance as measured by AYP—that it was a highly inaccurate measure of district performance on measured outcomes, as I discussed earlier. Indeed, the lawmakers who wrote the NCLB required states to use the AYP metric for accountability purposes.

Consider the general public as the external actor of interest. Research indicates that district residents generally rate their local schools highly (West 2014) and that the variation in their assessments correlates with published proficiency rates and report card ratings (Charbonneau and Van Ryzin 2012; Favero and Meier 2013; West 2014). Research also indicates that state-sanctioned performance ratings have a causal impact on district residents' views regarding the quality of their school districts (Chingos, Henderson, and West 2012) and that the introduction of school and district report cards was associated with a decline in public support for schools in New York City (Jacobsen, Snyder, and Saultz 2013). Importantly, public perceptions of school quality were responsive to proficiency-based performance measures despite the fact that much of the public clearly values administrative outcomes other than student achievement in math and reading (see, e.g., Jacobsen, Snyder, and Saultz 2014b).

These studies on district residents' information processing suggest that voters might use report card metrics such as AYP when making decisions that affect the administration of their local school districts. Indeed, Kogan, Lavertu, and Peskowitz (2015b) provide convincing evidence that the AYP metric had a causal impact on voter decisions to approve or reject district tax proposals in Ohio—provided that the report cards were released in sufficiently close proximity to the elections. Specifically, voters were 10 percent more likely to reject tax proposals if districts failed AYP. In another study, the authors document that such failed attempts to raise tax revenues caused significant reductions in district instructional spending and student achievement in the very subject areas that the federal government prioritized when establishing the NCLB accountability system (Kogan, Lavertu, and Peskowitz 2015a). Moreover, because AYP was arguably more a measure of student poverty than of school quality, this voter response likely undermined the NCLB goal of narrowing achievement gaps between student demographic groups.

District administrators, on the other hand, were far less likely to assign weight to the AYP metric when making administrative decisions. After all, they had access to and likely were far more familiar with the underlying measures used to calculate the AYP designation. The AYP designation itself provided them with no additional information about the performance of their students. Additionally, superintendents and school board members typically have access to qualitative and quantitative information on district staff and student performance that were not used in AYP calculations. It is unsurprising, therefore, that Kogan, Lavertu, and Peskowitz (2015c) found no causal relationship between an Ohio district's or school's AYP designation and the turnover of district superintendents and principals. Moreover, there appears to be no relationship between superintendent and principal turnover and value-added estimates of student achievement (Grissom and Andersen 2012; Kogan, Lavertu, and Peskowitz 2015c). This finding is consistent with research indicating that school boards and superintendents consider factors beyond student achievement when making personnel decisions (Grissom and Andersen 2012; Hess and Meeks 2010).

Although external political actors may lack some critical contextual information and may be more prone to misinterpret and misapply performance measures (see, e.g., Olsen 2015), developments associated with big data can introduce interpretative challenges for

administrators as well. For example, there is much confusion even among district personnel regarding what value-added estimates of school and teacher quality actually capture. Thus, as the expansion of data collection and storage capacity accelerates, data analysts increasingly are in a position to affect policy with the choices they make. With respect to value-added measures of teacher and school quality, for example, analysts make choices about which student observations to include, how many prior years of student achievement to account for, and how to specify the statistical models used to generate the estimates—choices that can have a significant impact on estimates of teacher value (see, e.g., Sass, Semykina, and Harris 2014). With respect to AYP, federal and state administrators, in consultation with technical experts, made decisions about what constituted a "proficient" student, the year-to-year changes in AYP requirements, and how student subgroups figured into the overall AYP designation. As a result, technical experts had an important role in making decisions about how to balance goals related to equity and effectiveness, and external political actors reacted punitively without understanding what these measures actually captured.

Big Data and the Role of Experts

Observers of all political stripes seem to understand that K–12 education policy and administrative decisions in the United States are increasingly based on measured outcomes that capture a very narrow subset of public goals. But there also seems to be some political agreement that basing decisions on some quantitative measures is better than basing them on no quantitative measures. Additionally, there is little appetite for generating nuanced pictures that do not reduce organizational performance to a single dimension. Quite the contrary, external political actors seem to prefer even less nuanced measures that assign schools overall letter grades—sometimes allowing public debate to bypass difficult issues about the proper balance among educational goals.[10]

Summary measures are not necessarily problematic; there are good reasons to simplify performance information in this way. For example, "information overload" (Radin 2000) is a very real problem that might lead people to ignore performance information entirely, particularly among voters who have other priorities and cannot afford to bear information collection and processing costs (Downs 1957). Providing one summary measure—such as a single letter grade for a school or district—can efficiently communicate a lot of information and might compensate for citizens' tendencies to use other informational shortcuts, such as school board candidates' race and gender (see, e.g., Lorinskas, Hawkins, and Edwards 1969; Matson and Fine 2006; McDermott 1997). But the creation and use of summary measures means that elite political actors and methodological experts have a growing say in which public goals performance measures prioritize.

In the case of education, many of the most influential policy experts have concluded that the administration of public schools should be focused on promoting achievement on testable knowledge and skills in

> As the expansion of data collection and storage capacity accelerates, data analysts increasingly are in a position to affect policy with the choices they make.

> Few people have the knowledge necessary to assess the validity of performance metrics, which likely introduces significant information failures in the market for school performance information.

mathematics and reading, largely because of the correlation with economic growth (e.g., Hanushek, Peterson, and Woessmann 2013). Additionally, as I mentioned earlier, measurement experts make many decisions about how to construct the summary outcome measures on which external political actors rely. When it came to setting yearly AYP proficiency cut scores, for example, states often put their faith in technical advisory committees composed largely of academic and professional psychometricians. Decisions about how to estimate school and teacher value added are sometimes delegated to private contractors charged with conducting the calculations, and economists and other social scientists have played a significant role in deciding on model specifications and on what data to base the estimates. To their credit, many of these experts have sought to educate nonexperts about the policy implications of various methodological choices (see, e.g., Harris 2011). But the point remains that because of advances in technology, experts increasingly exercise significant influence on the perceived value of the individuals and organizations administering public programs.

Thus, although the design of performance measurement systems is, in theory, an opportunity to deliberate over public goals (Moynihan 2008), the highly technical process of constructing simple summary measures is precluding some of this deliberation. Policy makers and experts—who, in this case, are inclined to emphasize the economic returns to schooling—can inject a skew toward certain goals into performance measurement schemes. Some might see the use of test-based report card measures as an opportunity to "nudge" school governance in the right direction (see, e.g., Thaler and Sunstein 2008; Weil, Graham, and Fung 2013). For example, value-added measures focused on testable skills in math and reading may help "compensate" for parents' prioritizing of safety and social considerations—priorities that may be attributable to those outcomes being relatively easy for parents to observe. But this sort of manipulation introduces some very serious concerns about the relationship between democracy and expertise.

How Public Administration Scholars Might Help

The private market is unlikely to supply performance information that promotes valid inferences about the value of organizations administering public programs (see Gormley and Weimer 1999). For example, with regard to U.S. primary and secondary education, consumers are likely to demand information about the private benefits that schools provide (e.g., student achievement and safety) as opposed to information that captures the public goals that schools are charged with realizing (e.g., citizenship and equal opportunity). Moreover, as I argued earlier, few people have the knowledge necessary to assess the validity of performance metrics, which likely introduces significant information failures in the market for school performance information. That means that government and private nonprofit actors should take seriously the task of constructing and disseminating performance information that promotes valid inferences about the value of organizations administering public programs.

Researchers have investigated many issues related to the generation and dissemination of organizational performance information (in particular, see Gormley and Weimer 1999), and one hopes that government and private nonprofit actors will apply this knowledge. For example, scholars across a number of disciplines are investigating the impact of various methods of information delivery on external audiences. With regard to K–12 public education, Jacobsen, Snyder, and Saultz (2014a) found significant causal differences in how people interpret school and district performance ratings depending on the scale used (e.g., whether the summary designation is a letter grade or a numerical performance index), and Kogan, Lavertu, and Peskowitz (2015b) found suggestive evidence that the timing of report card releases affects whether district residents take account of that information when casting a vote. Additionally, as I discussed earlier, economists continue to develop tools to address issues related to the statistical identification of schools' contribution to student learning. Unfortunately, those designing performance measurement schemes do not always use the insights that research has provided (Heckman et al. 2011), so there is some need for practice-oriented scholars to transmit some of that knowledge.

What is particularly lacking in scholarly efforts outside of public administration is a systematic study of the dimensions of quality on which public organizations should be judged. In education, there is some advocacy along those lines (see, e.g., Howe and Murray 2015), but those designing performance measurement systems need advice from experts who have a relatively comprehensive view of public value. As this article argues, rapid advances in data collection, analysis, and dissemination can further goal displacement if the range of outcomes that an organization is tasked with pursuing are not accounted for with comparable accuracy and precision—particularly when external actors who lack contextual knowledge are interpreting and using performance information. Identifying the range of goals that administrative organizations are tasked with pursuing is a necessary condition if one wishes to construct and disseminate performance measures in such a way as to minimize goal displacement, whether that entails drawing attention to unmeasured goals (e.g., by providing a sense for the proportion of organizational outcomes a summary performance designation captures) or formally incorporating those goals in performance metrics (e.g., by finding some way to capture them).

Public administration scholars have a significant contribution to make in part because they have taken it upon themselves to identify the range of politically legitimate goals that organizations are charged with pursuing. For example, researchers recently have sought to characterize the various dimensions of "public value" (see, e.g., Bozeman 2007; Bryson, Crosby, and Bloomberg 2014; Moore 2013, 2014). Significant difficulty arises when one tries to identify the range of legitimate public goals that organizations pursue (Lindblom 1959). Measuring performance at the organizational level compounds this problem because organizations typically have varying levels of responsibility and discretion in administering multiple programs. Consequently, identifying organizational goals and characterizing organizational responsibility and discretion is difficult—perhaps more so now than ever because of the complex governance arrangements in existence (Moynihan et al. 2011). Nevertheless, it is a goal worth pursuing if one is concerned about maximizing the value of organizations administering public programs. That it is difficult to identify the range of goals for which administrative agents are responsible is partly why public administration scholars stand to make an important contribution.

There are benefits to having multiple, competing frameworks for identifying and weighting the outcome dimensions on which to base measures of organizational performance. Indeed, promoting deliberation over what constitutes a set of politically legitimate goals for each and every organization administering public programs would be valuable. However, one should be able to convince a wide breadth of political actors that a framework is legitimate—otherwise, it has little prospect of influencing how organizational value added is estimated. Thus, for the purpose of performance measurement, the goals identified probably need to reflect the public policy commitments that formal democratic institutions have codified. These should probably include the non-mission-based goals that Rosenbloom (2007) and Wichowsky and Moynihan (2008) emphasize, for example, but some goals might be a tough sell to policy makers.

Additionally, public administration scholars probably would have greater impact if they went one step further and articulated procedures for capturing those dimensions of value in an organizational report card. A guidance document, for example, could describe procedures for (1) identifying the mission-based and non-missed-based goals that a public agency is responsible for (taking into account agency authority or discretion), (2) weighting or ranking those responsibilities (for the purpose of aggregation or presentation), (3) measuring organizational performance on each of those dimensions, (4) characterizing the validity of performance information on each dimension, and (5) communicating all of this information succinctly and clearly to external audiences. Such guidance could encourage the creation of better metrics as well as better communication about the quality of those metrics—both of which are important pieces of information for external political actors to consider. Researchers could also apply such guidelines to generate their own report cards in areas of substantive interest to them, or they could take it upon themselves to assess the quality of existing performance ratings (see Gormley and Weimer 1999; Hood, Dixon, and Beeston 2008), just as Arndt (2008) and Langbein and Knack (2010) have done with the World Bank's indexes.

Finally, public administration scholars might go even further by engaging political actors directly, to ensure that "big data" and the expansion of performance measurement serve to communicate a more accurate portrait of public value. Engagement could take many forms. Perhaps one of the field's professional associations or the National Academy of Public Administration could establish a committee of senior scholars and practitioners with the necessary expertise (in democratic theory, performance measurement, the communication of performance information, data visualization, the behavioral responses of various audiences, and so on) to develop a performance measurement guidance document targeted at lawmakers and agency officials. The diversity in academic training and views of public administration scholars and practitioners surely would make it difficult to draft such a document, but the payoff could be worth it in terms of drawing attention to these issues.

At the very least, we probably need more vigorous academic conversations on these matters and need to take seriously our function as a bridge between the academy and practice. So long as the academic

conversation taking place through journals concerns itself with understanding and solving performance measurement problems, it may be enough for scholars to transmit these insights through their teaching and training of public servants, engagement of political actors through legislative testimony or technical advisory teams, and work with other academics and private organizations (e.g., the Pew Charitable Trusts and its efforts to measure performance in election administration[11]) that have an impact on the design of performance measurement systems.

My purpose here is merely to stoke debate, as opposed to providing fully developed proposals. Whatever engagement strategy public administration scholars adopt, we bear some responsibility for mobilizing our interdisciplinary expertise to develop solutions for policy makers. By doing so, we stand to make a significant and positive contribution to public programming and societal outcomes. Many researchers see such applied work and advocacy as conflicting with efforts to develop a rigorous scientific enterprise (for a review, see Raadschelders 2011), but this problem-solving orientation is all that ever really set public administration apart from the positive social sciences. Indeed, even Herbert Simon believed that professional schools (such as schools of public administration) should engage in an applied design science (Barzelay and Thompson 2010).

Concluding Thoughts

From Luther Gulick (through the Brownlow Committee's report) to those seeking to shape the emerging field of policy analysis (see, e.g., Dror 1967), public administration scholars have long concerned themselves with improving the interpretation and use of administrative data by actors external to public agencies. Working to improve the extent to which organizational report cards communicate the value that public agencies deliver is a natural extension of those earlier efforts. Additionally, if we are in position to address the risks that "big data" presents, then it is our professional obligation to do so. By adopting a problem orientation and serving a role akin to Lasswell's (1951) "policy scientist of democracy" or Weimer and Vining's (2011) policy analyst working to promote the good society, public administration scholars can help ensure that the growth in policy-relevant information that external political actors receive through organizational performance metrics indeed promotes democratic accountability and efficient governance.

There are significant challenges associated with translating increasingly complex technical expertise for external political audiences (Fischer 2009). There is also some inherent tension between Lasswell's conception of an elite policy expert and the goal of empowering citizens (see, e.g., Farr, Hacker, and Kazee 2008). But, as I discussed earlier, "big data" is already increasing the role of experts in performance measurement. And many citizens and policy makers believe that the widespread dissemination of performance information is a great way to communicate expert assessments of the quality of public administration in an effort to improve accountability and efficiency. Thus, to address the problems associated with the mismeasure of public administration, it seems best to ensure that a more complete conception of public value is incorporated into the instrumental logic of modern public administration.[12]

Acknowledgments

Notes

1. Van de Walle and Roberts (2008) emphasize that political, economic, and cultural conditions feed the demand for performance measurement of decentralized delivery services, particularly in education and health. They quite rightly point out that this demand is a necessary condition for there to be a relationship between technological change and performance measurement. Additionally, they question why ratings systems have not been developed in some policy areas, such as immigration services, prisons, and welfare programs. They have, but these ratings systems are geared toward political principals as opposed to "customers."

2. The notion that technological progress will continue to accelerate has been promoted in popular discourse by a number of individuals, perhaps most prominently by Ray Kurzweil (2005). There are numerous detractors, of course, as extrapolating historical trends is a questionable enterprise. Nevertheless, the notion that the growth in our ability to store and analyze data is exponential—a "law" proposed by Intel's Gordon Moore—has been borne out. See, for example, McAfee and Brynjolfsson (2012).

3. See Gormley and Weimer (1999) for a careful review of these measurement problems.

4. Public administration research has thoroughly documented this problem as it occurs within public agencies employing performance management practices. For example, Radin (2015) argues that recent events in Baltimore may be attributable in part to the police department's internal use of crime statistics made possible by "big data" (see O'Malley 2014). But failure to account for context when using performance information is even more likely to occur when performance information is disseminated to actors external to public agencies.

5. For example, see Gerrish's (2016) meta-analysis of research linking management to measured performance. This is unsurprising, as researchers—including myself—are prone to relying on preexisting quantitative measures.

6. I refer to the NCLB performance regime in the past tense because most states have received waivers from the law to implement a different accountability system. It is worth noting, however, that NCLB's AYP provisions are still in place and that they apply to states if they fail to obtain and renew their federal waivers.

7. For example, the *Los Angeles Times*'s website makes publicly available teacher value-added estimates for all teachers in the Los Angeles school districts and provides rankings by teacher and school. See http://projects.latimes.com/value-added/ (accessed July 21, 2015).

8. For example, Deming et al. (2014) find that measures of quality such as value added are correlated with postsecondary student success, but so are other dimensions of school quality. Chetty, Friedman, and Rockoff (2014b) found that students with higher-quality teachers as measured by value added enjoyed greater incomes in adulthood.

9. For example, consider recent comments by Senator Pat Murray as Congress debated the reauthorization of the Elementary and Secondary Education Act. See Lauren Camera, "Sen. Patty Murray Outlines NCLB Reauthorization Priorities, *Education Week,* January 13, 2015, http://blogs.edweek.org/edweek/campaign-k-12/2015/01/sen_patty_murray_outlines_nclb.html (accessed July 21, 2015).

10. For example, as indicated by data from the Education Commission of the States, between 2002 and 2013, the number of states that assigned schools a single letter grade in their report cards went from 1 to 14 (see http://www.ecs.org/html/educationissues/accountability/stacc_intro.asp, accessed July 15, 2014). That said, some states are looking to develop some more nuanced report cards. For example, Virginia recently repealed its A–F rating system.

11. See, for example, Pew's Elections Performance Index, http://www.pewtrusts.org/en/multimedia/data-visualizations/2014/elections-performance-index (accessed July 21, 2015).

12. In other words, attempts to promote a more accurate picture of public programs probably must work within the "dominant rationality of neoliberal governance" that motivates "New Public Management" practices, as Dahl and Soss (2014, 496) put it.

References

Arndt, Christiane. 2008. Politics of Governance Ratings. *International Public Management Journal* 11(3): 275–97.

Barzelay, Michael, and Fred Thompson. 2010. Back to the Future: Making Public Administration a Design Science. Special issue, *Public Administration Review* 70: S295–97.

Bozeman, Barry. 2007. *Public Values and Public Interest: Counterbalancing Economic Individualism.* Washington, DC: Georgetown University Press.

Bryson, John M., Barbara C. Crosby, and Laura Bloomberg. 2014. Public Value Governance: Moving Beyond Traditional Public Administration and the New Public Management. *Public Administration Review* 74(4): 445–56.

Carlson, Deven, and Stéphane Lavertu. 2015. The Effect of School Closure on Student Achievement: Evidence from Ohio's Automatic Charter School Closure Law. Research paper, John Glenn School of Public Affairs, The Ohio State University. http://glenn.osu.edu/educational-governance/research/ [accessed July 14, 2015].

Chakrabarti, Rajashri. 2014. Incentives and Response under No Child Left Behind: Credible Threats and the Role of Competition. *Journal of Public Economics* 110(1): 124–46.

Charbonneau, Étienne, and Gregg G. Van Ryzin. 2012. Performance Measures and Parental Satisfaction with New York City Schools. *American Review of Public Administration* 42(1): 54–65.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014a. Measuring the Impact of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review* 104(9): 2593–2632.

———. 2014b. Measuring the Impact of Teachers II: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review* 104(9): 2633–79.

Chingos, Matthew M., Michael Henderson, and Martin R. West. 2012. Citizen Perceptions of Government Service Quality: Evidence from Public Schools. *Quarterly Journal of Political Science* 7(4): 411–45.

Cook, Thomas D. 2014. "Big Data" in Research on Social Policy. *Journal of Policy Analysis and Management* 33(2): 544–47.

Dahl, Adam, and Joe Soss. 2014. Neoliberalism for the Common Good? Public Value Governance and the Downsizing of Democracy. *Public Administration Review* 74(4): 496–504.

Davidson, Elizabeth, Randall Reback, Jonah Rockoff, and Heather L. Schwartz. 2013. Fifty Ways to Leave a Child Behind: Idiosyncrasies and Discrepancies in States' Implementation of NCLB. Working Paper no. 18988, National Bureau of Economic Research. http://www.nber.org/papers/w18988 [accessed July 14, 2015].

Dee, Thomas S., and Brian Jacob. 2011. The Impact of No Child Left Behind on Student Achievement. *Journal of Policy Analysis and Management* 30(3): 418–46.

Dieterle, Steven, Cassandra M. Cuarino, Mark D. Reckase, and Jeffrey M. Wooldridge. 2015. How Do Principals Assign Students to Teachers? Finding Evidence in Administrative Data and the Implications for Value Added. *Journal of Policy Analysis and Management* 34(1): 32–58.

Dixit, Avinash. 2002. Incentives and Organizations in the Public Sector. *Journal of Human Resources* 37(4): 696–727.

Deming, David J., Justine S. Hastings, Thomas J. Kane, and Douglas O. Staiger. 2014. School Choice, School Quality, and Postsecondary Attainment. *American Economic Review* 104(3): 991–1013.

Downey, Douglas B., Paul T. von Hippel, and Melanie Hughes. 2008. Are "Failing" Schools Really Failing? Using Seasonal Comparison to Evaluate School Effectiveness. *Sociology of Education* 81(3): 242–70.

Downs, Anthony. 1957. *An Economic Theory of Democracy.* New York: Harper.

Dror, Yehezkel. 1967. Policy Analysts: New Professional Role in Government Service. *Public Administration Review* 27(3): 197–203.

Farr, James, Jacob S. Hacker, and Nicole Kazee. 2006. The Policy Scientist of Democracy: The Discipline of Harold D. Lasswell. *American Political Science Review* 100(4): 579–87.

Favero, Nathan, and Kenneth J. Meier. 2013. Evaluating Urban Public Schools: Parents, Teachers, and State Assessments. *Public Administration Review* 73(3): 401–12.

Fischer, Frank. 2009. *Democracy and Expertise: Reorienting Policy Inquiry.* Oxford, UK: Oxford University Press.

Fung, Archon. 2013. Infotopia: Unleashing the Democratic Power of Transparency. *Politics and Society* 41(2): 183–212.

George, Gerry, Martine Hass, and Alex Pentland. 2014. Big Data and Management. *Academy of Management Journal* 57(2): 321–26.

Gerrish, Edwin. 2016. The Impact of Performance Management on Performance in Public Organizations: A Meta-Analysis. *Public Administration Review* 76(1): 48–66.

Gormley, William T., and David L. Weimer. 1999. *Organizational Report Cards.* Cambridge, MA: Harvard University Press.

Grissom, Jason A., and Stephanie Andersen. 2012. Why Superintendents Turn Over. *American Educational Research Journal* 49(6): 1146–80.

Grissom, Jason A., Demetra Kalogrides, and Susanna Loeb. 2015. Using Student Test Scores to Measure Principal Performance. *Educational Evaluation and Policy Analysis* 37(1): 3–28.

Hanushek, Eric A., Paul E. Peterson, and Ludger Woessmann. 2013. *Endangering Prosperity: A Global View of the American School.* Washington, DC: Brookings Institution Press.

Harris, Douglas N. 2011. *Value-Added Measures in Education: What Every Educator Needs to Know.* Cambridge, MA: Harvard Education Press.

Heckman, James J., Carolyn J. Heinrich, Pascal Courty, Gerald Marschke, and Jeffrey Smith. 2011. *The Performance of Performance Standards.* Kalamazoo, MI: W. E. Upjohn Institute for Employment Research.

Heinrich, Carolyn J., and Gerald Marschke. 2010. Incentives and Their Dynamics in Public Sector Performance Management Systems. *Journal of Policy Analysis and Management* 29(1): 183–208.

Hess, Frederick M., and Olivia Meeks. 2010. School Boards circa 2010: Governance in the Accountability Era. https://www.nsba.org/sites/default/files/SBcirca2010_WEB.pdf [accessed July 14, 2015].

Holmstrom, Bengt, and Paul Milgrom. 1991. Multitask Principal–Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. Special issue, *Journal of Law, Economics, and Organization* 7: 24–52.

Hood, Christopher, Ruth Dixon, and Craig Beeston. 2008. Rating the Rankings: Assessing International Rankings of Public Service Performance. *International Public Management Journal* 11(3): 298–328.

Howe, Kenneth R., and Kevin Murray. 2015. Why School Report Cards Merit a Failing Grade. Report, National Education Policy Center, University of Colorado Boulder. http://nepc.colorado.edu/publication/why-school-report-cards-fail [accessed July 14, 2015].

Jacobsen, Rebecca, Jeffrey W. Snyder, and Andrew Saultz. 2013. When Accountability Strategies Collide: Do Policy Changes That Raise Accountability Standards Also Erode Public Satisfaction? *Education Policy* 27(2): 360–89.

———. 2014a. 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.

———. 2014b. Understanding Satisfaction with Schools: The Role of Expectations. *Journal of Public Administration Research and Theory* 25(3): 831–48.

Kogan, Vladimir, Stéphane Lavertu, and Zachary Peskowitz. 2015a. The Impact of Local Tax Referenda on School District Administration and Student

Achievement. Research paper, John Glenn School of Public Affairs, The Ohio State University. http://glenn.osu.edu/educational-governance/research/ [accessed July 14, 2015].

———. 2015b. Performance Federalism and Local Democracy: Theory and Evidence from School Tax Referenda. *American Journal of Political Science.* Published electronically on May 6. doi: 10.1111/ajps.12184.

———. 2015c. Do School Report Cards Produce Accountability through the Ballot Box? Research paper, John Glenn School of Public Affairs, The Ohio State University. http://glenn.osu.edu/educational-governance/research/ [accessed July 14, 2015].

Kurzweil, Ray. 2005. *The Singularity Is Near: When Humans Transcend Biology.* New York: Viking Press.

Langbein, Laura, and Stephen Knack. 2010. The World Governance Indicators: Six, One, or None? *Journal of Development Studies* 46(2): 350–70.

Lasswell, Harold D. 1951. The Policy Orientation. In *The Policy Sciences: Recent Developments in Scope and Method,* edited by Daniel Lerner and Harold D. Lasswell, 3–43. Stanford, CA: Stanford University Press.

Lindblom, Charles E. 1959. The Science of "Muddling Through." *Public Administration Review* 19(2): 79–88.

Lorinskas, Robert A., Brett W. Hawkins, and Stephen D. Edwards. 1969. The Persistence of Ethnic Voting in Urban and Rural Areas: Results from a Controlled Election Method. *Social Science Quarterly* 49(3): 491–99.

Manna, Paul. 2006. *School's In: Federalism and the National Education Agenda.* Washington, DC: Georgetown University Press.

———. 2011. *Collision Course: Federal Education Policy Meets State and Local Realities.* Washington, DC: CQ Press.

Matson, Marsha, and Terry Susan Fine. 2006. Gender, Ethnicity, and Ballot Information: Ballot Cues in Low-Information Elections. *State Politics and Policy Quarterly* 6(1): 49–71.

McAfee, Andrew, and Erik Brynjolfsson. 2012. Big Data: The Management Revolution. *Harvard Business Review* 90(10): 60–66, 68, 128.

McDermott, Monika L. 1997. Voting Cues in Low-Information Elections: Candidate Genders as a Social Information Variable in Contemporary United States Elections. *American Journal of Political Science* 41(1): 270–83.

McGuinn, Patrick J. 2006. *No Child Left Behind and the Transformation of Federal Education Policy, 1965–2005.* Lawrence: University Press of Kansas.

Moore, Mark H. 2013. *Recognizing Public Value.* Cambridge, MA: Harvard University Press.

———. 2014. Public Value Accounting: Establishing the Philosophical Basis. *Public Administration Review* 74(4): 465–77.

Moynihan, Donald P. 2008. *The Dynamics of Performance Management: Constructing Information and Reform.* Washington, DC: Georgetown University Press.

Moynihan, Donald P., Sergio Fernandez, Soonhee Kim, Kelly M. LeRoux, Suzanne J. Piotrowski, Bradley E. Wright, and Kaifeng Yang. 2011. Performance Regimes amidst Complexity. Supplement 1, *Journal of Public Administration Research and Theory* 21: i141–55.

National Council on Teacher Quality (NCTQ). 2012. State of the States 2012: Teacher Effectiveness Policies. http://www.nctq.org/dmsView/State_of_the_States_2012_Teacher_Effectiveness_Policies_NCTQ_Report [accessed July 14, 2015].

Olsen, Asmus Leth. 2015. Citizen (Dis)satisfaction: An Experimental Equivalence Framing Study. *Public Administration Review* 75(3): 469–78.

O'Malley, Martin. 2014. Doing What Works: Governing in the Age of Big Data. *Public Administration Review* 74(5): 555–56.

Pirog, Maureen A. 2014. Data Will Drive Innovation in Public Policy and Management Research in the Next Decade. *Journal of Policy Analysis and Management* 33(2): 537–43.

President's Committee on Administrative Management (PCAM). 1937. *Administrative Management in the Government of the United States.* Washington, DC: U.S. Government Printing Office.

Raadschelders, Jos C. N. 2011. *Public Administration: The Interdisciplinary Study of Government.* Oxford, UK: Oxford University Press

Radin, Beryl A. 2000. *Beyond Machiavelli: Policy Analysis Comes of Age.* Washington, DC: Georgetown University Press.

———. 2006. *Challenging the Performance Movement.* Washington, DC: Georgetown University Press.

———. 2015. Baltimore: When Good Intentions Bring Negative Consequences. *Public Administration Review* 75(4): 511–12.

Reback, Randall, Jonah Rockoff, and Heather L. Schwartz. 2014. Under Pressure: Job Security, Resource Allocation, and Productivity in Schools under No Child Left Behind. *American Economic Journal: Economic Policy* 6(3): 207–41.

Reschenthaler, G. B., and Fred Thompson. 1996. The Information Revolution and the New Public Management. *Journal of Public Administration Research and Theory* 6(1): 125–44.

Rosenbloom, David H. 2007. Reinventing Administrative Prescriptions: The Case for Democratic-Constitutional Impact Statements and Scorecards. *Public Administration Review* 67(1): 28–39.

Rothstein, Jesse. 2014. Revisiting the Impacts of Teachers. Working paper, Goldman School of Public Policy and Department of Economics, University of California, Berkeley. http://eml.berkeley.edu/~jrothst/workingpapers/rothstein_cfr.pdf [accessed July 14, 2015].

Sass, Tim R., Anastasia Semykina, and Douglas N. Harris. 2014. Value-Added Models and the Measurement of Teacher Productivity. *Economics of Education Review* 38(1): 9–23.

Thaler, Richard, and Cass Sunstein. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness.* New Haven, CT: Yale University Press.

Van de Walle, Steven, and Alasdair Roberts. 2008. Publishing Performance Information: An Illusion of Control? In *Performance Information in the Public Sector: How It Is Used,* edited by Wouter Van Dooren and Steven Van de Walle, 211–26. Basingstoke, UK: Palgrave Macmillan.

Weil, David, Mary Graham, and Archon Fung. 2013. Targeting Transparency. *Science* 340(6139): 1410–11.

Weimer, David L., and Aidan R. Vining. 2011. *Policy Analysis: Concepts and Practice.* Upper Saddle River, NJ: Longman.

West, Martin R. 2014. Why Do Americans Rate Their Local Public Schools So Favorably? Paper, Brookings Institution. http://www.brookings.edu/research/papers/2014/10/23-local-public-schools-west [accessed July 14, 2015].

Wichowsky, Amber, and Donald P. Moynihan. 2008. Measuring How Administration Shapes Citizenship: A Policy Feedback Perspective on Performance Management. *Public Administration Review* 68(5): 908–20.