I am honored to receive APPAM’s Kershaw award. The list of people who deserve credit for it is very long; it includes my teachers and colleagues at the Maxwell School at Syracuse University where I received my degree, those who nominated me and those who selected me, and my coauthors over the years, and my colleagues at the La Follette School at the University of Wisconsin–Madison.

Rigorous research in the common ground between policy and management can be fruitful and offer profound research opportunities. This idea is somewhat at odds with an underlying tension in the field of public affairs between the study of policy and administration, an artificial separation more a result of scholarly division of labor than a reflection of the actual policy process.

Policy studies rest upon, sometimes implicit, assumptions of administrative quality and preferences, assumptions often not recognizable to those who directly study public administrators. In some cases, administration is considered as a variable only to the degree it represents a troubling deviance, for example as when we talk about fidelity concerns in the implementation of evaluation-inspired policies. Administration is also sometimes “rediscovered” in policy circles as Dwight Waldo observed in his characterization of implementation studies in the 1970s, and the current emphasis on implementation science as a distinct area raises similar concerns. At an extreme, policy designs are treated as self-implementing, assumed to have the same effect regardless of administrative context.

Public administration scholarship too often treats the policy setting as a given, not considering how policy design affects administrative variables. At an extreme, scholars treat policy as an object administrative actors implement, subvert, or ignore, a theme I will refer to later. Scholars give too little attention to the basic question of how administrative factors can be designed to facilitate the success of policy goals.

Three Areas of Interplay between Policy and Management

My research cuts across a variety of areas, looking at managerial aspects of policy areas such as public health insurance, homeland security, and the running of elections. To make sense of this work I will use the political science policy feedback model as an organizing framework that subverts the classic assumption that policy...
is a function of political choice. Instead the model conceives of the relationship between policy and politics as an ongoing interplay, where each shapes the other over time. In particular, this model helps to identify three ways in which policy and administration connect that I use to categorize my research.

The first area arises from unpacking the central claim of policy feedback that has sometimes been simply stated as “policy shapes politics,” a stark alternative to the claim that “politics shapes policy” that, in turn, implies the subclaim “administration shapes politics.” This assertion directs scholars to study not just how political forces impinge on administration but also how administrative organizations act on and transform different aspects of politics, ranging from basic access to public goods and services; political relations between the citizen and the state, including trust in government and the sense of civic standing the individual has in relation to the state; and individuals’ sense of political efficacy and civic engagement.

The second and third relationships reflect where, within the nexus of policy and management, the interplay resides: that administration shapes policy and that policy shapes administration. While at a superficial level, we might be equally willing to concur that “policy matters” and “management matters,” such an agreement masks different beliefs about which is more profoundly important in governance, and fails to consider the implications of how policy and management matter to each other. Rather than engaging in an exhaustive characterization of the ways in which this relationship works, I will explain how my work has offered some novel insights into the interplay of causal arrangements.

Relationship 1: The Interplay of Policy and Administration Affects Politics

Let me return to the first relationship. The concept of policy feedback suggests that policies can transform the political landscape. Policies, in this view, are political forces that reconfigure the underlying terms of power, reposition actors in political relations, and reshape political identities, understandings, interests, and preferences. Policies, and administration of those policies, have consequences for whether citizens receive resources they seek and how individuals understand their broader civic roles.

The effects of policies and administration in politics can occur in straightforward ways. Perhaps the most obvious ways are in how policies and administration structure democratic participation. With colleagues from the University of Wisconsin–Madison Department of Political Science, I published a paper in the *American Journal of Political Science* that examines how different election rules generated different turnout levels. Although current debate on election policy focused on voter ID laws, our work pointed to the relative importance of another aspect of policy design and administration. Allowing individuals to register on the same day they vote removes an administrative hurdle—registering to vote some weeks before the election—and turns voting into a one-stop shopping experience more likely to draw in voters who might become engaged in elections late in the race. We showed that greater access to this seemingly minor administrative service significantly increased turnout in elections, and other work suggests those turnout gains are higher for populations that are more likely to move frequently: poorer and minority voters.

Formal policies that structure access to administrative resources key to voting are one thing, but we identified in another paper the direct importance of discretion of election officials. We found that when local election officials in Wisconsin had ideological views different from the majority of their constituents, turnout was lower. For example, when these appointees were conservative, turnout was lower if these officials oversaw elections where voters tended to favor President Obama. While these individuals filled officially nonpartisan positions, the results illustrate how bureaucratic discretion in expanding the pool of registered voters and making access to voting easier can affect election outcomes.

How policies and administration shape citizens’ civic capacities can occur in less obvious ways. In her model of policy feedback, Suzanne Mettler hypothesized that policies matter to civic capacities by the types of resources they provide to citizens and by the experiences citizens face in receiving those services. Bureaucratic encounters can affect both of these variables. My recent work has examined how administrative burdens citizens encounter alter access to valued goods or services. These encounters may also teach citizens lessons about the state, mark them in politically consequential ways, alter their political capacities, and reposition them in relation to other citizens and dominant institutions.

Although aspects of administrative burden appear in different streams of prior research, the area is not conceptualized in a way that enables a consistent understanding of the topic, unifies existing knowledge, or lays out a coherent future research agenda. My research on this topic has involved several coauthors, including Pamela Herd, Tom DeLeire, Elizabeth Rigby, and Hope Harvey, and argues that this burden is large and composed of three types of costs:

- **Learning costs** arise from engaging in search processes to collect information about public services and how they are relevant to the individual. Large proportions of people eligible for a program are unaware of that program, believe incorrectly they are ineligible, or underestimate the benefits they would accrue.
- **Psychological costs** include the stigma of applying for or participating in a program with negative perceptions, a sense of loss of personal power or autonomy in interactions with the state, and the stresses of dealing with...
administrative processes. Individuals not wanting to experience disempowering processes or be seen as dependent may avoid such psychologically distressing encounters. Those who persevere may nonetheless experience a loss of well-being.

**Compliance costs** are the burdens, including time and money, of following administrative requirements, e.g., the costs of completing forms, or documenting status when applying for a service.

These burdens are important. First, *administrative burdens affect program take-up*. Compared to the near 100 percent take-up for universal programs like Social Security and Medicare, estimates of take-up rates by eligible beneficiaries of means-tested programs in the United States are much lower. The central difference between means-tested and universal programs is that the former must do more to distinguish between the eligible and ineligible, and their administrative processes impose greater burdens that even eligible applicants fail to negotiate.

Behavioral economics research points to cognitive biases that help to explain why burdens that seem relatively small compared to the benefits at stake are consequential in terms of whether a citizen accesses a service he or she may need. For example, in our study of Medicaid in Wisconsin, a seemingly modest requirement asked recipients to document that they lacked affordable employer-based insurance. Officials expected this change to affect 2 to 3 percent of applicants. It resulted instead in an enrollment drop of 20 percent for children and 17.6 percent for adults. Subsequent investigations found that the vast majority of those affected by the policy change were eligible. It was not eligibility, but rather the failure to negotiate the administrative processes that explains this dramatic loss of access to program benefits. More generally, we offer evidence that efforts by the state to reduce burden increased enrollment in public health insurance.

This issue is also important as it questions whether the creation of such *burdens are simply policymaking by other means*. Are the creation and reduction of burden, as well as the distribution of burden between the state and the individual, venues where politics plays out? Empirical studies of program take-up have not addressed this question, and in the field of public administration the existence of red tape has generally been attributed to historical accident or incompetence. Analysis of the unfurling of burdens in Medicaid in Wisconsin suggests that the preferences of political actors about a policy may shape what they perceive as the appropriate level of burden in that policy area—how successive governors viewed public insurance in turn reflected administrative decisions to reduce or increase burdens. Beyond Wisconsin, a simple analysis of the complexity of Medicaid forms across state governments suggest that partisan control of government predicts the compliance burdens that applicants face. Policymakers in such instances appear to use burdens as an alternative or complement to more overt forms of policymaking. The attractiveness of administrative burdens for such purposes is tied to their opacity. The details of administration that give rise to burden—such as the level of documentation an application form requires—may be largely invisible to the public and even most policymakers, and their impact may be poorly understood.

Finally, observing the magnitude of these burden costs raises the question, *how can burdens be reduced?* And when is it appropriate to do so? In some cases burdens may serve valid public values and should not be reduced. But in some cases the value of burdens are negligible or are vastly outweighed by their costs, motivating the search for methods to reduce burdens or ameliorate their negative effects. For example, our research on Wisconsin points to a variety of techniques. Outreach to systematically reach target populations reduced learning costs. Marketing campaigns to present Medicaid as health insurance for the working poor, rather than a form of welfare, were intended to reduce the stigma of such programs. Online applications, simplification of forms, and less frequent recertification helped to reduce compliance costs. A particularly promising avenue of research is to document practices and program designs that reduce burden without undercutting other values. Information technology and governmental data systems make it more feasible for states to shift burdens from the citizen to the state, using state data to identify and document the eligibility of individuals.

A practical implication of this work is the degree to which the state should monitor burdens. The move toward reliance on third-party government contractors has sometimes left the state struggling to articulate its role. The reduced oversight of third parties may give rise to their using burdens as a means to engage in cream-skimming if incentives align. Governmental actors largely contract for services on the basis of some measure of performance, and often some indicator of quality and financial indicators. However, these actors could also articulate administrative burdens as a normative criterion around which they monitor how citizens interact with the provider of governmental services. This work is recent but points to an area where policy and management scholars should have a shared interest: Burdens are co-created by policy designs and administrative implementation, and they have profound impact on policy outcomes.

**Relationship 2: Administration Affects Policy**

A second aspect of the policy-administrative relationship is the degree to which administrators take advantage of information asymmetries in shaping policy. This concern is a mainstay of street-level bureaucracy and implementation studies of how policies are implemented.

The design of policy may rest on incorrect or simplistic assumptions about how administrative processes matter, making close attention to these processes all the more important. My work has looked at the operation of administrative instruments intended to reduce information asymmetry and to encourage goal alignment between principals and agents. This effort is reflected in formal requirements to put in place systems that create and disseminate performance data. Governmentwide reforms, such as the Government Performance and Results Act (GPRA) of 1993, the Bush-era Program Assessment Rating Tool, and the current GPRA Modernization Act are prime examples of the creation of performance regimes. But the turn to performance measures is also embedded in policy-specific
changes, such as the No Child Left Behind Act and Race to the Top in the field of education.

In my book *The Dynamics of Performance Management*, I write “It is only a slight exaggeration to say that we are betting the future of governance on the use of performance information. The current era is also characterized by a willingness to adopt new structural forms of government and controls, such as networks or outsourcing, or to simply provide greater freedom to managers. New structural forms and modes of control raise difficult questions. How do we coordinate? How do we manage? How do we control? How do we account? How do we improve? How do we engage citizens? Performance information is frequently cited as the answer. We are told that performance information will allow elected officials and policymakers to set goals. It will provide the basis for accountability. It will be tied to incentives. It will allow innovations to be identified and diffused. It will enable the allocation of scarce public resources. It will allow citizens to give feedback on services. The one constant in future visions of government is the availability and smart use of performance information. If performance information does not prove to be the linchpin for the future of governance, we will have to return to the basic questions listed above, and find some alternative answers.”

If there is a single lesson from my body of work in this area, it is that the implementation of such systems is complex, and such complexity suggests that they cannot easily bear the heavy weight that they are assigned in contemporary models of governance.

My research on the adoption of performance systems in state governments documented one important way that actual policies depart from the New Public Management doctrine from which they are drawn. This doctrine called for managers to be held to higher expectations of performance and to be given more autonomy to achieve goals. By examining changes in performance reporting requirements, but also on basic personnel and financial controls, I find that state governments adopted the first of these prescriptions, but not the second, hence placing managers in a context where they face expectations for greater performance, while having gained little in the way of tools to achieve this goal.

The theory of such reforms is that more performance information will lead to better decisions, because performance information is objective, standardized, indicative of actual performance, and consistently understood, and it prompts a consensus about how a program performs and how it should be funded. Some observers have noted that such goals are unrealistic, and my book offers an alternative theoretical framework to understand the role of performance measures in policymaking, based on research at the state and federal level.

What I characterized as the interactive dialogue model argued that performance data were not comprehensive—for any complex program, there are multiple ways of capturing performance, some of which will and will not be formalized. People could look at the same policy area and reasonably disagree on what were appropriate measures. Performance data also have the quality of ambiguity—the same data can generate different interpretations. Performance information tells us nothing about context, implementation, or factors that shape how we interpret whether a program is effective. For example, people could look at the same data on schools and come to different conclusions on the meaning of the data and on what to do next. This element of ambiguity encourages subjectivity about performance, where actors associated with specific institutions, parties, or worldviews use performance data to present the version of the world they favored.

In short, performance data cannot be reasonably expected to replace politics or erase information asymmetries in the policy process. Instead, the most reasonable expectation was that it becomes part of the language in an interactive dialogue on the meaning of performance, shaped by different worldviews, interests, and power.

My work has also made the case that to understand the functioning of performance systems, we need effective ways of tracking the effects of these systems in terms of their behavioral effects on intended audiences. I made the case for studying the use of performance information as such a variable. While performance information use is a broad concept, there are a variety of types of uses, such as perverse use of data, that increase measured performance at the expense of underlying goals, or political use of data to advocate for or against programs. The key practical challenge for performance systems is how to encourage purposeful use that would foster greater innovation and effectiveness and limit perverse use.

A number of scholars have joined in this work, amassing a growing body of knowledge on a variety of environmental, organizational, job, and individual factors associated with performance information use. My own contribution focuses on two dimensions. One is to make the case that, consistent with the interactive dialogue model, politics did not undercut the use of performance data as some had suggested, but could encourage use. For example, one study of school superintendents in Texas showed that those with more fractious, diverse, and engaged stakeholders were more likely to use data, partly to respond to the variety of claims that stakeholders made, and to turn to what is perceived as a non-partisan source of information to justify their performance. In highly partisan contexts, however, politics can limit the effects of performance reforms.
In a series of papers, I showed that while the Bush administration’s Program Assessment Rating Tool (PART) initiative appeared to benefit conservative agencies in terms of encouraging performance information use, individuals exposed to PART saw an increase in performance information use relative to peers in their agencies not exposed to PART and relative to employees in liberal agencies. Employees of liberal agencies experienced PART as more burdensome, a finding based on subjective survey measures (even after accounting for individual ideological beliefs) and on objective measures of the number of performance assessments and improvement plans required of liberal agencies. I concluded in a 2013 summary of this work: “reforms are more likely to succeed if the political executive is preaching to the ideological choir and more likely to falter among ideological heretics. But the more negative possibility … is that reforms are used as a mechanism to investigate and burden the heretics.”

A second dimension of my contribution to this growing research area arose from an observation from The Dynamics of Performance Management: Although performance systems often promise lofty changes, they usually were limited to rules about measuring and disseminating data. They did little to institutionalize routines of using data. In aggregate, my research suggested that routines created by the 1993 GPRA and the PART did not facilitate greater use of performance data. Perhaps because of this troubling experience, the more recent GPRA Modernization Act explicitly sought to inculcate routines of use, for example in the form of requirements for quarterly meetings around key goals. In a paper presented at the 2014 Association of Public Policy and Management conference, I offered evidence that exposure to these organizational routines is associated with greater use of performance data for decisions at the individual level. This finding suggests the possibility for performance systems to be adapted over time by investing in organizational factors and professional values rather than putting faith in mechanisms of external control.

Scholarship on performance information use has increasingly evolved to move from reliance on cross-sectional survey-based designs to experimental work. While survey-based work remains critical for assessment of governmentwide reforms, experiments offer better causal assessments and can utilize more objective indicators of whether managers used performance tools. In recent work relying on survey experiments I offered evidence about how the framing of performance data influenced budget decisions that subjects were asked to make. While both treatment and control groups were shown the same data, treatment groups that were given additional information responded to performance data differently. For example, one experiment provided data that suggested good program performance to all subjects, but the treatment group was also provided target information that showed that although performance had been improving, it had not kept up with a very demanding target schedule. Respondents given somewhat arbitrary target information were likely to be less generous in making budget allocations, even though the underlying performance had not changed. Similarly, another experiment added comments from a policy advocate to the treatment group, which served to alter how respondents understood the performance data and how they allocated resources. The results are consistent with the interactive dialogue model characterization of performance information use as a form of behavior that is circumstantial, with individuals using performance data to alter allocations depending on the nature of the data and how they are presented.

There is even greater potential to understand how performance data are used via field experiments. Colleagues in Denmark and I found that randomized allocations of greater autonomy on personnel decisions to school principals was associated with greater tendency for those principals to download performance data—as part of the experiment, data are made available to principals, but they were not told that researchers were observing whether they downloaded it. A basic problem with performance data is that the data by themselves offer no sense of context about whether a level of performance is good or bad. We also found that if the data made available included comparisons with other schools, which provided such a context, principals were more apt to download more data. Consistent with the interactive dialogue model, we also found that individual preferences about the appropriate goals of education affected the tendency to utilize performance data.

There remains a need for more research in this area. One generally implicit assumption about the promise of performance systems is that concerns about the “black box” of management in the policy process will be made less relevant, since we can use external indicators to direct the black box without having to worry too much about its internal workings. This approach is largely wrong—elected officials pay little systematic attention to performance data, and so the value of the information depends upon how it is used by those inside the black box, and that process is as mysterious as most others.

To use a medical metaphor, the changes that sought to rewire the brain and surgically restructure the body of governance relied on primitive notions of how the brain worked. The success of policy initiatives and new structural approaches to government depends on a better understanding of how performance information is processed into shared beliefs and actions in policymaking. We also need to understand the neurological wiring—at the individual level and, in a metaphorical sense, the organizational and policy levels—that makes that the use of performance data more or less likely.

If the central goal is actually improved performance, an insight of my work suggests the use of techniques and frameworks differ from the dominant agency-theory inspired approach that emphasizes external accountability and extrinsic reward. Instead, the use of performance data seems to be greater for purposeful ends and less susceptible to perverse use when public employees are given autonomy to do their jobs, when such systems appeal to altruistic motivations, and when performance data become embedded in organizational routines that center on learning. Such a claim hits a discordant note in the prevailing chorus for how performance systems should be organized.
Relationship 3: Policy Affects Administration

The third relationship I consider is how policies shape administration. Much of administrative scholarship, including my own work, often treats policy as an unvarying background, or as an object of administrative action if considered at all. The potential for administration to shape policy is real, as reflected in the discussion of the second proposition above. But we should also be willing to complement this insight with its reverse. As organizations implement a policy, they transform it and are themselves transformed.

Administrators shape policy outcomes to be sure, but policies also have the power to disrupt and reconfigure administration. They can restructure authorities, change tasks, alter routines, redistribute resources, and reframe culture, identity, and motivation. The implication of this point is, as I write with Joe Soss, “that the study of administration today is distorted by a one-sided focus. The fruitful investigation of how administrators shape policy has blinkered researchers to the ways that bureaucracies are remade by the policies they administer.”

It is easiest to observe the effects of policy on administration in areas where there has been significant policy change. One such area is election administration, which has seen significant policy change since the 2000 election, with the passage of the Help America Vote Act (HAVA), and subsequent and ongoing state and local policy changes. The passage of HAVA helps to explain a puzzle that I had identified: What explained the popularity of electronic voting systems despite well-documented flaws? The answer is partly tied to incentives provided by HAVA, which encouraged such adoption. Once adopted, these technologies then enjoyed the benefits of path dependency. Adopters tended to display a series of cognitive biases to defend and maintain the selection of electronic voting machines. Using the first national survey of election officials, collected in partnership with the Congressional Research Service, I found that even after controlling for perceived benefits and costs of different technology, the perceived value of electronic voting systems was shaped by a series of individual biases, including faith in technology, confidence in one’s judgment not based on empirical inferences, a confirmation bias that screened out critical information about electronic voting machines, and a basic preference for the status quo.

HAVA changed the workplace of election administrators in other ways. My work with Carol Silva demonstrated that election officials saw a dramatic increase in work in the aftermath of HAVA. A more in-depth analysis of the lot of election officials in Wisconsin used a mixture of qualitative interviews, focus groups, administrative data, and surveys to demonstrate the degree to which HAVA had made the work of these officials more burdensome and less attractive.

Increased workloads associated with policy changes might result in organizational changes such as higher turnover and redistribution of organizational resources. But these changes might also alter bureaucratic beliefs and preferences in important ways. As election officials found Wisconsin election policies designed to make voting more convenient to be onerous, they tended be more critical of these policies on unrelated dimensions. They tended to perceive such policies as reducing electoral integrity, were more likely to incorrectly believe that such policies failed to reach their goal of increasing turnout, and were more apt to oppose additional policy changes that would have further increased access to the polls. The effects of electoral policies in one time period altered how administrators evaluated these policies, in some cases leading to demonstrable errors in evaluations, and cementing opposition toward new policy ideas. As public officials play crucial roles as policy implementers and as stakeholders, the effects of policies on bureaucratic beliefs and assessments matter.

Another aspect of my work on the influence of policies on administration comes from a National Science Foundation-funded study on disaster response. I identified the adoption, first voluntary but later mandated, of a managerial technology, the Incident Command System, by the federal government. Any federal agency with response capability (as well as state and local responders receiving federal dollars, and the private and nonprofit responders who interacted with them) was required to adopt a new way of engaging with one another. The Incident Command System changed how responders communicated with one another, the language they used, and how authority was deployed. My research relied primarily on in-depth coding of after-action reports, public testimony, and interviews.

This research revealed the tensions inherent in a mechanism of social coordination that featured the characteristics of hierarchical control but was used in the context of a decentralized network of responders from different types of organizations who are often unfamiliar with one another but must coordinate in a limited time-period. Despite the hierarchical characteristics prominent in formal presentations of the Incident Command System, the way in which response to large crises works in practice means that the network properties fundamentally mattered to operations. Case studies of different types of disasters, from forest fires, the outbreak of avian flu, hurricanes, and terrorist attacks, brought some common patterns to light. Larger networks of responders increased coordination costs and the potential for conflict even as they provided a wider array of resources. Network members frequently contest who directs the response, and trust based on prior working relationships matters a good deal to the success of crisis responses.

Let me turn to a deeper consideration of the tensions

---

One basic difference between networks for public goals and those found in private settings is that relationships in public networks are often a function of mandates put in place by policies, rather than a natural function of the goals and needs of the organizations.
between the formal framework of a new policy and the actual incentive structure that the policy places upon political relationships. New policies may serve to change the type of tasks that officials perform, the people with whom they work, how they are monitored, the goals to which they are held accountable, and the rewards and punishments they face. In crisis response, successive national response frameworks have directed the U.S. Department of Defense (DOD) to play a role in disaster response. This charge to engage in disaster response eroded a traditional and deeply held DOD cultural reluctance to accept tasks not directly related to war. A case study of the DOD role in Hurricane Katrina demonstrated how DOD leaders altered staffers’ pre-existing cultural beliefs to facilitate the embrace of disaster response. The initial slow DOD response arose because of self-imposed red-tape designed to limit engagement in crisis response, reflecting a cultural assumption of the need to maintain autonomy. DOD leaders altered the nature of the response by committing to another widely shared cultural assumption: a “can-do” approach to achieving difficult goals regardless of obstacles. As a result, the DOD responded on an unprecedented scale to a domestic disaster, but continued to act on its own terms, reflecting its preference for autonomy even in a networked environment.

One basic difference between networks for public goals and those found in private settings is that relationships in public networks are often a function of mandates put in place by policies, rather than a natural function of the goals and needs of the organizations. For example, national response frameworks require DOD to take directives from Federal Emergency Management Agency in any major disaster. In practice, tensions beset this arrangement. A basic network theory perspective is that coordination within a crisis response network depends a good deal on the capacity of network providers to develop norms of reciprocity and to maintain positive reputations with other network members. But network members also are concerned with the assessment of their reputation by non-network actors, such as their oversight committees or the general public. As actors in response networks perceive tension between protecting their reputation within and outside the network, they may choose one or the other. This potential tension is likely to be higher in public sector networks because network membership may be mandated (preventing voluntary exit as a solution to solving this tension) and where resources available to network members come primarily from non-network audiences (e.g., the appropriation committee). My case analysis pointed to this tension in the response to Hurricane Katrina, illustrating how the DOD dictated its approach by protecting its broader reputation even if doing so undermined reputation within the network, thereby weakening trust and making cooperation among network members less likely in future disasters.

The importance of policies to organizational variables reflects an assumption derived from some of my early work on public service motivation. Public employees score higher on a desire to serve others, which was largely assumed to reflect selection patterns into public service. My work also argued that organizational context affects the motivation to engage in public service. In some cases those factors could be negative: exposure to red tape and length of time in position (controlling for age) was associated with lower public service motivation. The overarching point is that as policies change tangible aspects of the organizational environment, so too will they change less tangible but no less fundamental aspects of organizational life, such as organizational culture and individual motivation.

Returning to the effects of performance-based reforms, the motivation to serve others may be weakened by reforms in ways that are actually detrimental to performance. Some of my research has used different measures of altruistic motivations in different datasets to offer evidence that public employees’ desire to serve others is associated with greater use of performance data to make decisions. In settings without strong extrinsic motivations, the use of performance data can be framed as a tool by which to help others.

By contrast, research derived from crowding-out theory suggests that introducing performance-based incentives can displace intrinsic and altruistic incentives, such as public service motivation. I have argued that this process may occur by changing the norms of existing public employees and by encouraging more publicly spirited employees to exit, and to discourage them from joining in the first place. If such patterns occur, and the public service ethos declines among public employees, it is also likely that performance will suffer, as a variety of meta-analyses suggest that intrinsic motivations are more important than extrinsic factors for the performance of complex tasks (that is to say, most public tasks), and that crowding out may result in a net lowering of performance. A second potential negative effect is that the moral judgment needed to resist the temptation for perverse use of performance data is also weakened as a public service ethos is crowded out. As performance regimes become perceived as unrealistic, unfair, and at odds with the underlying values of public service, even intrinsically motivated actors may find it easy to rationalize performance perversity.

Conclusion
I would like to conclude by returning to my central theme, which is that having a fruitful career working at the intersection of policy and administration is possible. Indeed, I might be willing to go further, to suggest that the space between policy and management is among the most interesting aspect of our field to study. That space allows us room to deal with the fundamental question of governance: the ability of the state to respond to challenges and the pressing problems of its peoples. Examining the interplay of relationships between the two fields allows us to develop more persuasive causal models and more relevant policy advice that speaks to broad governance issues, and across the divide between policy design and implementation. ✹
Assessing the Affordable Care Act: Current and Future Research Possibilities

By Barbara Wolfe and Maria Serakos

The Patient Protection and Affordable Care Act (ACA) has directly affected large numbers of Americans by expanding health-care coverage to millions of people, especially those with pre-existing conditions and those without the opportunity to acquire employment-based coverage, and by improving the quality of the health-care system. The sweeping national reform addresses increasing health insurance coverage, reducing inequality in access to health care, and controlling the very high and increasing costs of health care. In this paper, we outline the effects of the ACA in its first few years. We focus on the expansion of coverage, changes in access to care, and, to a lesser extent, health itself. We do not discuss changes to the health-care system itself.

Persons living at or below 133 percent of the federal poverty level in the 28 states and District of Columbia accepting Medicaid expansion (as of this writing) are now entitled to public health insurance coverage. Exchanges opened in every state, thereby improving the health insurance marketplace, and new mandates require coverage, though with some exceptions. The core provisions of the ACA on which we focus include expansion of a parent’s coverage to adult children up to age 26 and the constraint on insurers requiring coverage to those with chronic illnesses.

As a result of these provisions, millions of Americans have gained access to health insurance. In fact, by the spring of 2014, roughly 8 million to 10 million additional Americans gained coverage. Of all Americans with health insurance, nearly two-thirds are covered under private insurance, with just more than half of those receiving coverage through employer-sponsored insurance plans. Government insurance covers the remaining third of Americans, and this population is nearly equally divided between those receiving coverage through Medicaid or through Medicare.

Clearly, the ACA has increased health insurance coverage. These coverage gains may lead to a stream of effects, including changes in health-care access, health, and labor market outcomes. Some of these outcomes are relatively easy to identify, while others are more challenging given available data.

Important Research Areas

Expanding coverage, particularly among low-income individuals or those previously denied access to care, is at the core of the health-care reform legislation. Prior to the ACA, 47 million Americans, or 18 percent of the population, lacked health insurance coverage. As many of these individuals gain coverage, we would expect to see a series of downstream effects.

In addition to increasing coverage rates, the ACA is likely to affect access to and utilization of health services, trends in overall quality of health, and changes in labor market participation. Although medical cost and expenditure patterns and provider behavior may change as a result of the law, we concentrate on these trends centered on individuals. In Table 1, we indicate various sub-outcomes of interest within these four categories.

In accessing the ACA, it is important to evaluate trends in these outcomes both for the overall adult population and for subgroups of the population. Given the inequality in these outcomes prior to the ACA, comparison of how the legislation has differentially affected subgroups enables assessment of the extent to which the ACA has targeted low-income Americans.

The ACA likely affects low-income Americans in a unique way in terms of access to care, health, and labor market participation. For example, nearly two-thirds of the uninsured population prior to the ACA belonged to the “working poor,” corresponding to a family of four earning less than $44,000 per year. These individuals were less likely to have employer-sponsored insurance and often could not afford to purchase insurance on the private market. Under the ACA, these individuals may receive subsidies or gain coverage under Medicaid expansion. Low-income groups that reside in Medicaid expansion states are expected to be affected differently than those in states not expanding Medicaid. This is especially true for adults without dependent children who were not eligible for Medicaid in the vast majority of states prior to the implementation of the ACA Medicaid expansion in 2014.

The effects on young adults gaining coverage as a result of
the dependent mandate can be compared to the effects on low-income people. Obtaining coverage through a parent’s plan requires that one’s parent already has health insurance; therefore, collectively this young adult group is likely to be of a higher income level than the overall population gaining insurance as a result of the ACA. Additionally, the ACA may differentially affect the subgroup of elderly adults relative to other groups, as their health is typically of poorer quality and many are faced with labor market decisions around retirement age. In addition to studying how the reform has affected the broader adult population and subgroups of that population, it is important to assess additional targets of the ACA, namely individuals with pre-existing conditions and mental health needs.

Data Availability
The extent to which researchers can study the effect of the ACA on outcomes for various groups of the population rests on the availability of quality data. Three types of data are available to study the ACA: population surveys, administrative data, and provider data. Each has its own strengths and limitations in assessing coverage, health, and economic outcomes.

Since the ACA is a sweeping reform affecting the entire adult population, quality panel data are important for studying trends over time. While adults just above the age threshold provide a comparison group for studying the young adult provision and hence enable the use of cross-sectional data, there is no such comparison group when considering the adult population as a whole. As a result, developing and utilizing panel data that follow individuals over time is the only effective way of studying the effect of the ACA on the population at large and subgroups of the population.

To date, national population surveys, including the American Community Survey, Behavioral Risk Factor Surveillance System, Current Population Survey, Medical Expenditure Panel Survey, National Health Interview Survey, and Survey of Income and Program Participation, have been utilized most frequently in studying the ACA. However, these national data sources were not designed to study comprehensive health reform. While national surveys provide useful information on general health trends for the U.S. population, few are equipped to assess the intricacies of the coverage, access, health, and labor market outcomes of the ACA and to examine state differences in Medicaid expansion. The best of these for studying the broad

Table 1: Potential ACA Effects

<table>
<thead>
<tr>
<th>Coverage Rates</th>
<th>Access to Care</th>
<th>Health</th>
<th>Labor Market</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall population</td>
<td>Preventive care</td>
<td>Improvement in chronic conditions</td>
<td>Job mobility</td>
</tr>
<tr>
<td>Young adults</td>
<td>Availability of providers</td>
<td>Mental health</td>
<td>Self-employment</td>
</tr>
<tr>
<td>Individuals with preexisting conditions</td>
<td>Time to treatment</td>
<td>Physical health</td>
<td>Education-employment tradeoff</td>
</tr>
<tr>
<td>Subgroups: low-income, minorities</td>
<td>Patterns of utilization</td>
<td></td>
<td>Productivity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Retirement decisions</td>
</tr>
</tbody>
</table>

ACA will be the linked National Health Interview and Medical Expenditure Panel surveys dataset that covers the period just prior to 2014 into early 2015. However, this information will not be available until late 2016.

Despite these limitations, there are ongoing efforts to modify these data sources to reflect ACA-related outcomes. For example, the 2014 Current Population Survey Annual Social and Economic Supplement featured redesigned health insurance coverage questions corresponding to ACA coverage types. However, these data are not comparable to those gathered in previous years, limiting our ability to assess a variety of trends over time.

A few specialized surveys have been utilized or developed to gather information targeting ACA-specific outcomes, including the Commonwealth Fund Tracking Surveys, Gallup Healthways Well-Being Index, Health Reform Monitoring Survey, and RAND Health Reform Opinion Study. While these data sources provide the most in-depth information on the effects of the law, the timing of their development and data collection limits their ability to capture trends in the population prior to the law’s enactment. Their low response rates may create biases in the data as well.

Table 2 displays the capability of these national data sources and independent surveys to study coverage, access, health, and economic outcomes for those affected by ACA.

Although each of these data sources can be used to assess coverage outcomes, a number of them are appropriate for assessing subgroup differences. The American Community and Current Population surveys are among the most commonly used datasets for studying changes in coverage rates. The National Health Interview Survey provides particularly rich health insurance data, including length of time insured and information on intermittent coverage or churning throughout the year. The Urban Institute’s Health Reform Monitoring Survey has also provided a wealth of information on coverage trends, particularly for racial and socioeconomic subgroups.

Since the Health Reform Monitoring Survey was designed in 2013 after the enactment of the ACA, it also provides detailed information on how access to care and health have been directly affected by the ACA. However, the timing of this survey makes it difficult to assess pre- and post-reform trends, particularly for the young adult population.

Among the national population surveys, Behavioral Risk Factor Surveillance System, Medical Expenditure Panel Survey, National Health Interview Survey, and Survey of Income
and Program Participation all provide at least some information about access to care, including hospital stays and medical service utilization. However, while these data give a snapshot of overall access trends, they are insufficient for studying the nuanced access patterns of the ACA. Administrative and provider data, such as the Healthcare Cost and Utilization Project’s inpatient and emergency department databases, have instead been used to explore these outcomes.

By and large, the health data provided by the sources in Table 2 correspond to self-reported health measures; therefore, as in the case of studying access to care, we expect administrative and provider data to play an increasingly important role in studying the general health outcomes of the ACA.

Each of the national population surveys listed in Table 2, with the exception of the Behavioral Risk Factor Surveillance System, has potential to provide information on pertinent labor force outcomes. These may include type and duration of work, earnings, and educational attainment. Choosing which data to use is a function of the research aims, and as literature examining the labor force outcomes of the ACA is still emerging, there has been great variation in the data employed to date.

### Preliminary Research

The majority of social and economic research on the ACA to date has focused on young adults, as the dependent coverage provision was one of the first parts of the legislation to take effect in September 2010.

Because this provision affected only a small subset of the U.S. population, researchers can compare young adults ages 18 to 25 to adults slightly older (26 to 29, for example). Using this natural comparison group allows researchers to control for other changes such as the improving economy and changes in medical care in studying how the reform affected young adults in terms of various health, access, social, and economic outcomes.

Highlights of this research to date suggest that more than 2 million young adults have gained coverage as a result of the dependent coverage provision. Preliminary findings also suggest that these individuals have consequently sought more healthcare services, though early findings on routine and preventive service access have been inconsistent. There is some evidence of better self-reported health, including mental health, but no research targets specific conditions or detailed health outcomes. Preliminary labor market studies have produced mixed results, though there is evidence that suggests that females younger than 26 are less likely to marry as a result of the legislation.

### Future Research

Given the immediacy of the dependent coverage expansion and the relative ease with which researchers can study its effects, it is no surprise that studies of the young adult population dominate the ACA literature. Moving forward, we expect to see an emergence of research addressing the overall U.S. adult population that goes beyond increasing rates of coverage. Additionally, we can expect to learn much more about the other targets of the ACA and Medicaid expansion and their subsequent impacts on social and economic outcomes.

### Learning from Massachusetts’ Reform and Oregon’s Experiment

Research surrounding the Massachusetts health-care reform and Oregon’s health insurance experiment is relevant for studying the full effects of the ACA. In April 2006 Massachusetts enacted its third health reform legislation since 1998 (the previous two were an adoption of an employer mandate in 1998 and Medicaid expansion in 1996 and 1997). Key provisions

<table>
<thead>
<tr>
<th>Table 2: Select Data Sources for Studying ACA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>National Data Sources</strong></td>
</tr>
<tr>
<td>American Community Survey</td>
</tr>
<tr>
<td>Behavioral Risk Factor Surveillance System</td>
</tr>
<tr>
<td>Current Population Survey</td>
</tr>
<tr>
<td>Medical Expenditure Panel Survey</td>
</tr>
<tr>
<td>National Health Interview Survey</td>
</tr>
<tr>
<td>Survey of Income and Program Participation</td>
</tr>
<tr>
<td><strong>Independent Surveys</strong></td>
</tr>
<tr>
<td>Commonwealth Fund Tracking Surveys</td>
</tr>
<tr>
<td>Gallup Healthways Well-Being Index</td>
</tr>
<tr>
<td>Health Reform Monitoring Survey</td>
</tr>
<tr>
<td>RAND Health Reform Opinion Study</td>
</tr>
</tbody>
</table>
of the Massachusetts reform are similar to those of the ACA, namely subsidized insurance for low-income uninsured adults, an individual mandate requiring adults to have health insurance, employer responsibility to provide fair and reasonable insurance, Medicaid expansion, and insurance market changes.

An abundance of research focuses on coverage gains as a result of the Massachusetts legislation, with many studies stratifying their findings by subgroups, namely low-income and minority populations. Similar work assessing the effects of the ACA on coverage gains of subgroups with regard to work status and income level are likely to be forthcoming.

Several studies have focused on how the Massachusetts reform has affected health-care access rates. This research has informed us as to how this reform has led to changes in inpatient coverage rates prior to and following ACA implementation, individual emergency room use, coverage rates of individuals, admission rates for mental health diagnoses, access to care among children with special health-care needs, and changes in all-cause mortality and deaths amenable to health care.

Much of the Massachusetts literature emerged several years after the reform’s implementation. Given the time needed for health-care access and health outcomes to become evident, we expect similar lags in research surrounding the ACA. However, many of the Massachusetts studies took advantage of comparison groups in neighboring New England states in assessing access, health, and labor market outcomes. Without a similar natural comparison group for the overall ACA target population, it will be difficult to mimic the methodological techniques employed in the Massachusetts literature in assessing the ACA.

Studies on the social and economic impacts of the Oregon health insurance experiment may also predict ACA research. In 2008, more than 35,000 uninsured low-income adults in Oregon were selected by lottery to be given the chance to apply for Medicaid. This experiment functioned as a natural randomized control design, allowing researchers to study the effects of public health insurance coverage on health-care use, financial barriers to care, and overall health.

Individuals receiving coverage as a result of the Oregon lottery had significantly higher rates of health-care utilization, lower out-of-pocket medical expenditures, and less medical debt. There is some suggestion that these individuals also had better self-reported physical and mental health one year following the lottery. There was a significant increase in the likelihood of a diabetes diagnosis and decreased probability of positive depression screening among recipients, though there was no impact on the prevalence or diagnosis of hypertension or on related prescription medication use. Obtaining coverage through the lottery has not been associated with changes in employment, individual earnings, or earnings above the federal poverty level.

As with assessing the Massachusetts health-care reform, it is worth noting that these Oregon findings emerged several years after the lottery took place. Additionally, none of the data used in these Oregon studies came from national surveys; they were collected in person, via surveys, or through administrative data. Extrapolating these methodologies to the ACA underscores the importance of developing specialized datasets to assess access and utilization, health, and labor force outcomes.

Expanding Current Research

Research on the social and economic outcomes associated with the ACA is still in its early stages. Although there are an abundance of coverage estimates, few studies have specifically targeted important subgroups of the population, namely low-income individuals, individuals with pre-existing conditions, and minorities. Since these individuals were less likely to have coverage prior to the ACA, it is important to distinguish how their trends in coverage gains may or may not differ from the broader adult population.

Regarding access to care, we still know relatively little about overall patterns of health-care utilization. Although there appears to be an increase in office visits, emergency department and inpatient visits, and likelihood of having a primary care provider as a result of the ACA, these findings do not consider subgroup differences in terms of income or urban and rural populations, groups of individuals that may face additional barriers to receiving care. Furthermore, we are yet to learn whether the type of provider accessed for a health need or the time to treatment has changed following the ACA.

One of the key policy goals of the ACA was to improve quality of health; therefore, it is important to assess whether gaining coverage and increasing access to care has a positive effect on the overall health of adults. Most of the literature to date on health has been limited to self-reported data. While these findings have suggested a small improvement in overall health, they are not reliable for gauging changes in health among individuals with chronic conditions or changes in adverse health outcomes related to preventive care access, nor do they convey the more important dimension of longer term influences on health.

Understanding how the ACA may affect health outcomes will require time and additional data. Changes in public health take time to emerge, particularly among individuals with chronic conditions. As the Massachusetts reform and Oregon experiment suggest, we would expect to see more health research emerge in the coming years. Additionally, most national datasets are limited to self-reported health questions. Therefore, in studying health patterns, researchers will need to exploit administrative or survey data with sufficient information about health outcomes.

While there has been a recent emergence of labor market findings surrounding the ACA, the results to date have been inconsistent. More research is needed to corroborate or challenge the findings and isolate the underlying effects of the ACA on labor market outcomes. Employer-sponsored insurance has long been a significant component of many employment benefits packages; now that individuals and families can purchase coverage through the marketplace, one may expect to see a shift in employment patterns. These may include greater job mobility, changes in the tradeoff between seeking employment and additional education, particularly among young adults, and different rates of self-employment and part-time employment.
The availability of decent insurance outside of employment may also change retirement patterns. There may well be broader influences due to the ACA. For example, if the ACA improves the overall quality of health of individuals, there may be an increase in productivity, including more hours worked or a decrease in work missed due to illness. As with assessing health outcomes, labor market changes take time to become manifest, and assessing these trends requires specialized data sources.

One of the groups that may gain the most from the ACA is low-income individuals without children, especially in states that expanded Medicaid. Although some preliminary findings target expansion states, we still need a comprehensive research base surrounding this aspect of the ACA. Studying the difference between the experience of lower income individuals in Medicaid expansion states and those in states that did not expand will be important for understanding the effectiveness of the Medicaid component of the ACA.

Final Thoughts
The ACA has already significantly increased the health insurance coverage rates of American adults. We should be able to learn the extent to which the legislation has influenced access to care, and labor market outcomes in the next five or so years. The extent to which we will be able to capture fully these broader effects will depend to a large extent on availability of data, especially panel data which follow the same individuals in sufficiently large numbers to capture changes that occur due to the ACA. Since the timing of the ACA coincides with an improvement in the overall economy following the Great Recession, this represents a greater challenge to disentangling the effects of the ACA from behavioral and labor market shifts related to the economy. ◆
Reducing rates of smoking in the United States has become a stubborn problem. On one hand, a long view suggests success. In the 1950s and 1960s, before the Surgeon General’s warning on the health effects of cigarette use, nearly half of the adult population smoked. These rates were cut in half in a matter of a few decades, coincident with strong tobacco control policies. These public health policies have regularly been listed as some of the top ten public health successes in the last century in the United States. On the other hand, a shorter term view suggests limited progress. The prevalence of adult tobacco use has remained quite stable, at around 20 percent, for nearly 20 years, in the face of substantial increases in tobacco taxation and other control policies such as smoking bans in public places and some restaurants and bars. Indeed, some of the latest evidence suggests that the last round of tobacco tax hikes have generated no detectable reductions in cigarette use.

These trends are important in part because tobacco use is among the most common causes of mortality and morbidity in the United States and around the world. Tobacco use has been shown to harm nearly every organ in the human body and has been linked with more than 400,000 U.S. deaths each year, making it the leading preventable cause of death. In particular, smoking has been linked to pneumonia, many forms of cancer, lung disease, and risk of heart attack, stroke, and other illnesses.

An Economic Perspective

A fundamental concept in economics is the law of demand, which states that as the price of a good increases, the quantity demanded of the good will fall (keeping other factors unchanged). Following this logic, a primary focus of reducing smoking in the economics literature has been on increasing cigarette prices through taxation and regulation.

The current consensus is that cigarettes, like most consumables, do follow the law of demand. Most empirical studies estimate that the responsiveness in quantity demand is negative, but the range is quite large. So-called price elasticity estimates range from -0.5 to -2, meaning that a 1 percent increase in cigarette prices may be expected to reduce consumption by between 0.5 percent and 2 percent. (See Figure 1 for trends in taxes and tobacco use).

Our Study

The economic perspective on the determinants of cigarette consumption motivates our analysis. My coauthors and I aimed to extend the literature in two directions. First, we wanted to examine some of the potential mechanisms linking prices and quitting. That is, because smoking is addictive, any reduction in smoking that occurs when prices are raised is possibly a bit more difficult to achieve than when prices increase for, say, gasoline or milk. Indeed, the difficulty of quitting may be one explanation behind some estimates that show relatively small reductions in smoking when taxes are raised. Therefore, we wanted to take advantage of a relatively new set of pharmaceuticals that are meant to help people quit. Instead of examining the links between taxes and quitting behaviors, we asked whether we see increases in smoking cessation drug purchases when tobacco taxes are increased.

Second, we wanted to examine two types of “prices” when thinking about the price of continuing to smoke. The first price is the price of cigarettes themselves, shifted through taxation at the state level. The second price is the price of quitting—the price of smoking cessation drugs—and, in particular, whether these drugs are covered by state Medicaid programs for low income people.

To ask whether tobacco taxes and whether Medicaid coverage expansions were associated with the purchase of smoking cessation drugs, we worked with a company called IMS Health that collects data on more than 70 percent of all U.S.
prescriptions from 1999 to 2012—more than 50 million prescriptions for smoking cessation drugs. We then used regression analysis to link purchases of these products with state-level tobacco tax rates and state-level Medicaid coverage policies.

In brief, our preliminary findings did not suggest a link between taxes and smoking cessation drug purchases. We did, however, find evidence suggesting that states that provided Medicaid coverage for prescriptions had higher levels of smoking cessation drug prescriptions. One estimate suggests an increase of more than 20 percent.

We also examined whether states that created workplace smoking bans or smoking bans at restaurants/bars saw any changes in smoking cessation drug prescriptions. Indeed, we find large results for workplace bans (20 percent increase in prescriptions) but no impacts from restaurant/bar bans.

Why workplace tobacco bans induce greater demand for smoking-cessation drugs is not completely clear, but two contextual features may matter. First, such limits on smoking over the course of a full working day may enhance a smoker's desire to quit compared to bans on smoking in social gathering spots or to marginal increases in taxes. Second, smokers affected by workplace bans may hold private or public health insurance through their employers that facilitates lower cost access to otherwise costly interventions.

While our study is still in progress, we might speculate a bit on its potential implications. The clearest may be that the price of smoking cessation pharmaceuticals may be deterring quitting behaviors, even as taxation policies further inflate the costs of continuing to smoke. States that expanded Medicaid coverage to these pharmaceuticals saw very large increases in their use. If we assume that some proportion of this insurance coverage reduced smoking, then policies might consider subsidizing smoking cessation products in lieu of further taxation of cigarettes to pursue further reductions in smoking rates. At issue is that the budgetary and public health implications of these strategies may be working against one another. Some states have grown to rely on the large and stable flows of revenues into government coffers through increasing tobacco taxes on many people who are not going to quit smoking (economists would label this as a very efficient tax, where money is extracted from taxpayers with little/no behavior change). On the other hand, subsidizing tobacco cessation products is costly for states’ budgets in two ways—the money for the subsidy and the resulting losses in tobacco tax revenues. These are difficult questions to answer, but our study provides some new evidence of the potential impacts of shifts in policies. ✷
I presented research on how federal agencies handle statutes passed by Congress to see how those agencies act to translate those laws into regulations. It turns out Congress requires many rules that agencies never write. That means we might want to re-evaluate Congress’ delegation of power to those agencies and consider what is happening to Congress’ role of overseeing government operations.

Cary Coglianese, Edward B. Shils Professor of Law and professor of political science and director of the Penn Program on Regulation at the University of Pennsylvania Law School, and David Lewis, William R. Kenan Jr. Professor of Political Science at Vanderbilt University, were part of the closing roundtable that pushed us to look ahead and consider what our research should do next.

I can’t wait to get started.

Susan Yackee
Principal Effectiveness and School Accountability

By Robert H. Meyer and Bradley Carl

This paper provides an overview of a new evaluation measure for school principals. It is a value-added measure developed by the authors and has been adopted as part of the State of Wisconsin’s new educator effectiveness system. While the new metric posits that improving student academic achievement, as measured by the value the school adds, is a key objective for school principals, it recognizes that a principal may take several years to fully affect school performance.

Other states, including those supported by federal Race to the Top funding, generally make no distinction between school and principal performance, and in doing so may provide a strong disincentive for effective principals to accept appointments in schools with demonstrated low student growth. Wisconsin’s principal value-added measure addresses important distinctions between principal and school performance, resulting in a more accurate and fair measure of a principal’s impact upon schoolwide growth in academic achievement. Below, we define the concept of value-added and then introduce our measure of principal value-added.

Value-Added: An Overview

The term “value-added” refers to a class of student growth models and associated indicators used to measure the contributions of teachers, principals, and schools to improvement in student academic achievement over time. The models and indicators take into account differences across students in their starting points, as measured by prior achievement, and factors, such as poverty and other student characteristics, that often influence student academic growth. The objective is to facilitate valid and fair comparisons of student outcomes across schools, given that the schools may serve very different student populations. While value-added models utilize advanced statistical techniques and rely on high-quality data, they are conceptually fairly easy to understand.

In the first of three steps, a value-added model predicts levels of student achievement, given everything that is known about the students, including prior student achievement and student characteristics. Predicted achievement is simply the average of achievement for groups of students with the same characteristics. In the second step, actual achievement is compared to predicted achievement, yielding a measure of the degree to which students exceed or fall below predicted levels of achievement. In the third step, school or teacher value-added is calculated as the average of actual minus predicted student achievement for the groups of students in given schools and/or classrooms. The important point is that growth is measured in relation to schools or educators who serve similar students, as defined by prior achievement and student characteristics. Wisconsin’s school value-added model includes student achievement and these demographic factors: disability status, economic status (as measured by participation in the free or reduced-price lunch program), English language learner status, gender, and race/ethnicity. Figure 1 illustrates this definition of value-added.

Policy Context

Today, many states and school districts use value-added models as one component (within a “multiple measures” approach) to quantify the productivity of different actors (districts, schools, classrooms, and sometimes individual educators). Since student populations often vary widely across (and sometimes within) schools and classrooms in terms of prior achievement, family income, and other key non-school factors, a value-added approach helps create a more level playing field than is the case when attainment-based indicators (such as simple proficiency rates) are used to make high-stakes accountability decisions at the school and/or educator level. While simple attainment measures are very important in identifying standards of adequacy in student performance (typically in the form of college and career readiness) and for reinforcing the need to hold high expectations for the performance of all students, they almost invariably have a strong inverse relationship with student poverty (higher poverty rate = lower proficiency rate). This relationship raises important policy questions regarding the appropriateness of
simple attainment-based measures for teacher and principal evaluation (since schools and teachers typically don’t get to pick which students they serve). This relationship may also create unintended consequences (such as adding incentives for the most able educators to avoid working in the neediest schools).

Wisconsin and more than 40 other states have moved in recent years to create systems for school-level accountability (in the form of school and district “report cards”) and educator effectiveness (for evaluating the performance of teachers and principals, and providing them with targeted feedback for improvement). These systems often include value-added assessment methods.

This emphasis on educator effectiveness is driven by research demonstrating the impact of high-quality teaching and school leadership on student achievement, and has been codified into national and state policy through a combination of federal requirements and incentives. These include waivers granted to states by the U.S. Department of Education from provisions of the Elementary and Secondary Education Act (known in its most recent iteration as No Child Left Behind) and the Teacher Incentive Fund grant program at the federal level, and through state-level legislation (such as Wisconsin’s Act 166 of 2011).

Within Wisconsin’s new educator effectiveness system, which is being implemented statewide during the 2014-15 school year, teachers and principals will receive separate scores for professional practice and student outcomes rather than a combined overall effectiveness rating. The student outcome score will consist of three sub-scores: a small schoolwide reading value-added measure (for most elementary and middle school educators, designed to convey the importance of shared responsibility for literacy instruction), student learning objectives (based upon locally selected measures of student performance), and an individual value-added measure. The individual value-added score for teachers will not be applied until the 2017-18 school year at the earliest, due to the need to more fully develop state data systems that accurately attribute students to the appropriate teachers in terms of instructional “dosage.”

**Principal Value-Added: Key Features and Rationale**

For principals in most of Wisconsin’s elementary and middle schools, the individual value-added measure will be our principal value-added metric. In contrast to most other states for which information on growth measures can be discerned via state websites, Wisconsin is making a distinction within its educator effectiveness system between schoolwide growth and principal-level value-added, particularly for principals in their early years in new assignments. Most (and perhaps all) other states are, at least for now, treating schoolwide value-added/growth and principal value-added/growth as the same thing. At first glance, this melding may appear sensible, as the principal clearly serves as the leader of the school (and could, at first blush, be considered fully responsible for the entirety of a schoolwide growth measure).

Upon further consideration, however, situations arise in which this assumption (schoolwide growth = principal growth) is not consistent with what we know about how principals affect student learning. In contrast to teachers, whose impact on students is more direct and immediate (by virtue of being with their students each day), principals affect student achievement in very important, but somewhat less direct and immediate ways (see Figure 2). The “channels of influence” principals exert upon student achievement include making human capital decisions (hiring teachers and helping existing teachers develop through targeted professional development), establishing a school climate conducive to learning, and providing a safe learning environment, among other factors. Accordingly, one implication of constructing principal performance measures is that holding a new principal accountable immediately for schoolwide growth seems questionable, as schoolwide growth data in the early years of a principal’s tenure arguably reflect the effectiveness of a new principal’s predecessor(s) more than the new principal him/herself. Most research on this issue suggests that principals gradually assume full “ownership” of a school in three to five years, not immediately.

We have countered this assumption by building three distinct features into our measure of principal value-added:

- an “adjustment period” in the early years of a principal’s tenure that is designed to more accurately differentiate the impact of new principals from that of their predecessors
- accounting for the prior value-added history of a school
- incorporation of principal value-added from prior assignments (in the case of principals who change schools)
Adjustment Period

The first feature is an “adjustment period” through which principals assume full ownership over schoolwide value-added over approximately five years (incorporating four consecutive year-to-year growth cycles based on annual state tests). This element reflects research suggesting a three- to five-year “phase-in” period for principal leadership, after which time the principal is assumed to have full ownership of schoolwide value-added growth.

Our examination of data from Wisconsin public schools suggests that in any given year, approximately 20-25 percent of Wisconsin public schools have a different principal compared to the prior year, and about half have a principal whose tenure in his/her school is three years or less. As Table 1 shows, this finding suggests that far from being an isolated, infrequent challenge, the issue of principal turnover and tenure cannot be ignored when constructing meaningful principal-level growth measures, as a significant portion of Wisconsin principals arguably lack the tenure as leaders of their school to be held fully and immediately accountable for historical schoolwide growth data.

Accounting for Prior School Value-Added.

The second feature is an adjustment for the level of school value-added that existed prior to a principal’s assignment to a new school. For a principal who inherits a school with low prior growth, a reasonable expectation would be realistic and gradual improvement over time, rather than immediate, short-term turnaround (see Figure 3). Conversely, for a principal who inherits a school with high prior growth (see Figure 4), we would hope and expect this pattern to be maintained, since further improvement in value-added is essentially not possible at the highest levels of the growth scale. In short, the principal value-added metric has the common sense feature of classifying the following as excellent performance: increases in school value-added over time from a low starting point and sustainability of high value-added over time (even with no increase). Other important features of principal value-added include:

- The adjustment for prior school value-added is strongest during the early years of a principal’s assignment to a new school, which reflects the fact that it is most difficult during the early years of a principal’s assignment to disentangle the contributions of the current and past principals.
- Estimates of principal value-added exhibit the greatest statistical error during the first years of a principal’s new assignment (a consequence of the need to control strongly for prior school value-added).

- The range of principal value-added estimates is smallest during the early years of a principal’s assignment to a new school, a reflection that the estimates from these initial years are inherently less precise. This feature limits the possibility that the performance of the prior principal will unduly affect the estimated performance of the current principal.
- Over time, the adjustment for prior school value-added declines so that after approximately five years (the adjustment period discussed earlier), principal value-added is approximately equal to school value-added, and the principal assumes full responsibility for school performance.

Principal Value-Added from Prior Assignments

Principals who are new to their schools, of course, are not all equal with respect to the availability of prior performance data, which raises a third important point that is incorporated into our new principal value-added measure: the availability of prior data. Some principals are new entirely to the principal role in Wisconsin public schools, having been promoted from an assistant or associate principal role, a classroom teaching role, or from a school for which prior schoolwide growth data are not readily available (because, for example, the principal moved into Wisconsin from another state or previously served as the leader of a private school in Wisconsin or elsewhere). For other new principals who move from one Wisconsin public school with existing growth data to another, however, it is possible—and arguably highly desirable—to incorporate this information from prior assignments into the principal value-added measure.
Choosing not to do so would disadvantage a principal with a track record of producing high school-level growth and could unintentionally reward a principal with a history of low school-wide growth by allowing him/her to repeatedly start over with a “clean slate” in a new school.

The channels of influence through which principals affect student learning (as shown in Figure 2) can vary considerably across schools. In some schools, for example, the principal has considerable (or even complete) control over human capital decisions (such as which teachers to hire), while in others teacher hiring is done at a central (district) level by the human resources department. Similarly, the degree of control that principals exert over curriculum, budgets, and the identification of priorities for their schools (through school improvement plans, for example) also varies, from principals who have substantial authority in these matters to those with limited authority. While it is not currently possible to accurately quantify the degree to which principals influence these and other key channels of influence upon student learning, improving our ability to account for these factors through improved data collection is a key priority for research and development on principal-level growth models such as Wisconsin’s principal value-added measure.

Managing the Transition to New State Tests

As Wisconsin and other states transition to new state assessments in the 2014-15 school year, several challenges to the construction of a principal value-added measure must be addressed. By itself, the transition to a new test scale (from the Wisconsin Knowledge and Concepts Examination [WKCE] to the Smarter Balanced examination, and perhaps other assessments after that) is largely inconsequential for value-added models as long as the pre-test (the WKCE) provides good statistical predictions of the post-test in the transition year. One condition that could weaken the prediction between pre-test and post-test would be if substantial changes were made to the content standards to which state assessments are aligned; the status of content standards in Wisconsin going forward is somewhat uncertain as of this writing.

A more substantive challenge for calculating value-added in Wisconsin involves the movement of the state’s testing window from the traditional fall WKCE window to a spring window under Smarter Balanced. Wisconsin is considering several options for incorporating the change in testing windows into a multi-year principal value-added measure, including a simple re-weighting of the multi-year average to reflect the fact that growth in the transition year (from the Fall 2013 WKCE to the Spring 2015 Smarter Balanced) represents approximately 18 months of instructional time as opposed to the 12 months covered by prior growth cycles that are included in the multi-year average (Fall 2011-Fall 2012 and Fall 2012-Fall 2013).

A third example of a challenge associated with test transition involves principal mobility during the transition year. As described above, some 20-25 percent of Wisconsin public schools have a new principal in any given year based on historical data, which will presumably look similar for the 2013-14 and 2014-15 school years (which serve as the test transition years in Wisconsin). For principals who moved between schools with value-added data following the 2013-14 school year (that is, from School A in 2013-14 to School B in 2014-15), the test transition growth cycle (Fall 2013-Spring 2015) will involve two academic years and two schools. Options for handling these situations are being considered by a working group of Wisconsin practitioners and the state Department of Public Instruction.
Conclusion

As part of Wisconsin's new educator effectiveness system for teacher and principal evaluation, the University of Wisconsin–Madison’s Value-Added Research Center has partnered with the Department of Public Instruction to create the principal value-added measure, which will be applied as one component of a “multiple measures” approach that includes professional practice and student academic outcomes. In contrast to states that use schoolwide growth or value-added as a measure of principal effectiveness (regardless of principal tenure and the school’s prior academic growth history), the Wisconsin measure incorporates three distinct features. The first is an “adjustment period” in the early years of a principal’s tenure (in which principal value-added results are reported on a truncated scale) that is designed to more accurately differentiate the impact of new principals from their predecessors. The second accounts for the value-added history of a school to allow for the setting of ambitious, yet realistic, trajectories for improved student academic growth (in the case of principals who take over low-growth schools) or the maintenance of high growth (for principals who take over high-growth schools). The third feature is the incorporation of principal value-added from a principal’s prior assignments, where available. Failure to incorporate these features into a principal-level performance measure seems likely to inaccurately attribute the effects of a new principal to his/her predecessors and could create unintended incentives for principals to avoid serving poorly performing schools.

The La Follette School of Public Affairs

Congratulates
University of Wisconsin–Madison

Chancellor
Rebecca Blank

for winning the
2015 Daniel Patrick Moynihan Prize
from the American Academy of Political and Social Science