Performance Metrics as Formal Structures and through the Lens of Social Mechanisms: When Do They Work and How Do They Influence?

JEANNETTE A. COLOYVAS
Northwestern University

Our current educational environment is subject to persistent calls for accountability, evidence-based practice, and data use for improvement, which largely take the form of performance metrics (PMs). This rapid proliferation of PMs has profoundly influenced the ways in which scholars and practitioners think about their own practices and the larger missions in which they are embedded. This article draws on research in organizational sociology and higher education to propose a framework and research agenda for analyzing the relationship between PMs and practice. I argue that PMs need to be understood as a distinctive form of data whose consequences in use relate to their role as formalized systems of abstractions and the social mechanisms through which they exert change in practice.

I. Introduction

Performance metrics (PMs) are among the most salient links between data use and organizational improvement in education policy and management. Scholars and pundits are well aware of the importance of Adequate Yearly Progress standards mandated under No Child Left Behind (NCLB), the National Research Council’s (NRC) rankings of US doctorate programs, and SAT scores for college admissions. And the preponderance of such metrics extends beyond the public sector. The US News and World Report (USN) college rankings have become the gold standard for applicants comparing campuses, and many institutions have initiated evidence-based feedback systems modeled after those developed in hospitals and firms (Pfeffer and Sutton 2006). Per-
Performance metrics have many forms and sources, from locally generated measures used at work to professional standards proliferated by experts to top-down delegated rankings that benchmark individuals and organizations. Performance metrics are ubiquitous, as a lever in policy interventions and as a source of contention.

We use PMs every day, often with little reflection and understanding of the substance behind them and how they operate in deep ways to structure organizational life. Conventional approaches treat PMs as a facticity: if we measure what matters, and make those measurements public, people will adjust their behavior and practice will improve. This article provides an alternative, by not only challenging this assumption but also moving the agenda forward on the relationship between PMs and practice. I argue that PMs become a part of a formal organizational structure that shapes how people think about what they do and the patterned ways they use PMs. And yet, our tools for analyzing these relationships and examining their effects in practice are limited. In this article I articulate a framework and research agenda for analyzing the relationship between PMs and practice. I argue that PMs need to be understood as a distinctive form of data whose consequences in use relate to their role as formalized systems of abstractions and the processes through which their qualities influence change in practice.

A Distinctive Class of Data

Performance metrics are distinguishable from other data forms through their character, technical properties, and use. In their character, PMs are quantifiable classification systems that purport to connect practice, goals, and outcomes. Performance metrics are abstractions that delineate categories, compress information, and represent the essential substance of what individuals or organizations aim to accomplish (Stinchcombe 2001). As for technical properties, PMs enable commensurability because they gauge differences of degree of accomplishment within these categories that serve as points of reference among people, places, and things (Alonso and Starr 1987; Espeland 2001). In use, PMs’ defining factor is that they enable individuals or organizations to compare or benchmark achievements (Davis et al. 2010).

In education, test scores that assess student aptitude or school quality ex-
emply PMs, as do letter report card grades and statewide teacher/administrator evaluations. Higher education has numerous examples, including USN, Consumers Digest’s “Top 100 College Values,” or the government-sponsored Integrated Postsecondary Data System (IPEDS). Some data, such as the National Study of Postsecondary Faculty, by design preclude their use as PMs because they anonymize identities or are constructed to allow only aggregate analysis. Similarly, certain data are not used as PMs, despite being technically conducive to benchmarking, such as a regression analysis using IPEDS but reporting only coefficients. Some advocate for the use of existing data as PMs that, with minimal changes, could serve as more substantive alternatives to media-based rankings such as USN. The National Survey of Student Engagement, designed to measure conditions that promote learning, is one such example, but it faces the obstacle of institutions’ reluctance to making the results publicly available (Zemsky 2008).

Performance Metrics in Practice

Attention to PMs is critical to our understanding of data use and organizational improvement. As an intervention, PMs pervade education’s history, especially our contemporary accountability environment (Honig and Coburn 2008; Kézar 2010; Power 1997; Strathern 1997, 2000; Williams et al. 2005). From scientific management to total quality management to NCLB and beyond, many assessment movements have advanced metrics that define and gauge the state of our institutions and have become a cornerstone of practice.

Mobilizing PMs’ virtues without their vices, however, is our greatest challenge. Proponents in education view PMs as a technology that makes more information available, comprehensible, and useful through its ability to package and summarize information. Performance metrics facilitate comparison to oneself and others, motivate organizations, and infuse discipline into practice (Sauder and Espeland 2009). In higher education, PMs are touted as promoting a “healthy” competition, by directing attention and resources toward areas where institutions of higher education (IHEs) can build a comparative advantage. University of Pennsylvania’s rise in the professional school rankings or Northeastern University’s dramatic downsizing and move into the top 100 research universities are but a few examples in which rankings provided goals and focus for strategic planning (Zemsky 2008).

Critics, however, argue that these same properties obscure more than expose efficacious practices. As a form of feedback, PMs direct attention and resources to some areas while silencing others. As a means of making comparisons, PMs transform social structures, often undermining the identity, mission, and, ironically, accomplishments of organizations. As a discipline, PMs motivate gaming
strategies such as “teaching to the test” (Shulman 2004). College admissions provide the most notable examples of tactics used to boost selectivity scores by increasing applicants irrespective of fit or treating incomplete applications as “actionable,” with no intention of evaluating them (Speyer 2004; Sumner 2004, 71).

Regardless of one’s stance, PMs are ubiquitous—often invisible and taken-for-granted in ways that are treated as real, obvious, or without alternatives (Alonso and Starr 1987; Espeland 2001; Martins 2005; McDonough et al. 1998). An admissions process without GPAs or college decisions without rankings has become inconceivable, even among those numerous institutions that are not selective. Yet PMs are more than just a technology that becomes natural to many sectors (Davis et al. 2010). They are social, in that they require discipline, resources, and organization to subsist in practice (Espeland and Stevens 1998). For example, SAT achievement draws families into complicated and expensive routines outside of the classroom. Much of the SAT-related discipline, resources, and organization is directed toward practices that enhance scores and their use in very distant ways from the learning the exam is supposed to reflect. Performance metrics offer a “patina of objectivity” and “reorder the social institutions they are purported to assess” (Colyvas and Powell 2009, 79; Elsbach and Kramer 1996; Espeland and Sauder 2007).

Consider the example of US universities and their increasing role in formal technology transfer—defined as the movement of academic discoveries to industry through patents and licenses for commercial development (Haeussler and Colyvas 2011). Formal university technology transfer gained prominence as a means of maintaining international competitiveness and moving new knowledge to firms to convert it into socially and economically important products (Kaufman Foundation 2007; Powell et al. 2007). Bolstered by legislative and legal changes, technology transfer achievement is now expected for universities (Mowery et al. 2004; Slaughter and Rhoades 2004). The Association of University Technology Managers (AUTM) publishes an annual technology transfer survey across measures, notably the number of patents and start-ups produced, licenses to firms, and income generated (AUTM 2009). These rankings have become the “USN” of research universities, providing both useful and crude measures of university economic performance (Colyvas and Powell 2009; Colyvas et al., forthcoming).

Although some technology transfer metrics resonate—such as revenues, patents, licenses, and start-ups generated—others are silenced—such as citations to academic research or the movement of graduates into firms. Indeed, the mobility of graduates as carriers of tacit knowledge is arguably technology transfer’s most important form, and students’ cutting-edge scientific training is a vital economic contribution (Stephan 2008; Vallas and Kleinman 2008). Furthermore, the broader effects of increased university commerce are over-
looked, such as how technology transfer reshapes careers, the distribution of talent, or research agendas (Colyvas and Powell 2009). The director of a prominent university technology licensing office aptly notes: “If you measure success only by the amount of royalties and fees you bring in, then your licensing practices will reflect that. If you measure success in terms of social impact or awareness and . . . count things such as gifts, research collaborations, global impact, and boost to your reputation, it changes your orientation” (Check 2006, 413).

Nevertheless, universities persist in trying to move up the technology transfer ranking ladder by signaling achievement through existing measures, thereby devoting less effort to refining these measures in ways that would reflect more substantive contributions to the economy (Powell and Owen-Smith 2002). The Chronicle of Higher Education exemplifies this point: “three more universities join the $10 million club in annual revenue from licensing inventions” (Blumenstyk 2007). Technology transfer success has become a new fault line for universities, reordering academic institutions on the basis of these indicators, and creating a new currency for competition and prestige (Owen-Smith 2003; Stuart and Ding 2006).

A Framework and Research Agenda

Given these rewards and challenges, how can PMs meaningfully link practice, goals, and outcomes with minimal damage to settings in which they are introduced? My goal is not to document the range of uses or PM studies. Rather, I draw on organizational and management sociology to select theories, analyses, and examples toward a framework that can be used to investigate problems related to PM data use and organizational change—where increasing our understanding will point the way forward in policy, practice, and research. My objective is to facilitate the analysis of the relationship between design features of PM systems and their effects in practice.

Most research emphasizes how organizations measure up to PMs, treating the design and substance of these systems as given and impervious to change (Coburn and Turner 2012, in this issue). Recent work, notably by practitioners, has demonstrated the perverse consequences of PMs on everyday practices and work environments (Adams 2004; Speyer 2004; Sumner 2004). Present calls seek practical ways to ameliorate these challenges and manage performance more generally (Thacker 2004). I aim to shift the emphasis away from either the influence of metrics on organizational performance or the enumeration of dysfunctional consequences toward a focus on the relationship between the qualities of PMs themselves and everyday work. Here I provide a framework for analyzing PMs that attend to both design and practice—that
Performance Metrics as Formal Structures

is, to capture (1) efficacy and failure of PM systems in terms of their character and technical properties; and (2) direct and indirect effects of their use on practices and settings in which they are introduced. One can then examine variations across defining dimensions of PMs—character, technical properties, and use—to analyze the interrelated problem of the quality of PMs as interventions and their effects in practice.

I focus on higher education, defined as academic organizations, formally authorized to award postsecondary degrees—an area where many research advances on this topic originate, and that provides an apt setting to analyze the relationship between PMs and practice, especially through the lens put forth in this article (Brint 2002; Stevens et al. 2008). Among our most successful US sectors, higher education has a long legacy of persistence and change, generating numerous examples of PMs’ introduction, failure, and maintenance (Sauder 2008). Institutions of higher education transcend social domains, including the family, nation-state, law, and various professions, making higher education suitable for observing the reach and effects of PMs (Stevens et al. 2008). Constituting multiple organizational forms (i.e., public, private, and nonprofit), higher education provides a strategic setting for distinguishing effects across domains (Gumport and Snydman 2006). As a conduit for mobility and social reproduction, researchers can examine IHEs through the paradoxes and conflicts of designing interventions directed at improvement (Kerr 2001; Rosenbaum et al. 2006). Finally, PMs in higher education have multiple temporal and structural origins, such as the faculty-produced Gourman Report (beginning in 1967), Arizona State University’s Center for Measuring University Performance rankings of “Top American Research Universities” (2000), and for-profit New Mobility Magazine’s episodic publication of “Top Disability Friendly Universities” (1998). More recent efforts have included global perspectives, such as the “Academic Ranking of World Universities” from China’s Shanghai Jiao Tong University or the University of Western Australia’s pilot Teaching Quality Indicators Project. Most analyses focus on four-year IHEs, but the dynamics, framework, and agenda put forth in this article are applicable to others and education more broadly.

The following sections develop two approaches to analyzing PMs, which make up the conceptual framework put forth in this article: as a form of formalization that generates abstractions and governs behavior (Sec. II), and through the lens of mechanisms that prompt organizations and individuals to react (Sec. III). The former permits examination of the quality of PM systems, and the latter enables us to examine the multiple ways in which they exert effects in practice. Together these lenses comprise a formality-mechanisms framework that can characterize key constructs for analyzing features of the PM systems themselves and those causal processes that alter the practices and setting in which they are introduced. I delineate the application of each lens...
to PMs and their analytic benefit. Section IV concludes with a synthesis and research agenda.

II. Performance Metrics as Formalized Abstractions

If the quality of PMs is critical to practice, how can we systematically gauge their efficacy? The first step requires understanding PMs through the lens of formal structures in organizations—that is, as officially sanctioned arrangements that delineate goals and processes for conducting work (Scott 2001). The second step is applying a framework for analyzing distinctive features of formal structures to the context of PMs. I draw on insights from organizational sociology to illustrate the implications of seeing PMs as formal structures, followed by a framework developed by Stinchcombe (2001) to analyze formal structures. We can thus emphasize the link between formal structures and practice and also identify features that differentiate PM systems and evaluate their quality along measurable dimensions. Doing so enables us to understand how PMs might be better developed to complement practices, and in turn how practices may be addressed through systematically analyzing PMs.

Understanding PMs through the Lens of Formal Structures

Treating PMs as formal structures shifts our focus to their ability to delineate goals and processes in practice. For example, college rankings, learning assessments, or AUTM’s technology transfer survey purport to provide approximations that can focus practices that would lead to better college quality, teaching, and economic contributions. Each PM is officially sanctioned, often at multiple levels, through professional associations, public media, and local management. Yet each plays various roles in practical application and has been subject to numerous critiques. By seeing PMs as an element of formal structure, we can thus treat PMs as a system of abstractions that approximates the meaningful qualities of an objective, such as intentions about performance, and shapes the activities of those subjected to that objective, such as higher education administrators and faculty (Stinchcombe 2001).

Acknowledging the role of formality in practice is important because “building formalities that work” has been one of the central challenges in organizations, certainly in higher education (Stinchcombe 2001, 2): “Much of the time formalities are designed to be rituals, and not to work. Very often they are perverted from their intended purpose to serve corrupt purposes for which they were not designed.” We know that a formal structure can be developed for one reason and mobilized for others, or it may take on a life of its own.
Consider the SAT, conceived as a means of identifying talent to diversify society’s elite. Many caution that it has become a mechanism for social stratification, because youth with resources and privilege are better able to maximize gains through access to specialized information, directed preparation, and testing strategies (Lemann 1999). Once a formal structure takes hold, it can be difficult to undo and can outlive its original functions. Sometimes formalities themselves can be built for purposes that we later deem inappropriate, such as state laws that once authorized racial segregation in IHEs. Formality has been the cornerstone of educational policy and improvement directives, yet with mixed results.

Meyer and Rowan (1977, 1978) demonstrate this point in their study of myth and ceremony in organizations that adopt structures more to maintain legitimacy than for substance. Educational institutions buffer their practices from the content of these structures—what scholars term decoupling—through a range of strategies, notably by symbolic gestures of compliance such as task forces, job titles, and public declarations (Rowan and Miskel 1999). Indeed, much higher education research has identified numerous ways in which practices are decoupled from structures, from teaching (Kezar 2010) to management fads (Birnbaum 2000) to admissions (Delucchi 2000).

Evidence suggests, however, that PMs may be distinctive in their practice-related influence. Recent work emphasizes “coupling” rather than “decoupling” processes, largely through analyses of PMs’ effects as higher education benchmarks: “Rankings, as commensurate, relative, and broadly circulating measures, are more difficult to buffer than . . . other types of institutional pressures” (Sauder and Espeland 2009, 65). This point makes PMs especially consequential as formal structures, since organizations cannot deploy the same buffering strategies to maintain their legitimacy and attend to technical exigencies of work. One might argue that PMs can facilitate decoupling in other ways, since conformity to a faulty proxy may actually enable an IHE to signal compliance without any substantive change. Nevertheless, PMs’ relationship to “coupling” processes makes the problem of devising efficacious interventions less about successful implementation, or adherence to PMs, than about how PMs, given their design and characteristics, operate in practice.

Viewed as formalized abstractions, the critical attribute for PMs to operate effectively requires substantively preserving the essential elements of the objectives that give rise to them (Stinchcombe 2001, 3). For example, for college rankings to work well, we would need them to reflect meaningful institutional differences in ways that enable student choices about which college to attend (Bowman and Bastedo 2009). When PMs lack alignment with the actions to which they are directed—for example, if the same rankings were used to allocate government research funding—this perspective predicts they are more likely to fail (Stinchcombe 2001). Such instances can give rise to informal
patterns that can at the same time patch up a faulty abstraction and also be at odds with it.

Stevens (2007, 49) demonstrates this tension between patches and contradictions in college admissions. The author recounts a conversation with an administrator about the annual prospective visitors’ data:

He went on, “Another thing that happened with [institutional research] was about the percentage of students in the entering class in the top 10 percent of their [high school] class.” This was a very important figure, because USN uses it as part of its institutional ranking scheme. “When [one of our staff] calculated this figure, she calculated it from 90 percent. But [institutional research] uses 89.5 percent. Which they say we can do because that’s how the estimate works. If we use 90 percent, the percentage of students in the top 10 percent of their classes is 53; if we use 89.5 percent, that number changes to 55.” . . . Simply by using the lowest possible number that could legitimately be rounded to 90, the college nudged a consequential figure a little further in the right direction.

Thus, seemingly minor statistical differences generate consequences in the college’s rankings and image. Yet the goal of promoting the college can undermine other forms of improvement. The statistic that prompts the organization to learn from previous performance can be obscured when generating numbers for comparison to others. The lower figure might motivate the school to learn, whereas the higher one confers legitimacy and resources (DiMaggio and Powell 1983, 1991; Kezar 2005). Scant research considers the trade-off between internal organizational learning and increasing external standings (Ewell 2009).

Informal patches take multiple forms, notably in situations where PMs as interventions fail or warrant revision (Stinchcombe 2001). Doctoral programs often require qualifying exams intended to reflect student command of a body of knowledge. Whether a test question reflects the substance that students must master is often disputed. Failure of one student on an examination could prompt a committee to retest the student or take up remediation. Others may choose to address the examination process and question the exam’s efficacy: “One sort of informality . . . is making a patch on the system to get the right results for the case at hand, but not modifying the system of abstractions for future formality” (Stinchcombe 2001, 7). Often programs develop policies or routines to counteract situations where the exam’s inadequacy generates failure, which can include multiple assessment mechanisms, committees that customize remediation, or reexamination procedures. From this perspective, such elaborations can function more to institutionalize the PMs’ shortcomings than to improve the PMs’ use.

Little is known about the conditions that give rise to patches built around
PMs rather than revisions of the PMs themselves. The variation across settings is notable. Evidence suggests that once a measurement system is used publicly across organizations, it is very difficult to undo (Colyvas and Jonsson 2011; Espeland 2001). Universities are ranked nationally on their number of patents, start-ups, or licensing revenues, despite the obvious deficiencies in capturing the actual movement of new knowledge through graduates or contributions to high-priority regions or industries. Colleges are now subject to metrics of selectivity relative to similar others across the country rather than a socio-demographic they traditionally reach (Adams 2004; Espeland and Sauder 2007; Speyer 2004).

Once such metrics become institutionalized, patches are more likely to involve the elaboration of practices rather than substantive revision of existing structures (Berger and Luckman 1967). Colyvas and Powell (2006) show how technology transfer practices were built up with the establishment of roles to accommodate the new objective of patenting and licensing, such as “technology transfer associate” job titles and accounting routines. An insight from firm-based performance analyses is that organizations tend to tack on more measures rather than tackle the complicated task of devising more appropriate ones—that would provide a simplified yet comprehensive look at the disparate elements of the organization such as finance, operations, and core constituents (Kaplan and Norton 1992). The conditions that promote revisions to the metrics themselves over routines that build up or work around PMs would provide important insight into the problem of removing outdated performance structures or making existing ones more effective.

Framework

Treating PMs as formalized abstractions enables us to draw on Stinchcombe’s (2001) framework for analyzing formal structures by identifying features that differentiate different kinds of abstraction systems and converting those features into measurable variables (summarized in table 1).1

According to Stinchcombe, for abstractions such as PMs to “work”—that is, to reflect the substance of practice given particular goals—they require three features: cognitive adequacy, communicability, and a means of improvement. Cognitive adequacy refers to the ability of an abstraction to capture enough of the important qualities of the substance it is supposed to represent to guide action. “The main argument is that, if formality is to govern . . . better than informality could, it first must have mechanisms that make the abstraction in it an effective representation of the problems and solutions in an area of social life” (Stinchcombe 2001, 10). Gauging abstractions’ cognitive adequacy requires assessing four core attributes: “their accuracy in description, their sufficiency
<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Variables</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive adequacy</td>
<td>Accuracy</td>
<td>Accuracy relates to the ability to describe or portray the substance of the abstraction.</td>
</tr>
<tr>
<td></td>
<td>Economy</td>
<td>Economy reflects the cognitive aspects of compressing and providing enough information (free of noise) to generate the appropriate “diagnosis” of a situation and “remedy” for it.</td>
</tr>
<tr>
<td></td>
<td>Sufficiency</td>
<td>Sufficiency represents the quality of covering enough information so that “relevant action can be taken.”</td>
</tr>
<tr>
<td></td>
<td>Scope</td>
<td>Scope insures that the abstraction captures most situations to which it would apply.</td>
</tr>
<tr>
<td>Communicability</td>
<td>Transmissibility</td>
<td>Transmissibility is the ability of the abstraction to aptly transfer to its audience, including features of its plausibility and comprehensibility.</td>
</tr>
<tr>
<td></td>
<td>Transparency</td>
<td>Transparency needs to be impervious to corruption by interests and ideologies.</td>
</tr>
<tr>
<td></td>
<td>Durability</td>
<td>Durability includes the ability to withstand degradation by noise and confusion.</td>
</tr>
<tr>
<td>Improvement trajectory</td>
<td>Correctability</td>
<td>Correctability often takes the form of mechanisms and routines that track changes in the world.</td>
</tr>
<tr>
<td></td>
<td>Robustness</td>
<td>Robustness reflects both the flexibility of the abstraction and the potential to increase its scope.</td>
</tr>
<tr>
<td></td>
<td>Algorithms and reasoning devices</td>
<td>Algorithms and reasoning devices support the ability to relate abstractions to each other.</td>
</tr>
</tbody>
</table>


to grasp all that is necessary to govern the action, their *cognitive economy* that makes them easy to work with, and their *scope* so that they can govern a wide variety of situations*” (Stinchcombe 2001, 18). Thus, test scores must reflect the problem of student learning and the solution of having been taught (cognitive adequacy). Test score categories must appropriately describe what they measure (accuracy), such as reading, math, and writing ability on SATs, and must reflect enough of the substance of that category representing teaching
and learning (sufficiency) without omitting essential elements of it. Test score categories must also be easy to understand without the inclusion of unnecessary or redundant information (cognitive economy), such as whether a measure of multiplication skills accompanied GRE scores that were already covered in other aspects of the quantitative category. They must also accurately reflect the situations that they are meant to cover (scope), such as graduate school readiness in the GREs and not, for example, civic engagement or emotional maturity.

However, ample research raises questions about test score accuracy on grounds of validity, measurement error, consistency in results, and overall bias by “producing systematically incorrect estimates of the performance that accurately represents achievement” (Koretz 2008, 13). Understanding PMs in terms of cognitive adequacy provides the added benefit of disentangling the sources and degree of accuracy, sufficiency, economy, and scope above and beyond the biases that particular metrics produce. In Stinchcombe’s theory, abstractions that are low in cognitive adequacy are not serious, meaning that they are likely to be more ritualistic than efficacious.

In addition, an effective abstraction needs communicability: it must be transmissible, transparent, and durable. This dimension addresses the social and discursive aspect of abstractions that many argue is consequential in analyzing public benchmarks (Espeland and Stevens 1998). Receivers of PMs require the ability to access, understand, and believe what they represent (transmissibility). Us News and World Report’s Web-based rankings have a user-friendly matrix interface, providing easy access to information and the ability to search and customize criteria. Such features heighten transmissibility. In contrast, many state and government sources of educational data are difficult to navigate or require onerous licensing and security arrangements, decreasing transmissibility. Performance metrics must also resist corruption and cooptation by ideologies and interests, largely supported by the open disclosure of methods and modifications (transparency). If USN concealed the disparate ways IHEs were ranked, or allowed input from advertisers or politicians, rankings would look very different. Yet transmissibility has its challenges. The peer review method in USN has generated gaming tactics whereby deans deliberately downgrade competitors and first-rate universities, superficially ranking themselves at the top (Bastedo and Bowman 2010). Performance metrics also require the quality of resisting degradation through noise or confusion (durability). A testing category that ranks a student’s ability to perform on standardized exams would provide extraneous information and detract from the substance of what is supposed to be learned and measured as an outcome. Imagine how prospective students, families, or IHEs would interpret and respond if USN included such a metric.

Finally, an abstraction must have improvability through correction, robustness,
and reasoning devices. “One way to check whether [formality] is likely to be fraud and ritual is to ask whether there is an effective set of devices to improve the abstractions in the system” (Stinchcombe 2001, 36). The improvability of an abstraction depends on some means of tracking changes in the world (correction), “so that it will be better in the future than it is now” (37). Peer review provides a way of updating the baseline for assessing academic findings, helping us determine a contribution based on current research. Efficacious abstractions must also remain flexible enough to increase their scope and withstand future changes in the world (robustness). Stinchcombe (2001) emphasizes how US case law provides a degree of ambiguity for categories, which can be updated through precedent. As a result, legal categories can absorb changing norms and technologies, as in privacy rulings in the context of the proliferation of the World Wide Web. An abstraction’s improvability requires some means of facilitating relationships among formal structures (reasoning devices), often as algorithms, such as tenure-line faculty promotion metrics that relate research, teaching, and service in quantifiable forms. Allocating percentages of effort among these categories in sponsored research reports and counts of activities (e.g., papers published, classes taught, and committees served) reflect procedures for calculating faculty performance. Some abstractions have no algorithms, in the sense that one cannot calculate with them, such as flow charts that reflect tracks for employee advance or boilerplate contracts used for conflict of interest reports or licensing technologies.

As of yet, no studies apply this framework to education or PMs. Most scholarship in these areas either focuses on PMs’ effects on outcomes, presumably ones they were designed to influence, or highlights their deleterious consequences. Those that do address PMs’ attributes largely focus on testing and emphasize technical features, notably psychometric and statistical characteristics such as measurement and sampling errors, reliability, validity, and forms of bias (Koretz 2008). But, this approach relies on an incomplete lens of statistical measures, rendering most practice-related failures to consumers’ intentions or understandings, with less regard for the social and organizational aspects of PMs’ use. Stinchcombe’s framework provides a promising foundation for filling this gap and analyzing the relationship between PMs’ form and quality in substantively linking practice, goals, and outcomes.

III. Performance Metrics through the Lens of Mechanisms

If the use of PMs is consequential to organizational improvement, how can we analyze PMs influence on the practices and settings in which they are introduced? Although we can treat the design of PMs as abstractions that link substance to action, in practice such links can operate in numerous ways with
Performance Metrics as Formal Structures

profound effects, some of which may be out of alignment with PMs’ substance or use as an intervention. Here I delineate a lens for analyzing the processes through which PMs generate particular effects on practice by utilizing the construct of social mechanisms developed in analytic sociology (Hedstrom and Ylikoski 2010). A mechanisms-based approach affords the ability to link the substance of what PMs represent (intervention) to what organizations and individuals do (practice). I next review a mechanism-based framework developed in the sociology of higher education for examining the relationship between PMs and practice to detail its range of application and suggest areas for expansion. Unlike Stinchcombe’s framework for analyzing formalized abstractions, mechanism-based inquiry into the effects of PMs on practice has been taken up in recent years in higher education. I emphasize, however, that the promise for utilizing this processual construct has numerous applications that have yet to be examined.

Understanding PMs Use through the Lens of Mechanisms

A mechanism is a type of process, identified through its relationship to an outcome or effect (Hedstrom and Ylikoski 2010, 50): “The basic idea of a mechanism-based explanation is . . . that proper explanations should detail . . . the causal process through which the outcome to be explained was brought about.” The simplest conception of mechanisms is the relationship between two variables. For example, Lutfey and Freese (2005) examine the link between lower socioeconomic status (SES) and adverse health outcomes, whereby SES has been rendered a “fundamental cause” because of its persistent significance in statistical analyses. The authors demonstrate multiple pathways that associate SES and health, notably the organizational factors embedded in the health system and realities of day-to-day life for lower-income people, as opposed to cognitive and motivational claims. Therefore if variables are the nuts and bolts of an explanation, mechanisms can be understood as the cogs and wheels (Davis and Marquis 2005; Hedstrom and Swedberg 1998).

If we think of PMs and practice as two such variables, then we can understand mechanisms as those processes that explain the relationship between these two variables, specifically how PMs influence practice and, in turn, how practice might (or might not) influence PMs. The challenge, as noted in debates about mechanism-based scholarship, is that “a mechanism is always a mechanism for something,” and analyses can become mired in detailed descriptions and storytelling (Hedstrom and Ylikoski 2010). One solution to this problem is to take up mechanisms more conceptually, whereby the focus and investigation of specific processes is analyzed in terms of observable effects.

For example, mechanisms can represent “the structure or process S by which
some event or variable $I$ leads to or causes a change in the state of variable or event $O$ (Gross 2009, 360; Hedstrom and Swedberg 1998). Performance metrics can be understood as the structure or process ($S$) by which their introduction ($I$) leads to or causes a change in the setting in which they are introduced ($O$). The challenge for research, policy, and practice is whether that change is aligned with the goals for which certain PMs are constructed.

For higher education, understanding mechanisms is especially important, because the plurality of roles that IHEs take suggests that PMs in one domain can have effects in others, potentially with profound social consequences. Stevens et al. (2008) delineate four higher education roles that have concrete distributive effects in society: (1) sieves that regulate social mobility; (2) incubators that develop and shape the competence of social actors; (3) temples that legitimate and authorize knowledge; and (4) hubs that connect multiple institutions. All four link IHEs and social outcomes that we can identify empirically.

Yet most PM research focuses on specific settings, such as law schools, or on particular goals, such as improving links to the economy, teaching, or learning. Less is understood about the mechanisms through which the introduction of PMs in one higher education function influences practices and structures in others.

For example, many express concern about the effects of technology transfer policies—a more recent practice as a “hub” that connects multiple institutions on the humanities and social sciences—a foundational “temple” function in the legitimation and authorization of knowledge, because the potential for biomedical blockbusters directs resources toward the sciences (Mowery et al. 2004; Slaughter and Leslie 1997). The ability to measure, benchmark, and transmit PMs in technology transfer reinforces a broader claim about the practical applications of university research and more local attempts at securing revenue (Powell et al. 2007). Patenting research in the humanities or social sciences is difficult, with even fewer opportunities for licenses. The fact that there is no analogue for technology transfer performance for these disciplines (e.g., one cannot patent an ethnographic claim) reorders statuses within universities, often shifting practices to applied areas (Owen-Smith 2003). The transposition of PMs from one domain into another can have profound effects, particularly when organizations span multiple institutions (Hwang and Powell 2005). We can thus understand transposition as a mechanism that is consequential to the effects of PMs on practices, such as changes in research programs to more applied, commercializeable areas.

Social mechanisms can be understood and analyzed at multiple levels, depending on the specificity required for an analysis. For example, whereas the transposition of PMs from one domain to another may be a more macro mechanism, we can further focus the lens of mechanisms and capture the
specific processes through which certain PMs, such as USN rankings, influence certain effects, such as admissions decisions. One can easily observe the numerous ways in which the introduction of PMs to improve one area of practice can have effects in others. At their core, mechanisms are organizational and remind us that the role, meaning, and use of PMs, however technologically sophisticated or taken-for-granted, are organizational as well (Espeland and Stevens 1998). Much research is needed to unpack the relationship between the transposition of PMs and the effects on practices across higher education’s multiple roles and activities, as well as the specific processes and effects related to features of PM systems themselves.

Framework

This approach facilitates a multilevel framework for characterizing the structures or processes ($S$) by which the use of PMs ($I$) exerts effects across educational roles and practices ($O$). One example of such work comes from the research programs of Espeland and Sauder (2007), who have studied the effects of USN rankings on law schools. According to Espeland and Sauder, public rankings can be understood in terms of reactivity—the ways in which individual behavior changes when being measured, observed, or evaluated: “Because people are reflexive beings who continually monitor and interpret the world and adjust their actions accordingly, measures are reactive . . . [and] elicit responses from people who intervene in the objects they measure” (2). We can use this conception to address different forms and patterns of reactivity and link them to specific effects, especially those that are downstream and often overlooked when advancing PMs as interventions.

Table 2 summarizes Espeland and Sauder’s (2007) framework, distinguishing between mechanisms of reactivity and patterns of effects associated with them. One mechanism relates to self-fulfilling prophecies—that false interpretations of situations generate behaviors that make them true (Merton 1973). The authors expand this definition beyond spurious interpretations to how an expectation becomes amplified or confirmed after being rendered as real or true: “processes by which reactions to social measures confirm the expectations or predictions that are embedded in measures or which increase the validity of the measure by encouraging behavior that conforms to it” (11).

According to this perspective, self-fulfilling prophecies alter practices in four ways. First, they shape audiences’ perceptions, which can make statistically small differences have much larger consequences. Stevens’s (2007) prior-mentioned admissions research demonstrates the sensitivity of rounding up or down in updating statistics, which has been observed across IHEs (Monks and Ehrenberg 1999; Sauder and Lancaster 2006).
Second, rankings produce biases, as prior survey responses influence subsequent ones: “past judgments about a school’s reputation, embodied in rank, may influence current evaluations, which reproduce and strengthen these judgments,” even when they are subconscious (Espeland and Sauder 2007, 14). This effect is not limited to USNews alone, where many professional school metrics rely on evaluations from other deans. Many challenge the validity of NRC rankings because of their reliance on peer reports for their subjective quality, yet this perspective predicts that the consequences on future updates are amplified through prior survey response bias.

Third, public measures influence the allocation of resources, as PMs provide feedback for internal budgeting decisions. For example, a primary concern in developing university technology transfer entailed curtailing potential stratification across departments by incurring licensing revenues (Colyvas 2007). Despite these claims, universities have made concerted efforts to seed technology transfer and reward returns with new investments such as capital expenditures and faculty billets.

Fourth, rankings “encourage schools to become more like what rankings measure, which reinforces the validity of the measure” (Espeland and Sauder 2007, 15). In law schools, “rankings impose a standardized, universal definition . . . which creates incentives . . . to conform to that definition” (15). Higher education has numerous examples, from the effects of SATs in undergraduate education to GREs, LSATs, and recent efforts to introduce learning assessments (Ewell 2009; Kezar 2010). Espeland and Sauder (2007) provide an apt

### TABLE 2

**Espeland and Sauder’s Reactivity**

<table>
<thead>
<tr>
<th>Mechanisms of reactivity:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-fulfilling prophesy:</td>
</tr>
<tr>
<td>Effects on external audiences</td>
</tr>
<tr>
<td>Influence on prior rankings on survey responses</td>
</tr>
<tr>
<td>Use to make organizational decisions</td>
</tr>
<tr>
<td>Realization of assumptions embedded in rankings</td>
</tr>
<tr>
<td>Commensuration:</td>
</tr>
<tr>
<td>Uniting and distinguishing relations</td>
</tr>
<tr>
<td>Eliciting reflection on meanings of numbers</td>
</tr>
<tr>
<td>Patterns of the effects of reactivity:</td>
</tr>
<tr>
<td>Maximization</td>
</tr>
<tr>
<td>Redefinition</td>
</tr>
<tr>
<td>Manipulation</td>
</tr>
</tbody>
</table>

example of such conformity through a law school dean of a self-defined “access university”:

A student can have a very high GPA but a very low LSAT score. That student is not going to help us in the rankings. But let’s assume for the moment that that student is a minority—that’s the kind of student we like to admit and to give an opportunity to, even though that’s going to adversely impact our rankings, and even though that student is not going to pass the bar exam the first time around . . . that is our mission. We are going to have to change our mission and change our thinking if we are going to take that student’s place and give it to a student with a higher LSAT score, if we are going to be driven by the rankings. Our problem, and our challenge and our opportunity, is that we are poised at the top of a tier. So it is almost impossible to resist the temptation to move into that [higher] tier [and] then that becomes a self-perpetuating situation. (15)

Another reactivity mechanism reflects the cognitive aspects of quantification, labeled *commensuration*—the transformation of qualitative differences in form to quantitative differences in degree (Espeland 2001). The fact that PMs come as categories and numbers generates powerful conduits for reactivity: “Where self-fulfilling prophecies operate primarily by changing behavior in relation to altered expectations, commensuration works mainly by transforming cognition; it changes the locus and form of attention, both creating and obscuring relations among entities” (Espeland and Sauder 2007, 16). According to Espeland and Sauder, these processes take place in three observable ways.

First, commensuration’s ability to reduce, simplify, and integrate information renders some data more relevant than others, often at the expense of more difficult or qualitatively rich information. This mechanism defines a shared metric across institutions, making information appear more authoritative. Through commensuration, PMs become portable and decontextualized from the circumstances that generated them, taking on other roles and purposes. Ewell (2009) raises this point with state uses of ACTs and other cognitive examinations for institutional assessments, designed for very different purposes.

Second, commensuration defines organizational relations and dimensions for evaluation. This process of uniting and distinguishing renders distinctiveness a matter of comparison and magnitude. Many Collegiate Learning Assessment programs (CLAs) have tied funding to performance, particularly to graduation rate incentives (Ewell 2009). This approach puts IHEs in direct comparison and competition for scarce funds, around precise measures, which stratifies them. As in previous examples, IHEs that were able to maintain distinction by serving unmet needs and achieving regional goals are redefined.
in national standings across criteria that are detached from their missions of representation and local service.

Third, commensuration prompts reflection on the meaning of the numbers on which schools are evaluated. “USN requires administrators to prepare numbers which it turns into new numbers called rankings. Because these new numbers are important and dissected by many parties, the relationship between what the numbers are supposed to represent about their target and what members believe is real about their law schools elicits explanation” (Espeland and Sauder 2007, 21). Administrators may choose to eschew or embrace rankings, yet we have little understanding of the conditions that shape the allocation of scrutiny at the level of practice.

Although commensuration and self-fulfilling prophecies take multiple forms, Espeland and Sauder argue that in practice reactivity’s effects can be observed in patterned ways—through maximization, redefinition, and manipulation. Institutions of higher education change their method of allocating resources (maximization)—largely to improve their standing. Routines and scripts emerge through PMs (redefinition), such as job descriptions, admissions policies, and applicant evaluation methods. Reactivity also occurs as gaming strategies that enhance appearances (manipulation), often disconnected from the substance behind what rankings seek to measure. Institutions of higher education often reclassify students to remove scores from what they report, sometimes at the expense of characteristics that may be more meaningful to their variegated local and regional missions, such as civic engagement, diversity, or reaching community and business needs (Speyer 2004; Sumner 2004). From this perspective, quantitative differences of degree repress qualitative differences among students that support the school’s self-identity.

These insights apply to other educational domains, from admissions to university technology transfer to transformations in the broader higher education field (Colyvas and Powell 2009; Sauder 2008; Stevens 2007). Espeland and Sauder’s framework provides a first step, but the opportunities for more mechanism-based work are numerous. For example, although most analyses have highlighted deleterious effects of PMs, this lens need not apply solely to negative consequences. Its benefits reside in the ability to highlight effects that are multilevel, secondary, or beyond the cognitive reach of practitioners and policy makers making real-time choices. According to Espeland and Sauder (2007), much research is needed on the multiple forms that reactivity takes and the patterned ways in which they manifest and resonate in different settings. We also need research that distinguishes when and how these mechanisms and patterns of effects might protect versus transform the range of functions across higher education.
IV. Conclusion and Research Agenda

Our current educational environment is subject to persistent calls for accountability, evidence-based practice, and data use for improvement, which largely take the form of PMs. This rapid proliferation of PMs has profoundly influenced the ways in which scholars and practitioners think about their own practices and the larger missions in which they are embedded. A framework for analyzing PMs can help disambiguate those aspects of PM systems’ character, technical properties, and uses that promote or defy PMs’ ability to meaningfully link practices, goals, and outcomes.

Higher education research provides a wealth of insight into PMs. We have numerous accounts of the perverse tactics that engagement with PMs produces, through competition, resources, and even high-stakes coercion. We also learned of successes as IHEs parlay PMs’ discipline and focus into advance and specialization. Discourse often partitions this puzzle into discrete debates: proponents versus critics of PMs’ use, positive or negative effects, or intended versus unintended consequences. Such framings, however, overlook the complexity that practitioners aptly explain: not all PMs are created equally; certain PMs are difficult to abandon without substantial consequences; and PMs potentially help as well as hinder. Furthermore, any intervention involving PMs is only as good as the individuals and organizations using them. Where does a research agenda need to go from here?

One alternative is to focus on the organizational aspects of formal structures and those mechanisms that reinforce, hinder, or reside in the shadows of PMs’ use—the formality-mechanisms framework described here provides a first step (fig. 1). A theoretically consistent framework that combines the design features that make PMs work and the mechanisms through which PMs influence enables us to capture PMs’ efficacy and failure as well as their primary and secondary effects in practice. I described PMs (1) as formal structures comprising abstractions that govern behavior to assess their character and technical properties, and (2) through the lens of mechanisms that prompt organizations and individuals to react to assess PMs’ use. The formality lens focuses attention on measurable attributes that can promote or undermine the linkage to practices, insofar as they align and aptly represent goals and outcomes. The mechanisms lens provides a means of identifying and categorizing those processes through which such practices, goals, and outcomes may change through PMs’ use. Combined, we can address how PMs’ features and effects reflect abstract elements and processes that can reside at multiple levels, often beyond the cognitive understanding of those who participate in them (Alonso and Starr 1987).

This formality-mechanisms approach also highlights gaps in higher education literature: (1) how PMs generate practices that are synergistic or con-
tradictory to organizational goals; (2) the conditions that give rise to patches in practice versus revisions of the PMs themselves; (3) the transposition of PMs’ use or effects across different IHE functions; and (4) the adoption and scrutiny of some PMs over others. This framework provides a preliminary set of tools but also provides opportunities for refinement and elaboration.

Application of Formality Analysis to Higher Education

Treating PMs as formalized abstractions permits us to more broadly examine PM’s qualities and features that make them more or less efficacious in practice. Formal structures entail coordination and control, especially in settings where work is embedded in complex, interdependent networks of people, tools, and tasks (Meyer and Rowan 1977; Poole 2004). Stinchcombe (2001) reminds us that although so much attention goes to processes that circumvent and impinge upon formal ones, scholarship tends to treat formality as epiphenomenal. Yet, IHEs have a long legacy of formal interventions—for example, efforts at professionalizing and infusing managerialism into practice has transformed universities (Altbach et al. 2005; Gumport 2000). Our challenge is to under-
Performance Metrics as Formal Structures

stand “how formality works to accurately and usefully reflect substance . . . [Then we can see] . . . that at its best it does better with the substance than we could do informally” (Stinchcombe 2001, 2). Since Stinchcombe’s formality/abstraction perspective has yet to be deployed in PMs and higher education, systematic application, operationalization, and development of variables provide a necessary step. How does formality/abstractions analysis apply to higher education? What are the challenges of and solutions to its analytic application? What is the relationship between failures in PMs’ qualities (cognitive adequacy, communicability, improvability) and practice?

Since so much extant research into the design of PMs has developed in the realm of testing, a research program in higher education might do well with a specific comparison to existing approaches: How does formality/abstractions analysis compare to other modes of evaluation, notably psychometrics? How might PMs operate effectively in terms of psychometric criteria of measurement and sampling errors, reliability, validity, and bias yet exhibit vulnerabilities in terms of cognitive adequacy, communicability, and a means of improvement? Under what conditions might we see certain relationships between psychometric criteria and formalized abstraction criteria?

Extension of Mechanisms Analysis

Mechanisms-based approaches help expand our range of PMs’ effects, analyze the processes through which such effects occur, and identify places where interventions may more substantively help PMs work. Research from sociology and higher education demonstrate that multiple, often nested mechanisms contribute to PMs’ effects. One extension to higher education research would be to delineate different mechanism levels, types, and relationships among them. For example, we can distinguish those identified thus far as (1) perceptual mechanisms, such as commensuration and self-fulfilling prophesy, reflecting cognitive and social psychological processes, and (2) practice mechanisms, reflecting what individuals do. Further analysis might observe additional perceptual or practice mechanisms along with patterns of their effects.

Mechanisms in organizational sociology are meant to reflect processes that explain empirical relationships more broadly and “see beyond the surface-level description of phenomena” (Anderson et al. 2006, 103). Thus, higher education mechanisms need not be confined to perception or reactive forms of practice. We can investigate, for example, mechanisms that reinforce or hinder the use of new PMs—such as existing IHE data management routines that are more closely aligned with some PMs and not others. Or, if concerned about patterns of effects, we can investigate mechanisms of abandonment of existing arrangements as a result of PMs’ introduction. Numerous IHEs report
dropping SATs, yet to varying degrees, and rationales, including the effects of USN (http://www.fairtest.org/university/optional; Thacker 2004). We could take up an analysis that distinguishes among those processes that influence different higher education levels—what some scholars term situational, action-formation, and transformational mechanisms (Anderson et al. 2006; Coleman 1990). For example, how do PMs effect practice (1) through external environmental conditions (situational); (2) when operating solely within IHEs (action formation), or (3) in ways that influence the broader field (transformational)?

Comparisons across PM Systems and Mechanisms

Although we have accumulated a stock of domain-specific and goal-specific studies, the analysis of multiple PM systems or their comparative quality and effects remains unexamined. Systematic comparisons across PMs, mechanisms, and patterns of effects are a necessary step, as comparative approaches help identify associations often lost in single-case analyses. The formality-mechanisms framework could be deployed in three ways, as single-case or comparative-case designs: (1) as a comparative analysis of the quality of multiple PM systems; (2) as a comparative analysis of reactivity mechanisms prompted by one or more PM systems; or (3) as a two-stage formality-mechanisms analysis of one or more PM systems across IHEs. What relative features make certain PMs more or less efficacious in their application and everyday use? What are the comparative circumstances in practice that give rise to PMs’ revision as opposed to patches that ameliorate their shortcomings? How do different PMs reach into IHEs multiple roles?

One of the great advantages of this formality-mechanisms approach rests in its potential to capture the recursive relationship between features of formal structures and practice (fig. 2). Formal structures by design or development play an important role in shaping practice, and yet the execution of everyday work can reshape structures (Powell and Colyvas 2008). From the formality insights, we can conceptualize the relationship between attributes of PMs and organizational change through measurable variables. Through the mechanisms insights, we can conceptualize the multiple processes that link these two sets of variables—PM attributes and practice. Both perspectives concern themselves with the relationship between formal structure and practice, as well as patterned effects that PMs might produce. And both provide useful conceptual distinctions that help to disambiguate consequential characteristics of PMs design or use.

Yet each place their explanatory emphases on different aspects of the relationship. Whereas the formalized abstractions lens details constructs and
variables associated with characteristics of the PMs, the mechanisms lens emphasizes details and constructs associated with the processes through which PMs exert effects. The formalized abstraction lens identifies characteristics that PMs must have in order to meaningfully link practices, goals, and outcomes; the mechanisms lens identifies the processes through which PMs alter behavior, often in ways that extend beyond the focus of the PMs. In combination, the formality-mechanisms framework suggests that some mechanisms produced by features of the PMs, might interfere with other important characteristics of the PMs. For example, Espeland and Sauder’s (2007) work provides numerous examples of mechanisms of self-fulfilling prophesies that undermine the improvement trajectory of the formalized abstractions. If the metric is made to be true, then there is little impetus to update it. How do PM characteristics as formalized abstractions shape or activate distinctive mechanisms of reactivity? How do mechanisms of reactivity reinforce or impede features of cognitive adequacy, communicability, or improvability?
Sources of PMs as Public, Private, Nonprofit Forms

Most research examines externally mandated or publicly legitimated PMs, with little reference to their institutional form. Performance metrics are not limited to national policies or media-based public benchmarks. We must investigate PMs’ qualities and effects depending on their organizational origins: whether developed from public, private, or nonprofit organizations; from outside the higher education field or homegrown inside IHEs. In other sectors such as health care, analysis has demonstrated significant differences in behavior and performance among different institutional forms (Weisbrod 1998).

Also, much higher education policy and scholarship emphasizes accountability, notably from state and government forces (Ewell 2009; Fitz-Gibbon 1996; Kezar 2010). Resource pressures have empowered donors and foundations to dictate PMs’ use and form, from logic models to implementation plans to balanced scorecards, whereby measurable outcomes are contingent for support. Note the sharp contrast to other forms of accountability, such as National Science Foundation (NSF) grants that mandate intellectual merit and impact, yet do not require measures of such merit and impact. Instead, the NSF requests descriptive statements, such as how knowledge dissemination will occur. Distinguishing PMs’ sources can help explain persistent design successes and failures for specific higher education needs and also help direct policy making and implementation. How do PMs from public, private, or nonprofit sources differ in their dimensions of cognitive adequacy, communicability, and improvement trajectory? How do these dimensions differ when publicly mandated, homegrown, or contingent on resources? To what extent do the PMs’ qualities versus the stakes associated with them independently influence reactivity mechanisms or patterns of their effects?

Performance Metrics’ Higher- and Lower-Level Institutionalization

In higher education, PMs are supported at multiple levels, but does the degree of institutionalization matter—that is, the extent to which practices and structures become integrated into systems of rules, modes of organizing and beliefs (Greif 2006; Scott 2001)? For structures to be institutionalized, they must be self-reinforcing and self-reproducing, which can take place at micro and macro levels—that is, at the level of an organizational routine or at the level of an entire field (Colyvas and Jonsson 2011; Jepperson 1991).

Treating PMs as formal structures enables us to distinguish between their level of institutionalization—that is, whether PMs are public and institutionalized in the broader higher education field but only weakly integrated locally,
or local to particular IHEs and scantly integrated field-wide. Practices can emerge and become institutionalized at the local level to support public PMs, or local ones can be developed to buffer broader, more ambiguous policy mandates. We may observe USN’s effects inside colleges, yet acknowledge that these forces emanate from macro-level environments. In such instances, we may be interested in mechanisms through which PMs are pulled down from the institutional environment and local practices are built up to support them (Powell and Colyvas 2008).

Macro and micro-level structures shape the range of actions permissible for individuals and organizations in diverse ways (Scott 2001). Stinchcombe’s theory suggests that public PMs can prompt more elaboration of practices and patches to ameliorate their shortcomings, whereas local measures may be less rationalized, more idiosyncratic, and therefore more conducive to customization. Yet local measures may have a higher likelihood of linking to the substance of what people do, because top-down metrics and efforts toward standardization can undermine the cognitive adequacy of PMs, make them more vulnerable to corruption, and distance them from any means of improvement. An alternative hypothesis could be that public measures, especially those attached to formal or reputational sanctions, more likely induce reactions that speak to their design. Yet they can be more difficult to control and limit their reach into other organizational functions. Empirical analyses testing such hypotheses would enable better PM design, implementation, and practical use.

We also need more generalizable understandings of how PMs are reinforced and reproduced within work settings. Consider the practices identified so far—maximization, redefinition, and manipulation—that can boost IHEs’ public face, sometimes without substantive change or at odds with organizational learning. We also observed patches that build around abstractions, in effect supporting their shortcomings, rather than refining PMs. How do PMs differ in formality and reactivity based on ways they are implemented into practice? How does formality and reactivity differ under different levels and means of reinforcement and reproduction—for example, whether mandated publicly, embedded in technologies, marshaled into IHEs by consultants, or initiated locally by entrepreneurs?

Ultimately, our goals are best aligned when in service to advancing theory and informing policy and practice. This formality-mechanisms framework is an important start. Education has a broad range of PMs that often go overlooked or have fallen out of use. But we must not assume that their silencing is due to lack of efficacy or utility. Some practitioners have taken steps in developing better PMs and education has numerous examples of the initiatives of professions and institutions. My hope is that this article can help to focus attention on a set of empirical questions and tools to investigate them that will deepen our understanding of the relationship between educational data
use and organizational change, and ultimately lead to better metrics themselves.

Notes

I thank Andrea Beuschel, Cynthia Coburn, Adrianna Kezar, Erica Turner, Walter Powell, and five anonymous reviewers for invaluable comments on this article. I also thank Burt Weisbrod and colleagues participating in the working group on performance metrics in the nonprofit sector at Northwestern’s Institute for Policy Research. Support for this research came from the Spencer Foundation and the National Science Foundation (SES-0849036). This article also benefited greatly from the network of scholars and intellectual resources of the W. T. Grant Foundation’s program on the use of research evidence in policy and practice.

1. We might view these components as discrete—as a matter of their presence or absence—but they are rather meant as a matter of degree to be treated as variables.

References


Performance Metrics as Formal Structures

Stephan, Paula. 2008. “Tracking the Placement of Students as a Measure of Technology Transfer.” In Measuring the Social Value of Innovation: A Link in the University Technology Transfer and Entrepreneurship Equation, ed. Gary D. Libecap. Advances in the Study of
Colyvas


