1. Assessing performance and public value in the hollow state

Kelly LeRoux and Nathaniel S. Wright

More than 25 years have elapsed since Milward and Provan (1993) introduced the term “hollow state” into the lexicon of public administration. These authors describe the hollow state as a metaphor for “the increasing use of third parties, often nonprofits, to deliver social services and act in the name of the state” (Milward and Provan, 2000, p. 359). While the process of government’s hollowing out began decades earlier with unprecedented increases in United States (US) federal spending for health and human services in the 1960s, it accelerated throughout the 1980s and 1990s as policymakers embraced new public management (NPM) reforms emphasizing devolution, decentralization, and privatization. Indeed, the number of nonprofits in the US roughly doubled every decade between 1960 and 2000, a trend that is largely attributed to the increased availability of government funding to support these organizations (Salamon, 1995).

Thus, by the time Milward and Provan coined the term “hollow state” to describe the diminished role of government in delivering public services and the ubiquitous reliance on nonprofits for this purpose, the phenomenon had already been well established.1 As the discipline’s leading scholars began to take an academic interest in this reality and contemplate its consequences for public management, similar terms and concepts made their way into the discourse of the field, including governance (Frederickson, 1999; Lynn et al., 2001), hybrid governance (Skelcher, 2005), and collaborative public management (Agranoff and McGuire, 2004). While there are nuances among these terms and the policy tools they emphasize, they share a consistent understanding about the role of government and public organizations in contemporary public administration and management. Across these terms, government is recognized as having political and legal authority to determine policies and their methods of implementation, and to appropriate resources for this implementation. Across these terms, bureaucracies are understood to labor in the pursuit of the public interest, which often requires engaging other governments, citizens, and private organizations (for-profit and nonprofit) in policy debates and in solutions for service delivery. Across these terms, it is recognized that
as public organizations navigate complex policy problems, complex political dynamics, and complex resource deployment decisions, they will strategically form interdependencies with third party actors while steering the policy course and retaining oversight of those actors. While multi-sector collaboration is implicit in each of these terms, nonprofits figure most prominently in the concept of the hollow state. Milward and Provan’s arguments were premised upon the notion that government’s capacity to directly deliver services, mainly health and human services, had been largely supplanted by nonprofits. In this book, we examine the consequences of this reliance on nonprofits and to what extent it has delivered positive (or negative) outcomes, including public value for citizens. Therefore, “hollow state” is the conceptual frame from which we orient our analysis, and the term we engage throughout this book.

A number of different policy tools and financing instruments enable the state to function in this hollowed out condition. One of the most widely used mechanisms for delivering publicly funded services through nongovernmental organizations are fixed price and fee-for-service contracts. In the US, roughly 40 percent of the federal government’s discretionary spending goes to contracts for goods and services, with the federal government spending more than $550 billion on these contracts in 2018, an increase of more than $100 billion from 2015 (US Government Accountability Office, 2019). While contracting is one of the more visible forms of outsourcing public functions, there are many other policy tools that facilitate a functioning hollowed out state, including grants, vouchers, and co-production. Grants are lump sum payments, often given to help subsidize administrative costs, to finance start-up costs of a new program, or pilot an innovative program model. Vouchers enable citizens to dictate the allocation of public funds, by choosing where to receive services among a set of providers in a community. Co-production is a process of engaging citizens, usually working as volunteers in the delivery of public services. Pestoff et al. (2012) describe co-production as a mix of activities in which bureaucratic professionals or “regular producers” act alongside citizens as voluntary “citizen producers” to enhance the quality and/or quantity of services they receive.

While the notion of hollowing out carries an undesirable connotation, many have theorized about the virtues of government reliance on nonprofits to perform public services. In many ways, the appeal of nonprofits as public service providers is intuitive. Nonprofits benefit from a “halo effect” in that they are viewed by citizens as among the most trustworthy institutions in society, and they enjoy bipartisan support. In a thorough and compelling fashion, Salamon (1995) articulated the benefits of the government–nonprofit service delivery partnership, describing it as a “win–win” for government as well as the nonprofit sector. Policymakers benefit by being able to meet constituents’ needs through expanded services and increased government funding without increasing the size of the government workforce. Public bureaucracies
benefit by capturing the flexibility, social innovation, and creative programming that may be highly effective in mitigating social or economic problems, but infeasible to implement in legally constrained, rule-bound bureaucracies. Bureaucracies contracting with nonprofits for services have the advantage of accessing professional expertise and capacity to address social problems that public organizations cannot address effectively on their own. As another form of added value, nonprofits can raise money through fundraising efforts, bringing their own resources to bear on social problems. Similarly, they provide spaces for citizens to volunteer and donate, generating “value added” labor and creating social capital in local communities. From their perspective, nonprofits benefit from the infusion of financial resources. Government funding enables nonprofits to expand the scope and scale of their operations and extend their missions to more people in need. It allows nonprofits to professionalize, build capacity, and to hire the most qualified employees. Government funding also confers upon nonprofits a degree of legitimacy, which positions them to compete more effectively for other types of institutional support such as foundation funding.

Despite these benefits there remains a persistent, underlying concern about the fragmentation of accountability under conditions of hollowed out service delivery, and questions remain about how best to assess the performance of these arrangements. In 2007, Frederickson and Frederickson published *Measuring the Performance of the Hollow State*, which offered an in-depth analysis of how five federal agencies within the US Department of Health and Human Services attempted to measure performance under severely hollowed out conditions. The public organizations in their study shared a common need to demonstrate responsiveness to the reforms of the federal Government Performance and Results Act (GPRA) of 1993, while faced with the reality that nearly all of their service delivery had been contracted out to third parties, namely nonprofit organizations. In their research, Frederickson and Frederickson (2007) found that these federal agencies collected a variety of performance data, both quantitative and qualitative, and relied on a wide variety of performance measures including process measures, outputs, fraud control measures, and annual reporting up the hierarchy and to political principals. Their study found that only one agency attempted to measure outcomes of service delivery. Thus, while we know that public organizations aim to capture various types of performance data even under hollowed out conditions, we ultimately know very little about the performance consequences, or outcomes that result from relying on nonprofits to deliver or co-produce public services.

Since the publication of Frederickson and Frederickson’s study, demands for both public and nonprofit organizations to demonstrate accountability and performance have only intensified. As the hollow state became a recognized phenomenon and as residual NPM values linger in today’s landscape of public
service provision, nonprofits have faced increasing pressure to demonstrate their impact and outcomes. These pressures have come not only from the government agencies that fund nonprofits through taxpayer revenues, but also more recently from other institutional funders such as foundations, local intermediary funders such as United Ways, and from private citizens who donate precious time and money to these organizations.

Despite these pressures, and despite the growing adoption of performance measurement regimes within the nonprofit sector, there has been surprisingly little scholarly effort to examine the performance of nonprofit organizations as hollow state actors. Researchers have yet to demonstrate in any systematic way whether hollowing out the state and relying on nonprofits as service delivery partners yields positive benefits for service users, citizens, or the public more broadly. This is the central goal of our book. In the chapters that follow, we are concerned with the overarching questions: “How are nonprofits performing?” and “Does the presence or involvement of nonprofits in a public service improve public outcomes or otherwise add public value?” Through the chapters that follow, we aim to shed some light on these questions. We accomplish this through a diverse set of empirical studies spanning five local policy domains: public education, public health, urban sustainability, public parks and recreation, and public social welfare services. Our goal is to examine the performance of the nonprofit sector across these policy areas, and along the way to identify the factors that contribute to their success in mitigating public problems. We also identify challenges nonprofits encounter in their roles as government partners, and challenges that government organizations face in holding them to account for outcomes. In the next section, we review the literature on measuring performance in the nonprofit sector, followed by a discussion of what it means for nonprofits to generate public value. We then provide a brief overview of each chapter to guide readers through this book.

CHALLENGES OF DEFINING AND MEASURING PERFORMANCE

Shaped in large part by the pressures encountered by nonprofit managers, no other topic has garnered more attention over the last decade in the nonprofit literature as the issue of performance. Researchers have approached the issue of nonprofit performance from a staggering variety of perspectives, examining questions that range from the technical and pragmatic, to the normative and philosophical. While the empirical studies are many and nuanced, we distill the nonprofit performance literature into three debates we regard as the most salient: (1) debates over the best methods and measures for evaluating performance; (2) debates over which actors should define performance, or who gets
Assessing performance and public value in the hollow state

a say in how performance is to be assessed; (3) debates over how performance information should be used.

Methods and Measures for Evaluating Performance: Is There a Best Way?

The most widely debated issue in the literature relates to methods and measures, and is fundamentally concerned with the question of how best to assess nonprofit performance. The question is both a technical and a normative one. To the extent they have measured performance at all, nonprofits have historically relied on output measures, counting for example the number of hot meals served, vaccines administered, or the number of free books or tutoring hours provided to low-income children in an after school program. Increasingly, however, nonprofits are pressured to measure outcomes, to demonstrate the impacts or effects of their programs and services (Mosley and Smith, 2018). One of the central challenges with designing performance systems to adequately capture outcomes is the difficulty of isolating the nonprofit program or service as the cause of improvement in the social condition that program is working to address. For example, the children participating in the after school tutoring program may indeed show improvements in academic performance, but a variety of factors might have led to this outcome, including increased support or help at home from caregivers, support from teachers, improved test-taking abilities, or their knowledge may simply have increased as a result of maturation. Or quite possibly, the children enrolled are inherently different in some way; having an advantage, for example, of their parents being sufficiently informed and motivated enough to enroll them in the program. In short, there are too many confounding influences to link with certainty improved school performance to the nonprofit after school program. This makes it difficult for the nonprofit to claim credit for its perceived program successes.

One way to address this outcome measurement problem is to restrict the program to a randomly chosen group of eligible children, and compare their outcomes against those who were not chosen to participate. Formal program evaluations and randomized control trials (RCTs) are methods of outcome measurement that use experimental designs, involving a set of clients or service recipients who have access to a program or service, where their outcomes or conditions are compared against those of an equivalently matched comparison group who have not had access to the program or service. These types of outcome measurement are superior in the sense that they can isolate the nonprofit program or service as the cause of improvement, but they are problematic for nonprofits in many ways. Nonprofits are mission-focused and typically want to extend their programming and services to as many people as possible; the notion of restricting their programming to a randomly
chosen group for the sake of proving how well their program works is one that challenges the professional ethics of most nonprofit boards, managers, and frontline workers. Even if nonprofit leaders could be persuaded to adopt this type of measurement, the cost of carrying it out represents a formidable barrier. MacIndoe and Barman (2012) found that more than half of nonprofits implementing performance measures of some sort do so without any resources budgeted for them. Moreover, the vast majority of nonprofits lack the expertise and technical capacity in-house to carry out an RCT, and thus have to outsource this program evaluation, often to a university researcher or some other type of evaluation expert, at a substantial cost to the organization. Experimental approaches can more readily isolate cause and effect relationships between nonprofit services and outcomes, but the use of these methods has been slow to diffuse.

Another promising method of assessing impact for nonprofits involves social return on investment (SROI) measures, which rely on a modified form of cost–benefit analysis. The benefits of this approach include monetizing social value generated by nonprofit programs, forcing nonprofits to take a critical look at their own costs, conveying to legislators (who appropriate public funds for the services) the long-term cost savings to the public as a result of investment, and signalling to donors a clear return for each dollar invested. However, this form of performance measurement has been fraught with challenges and slow to catch on in the US, and adopted with only slightly greater frequency in Europe. In addition to resource limitations and capacity constraints, Moody et al. (2015) cite some of the challenges associated with implementing SROI measures, including “selecting the right group of stakeholders to define impact measures, building consensus around indicators and measurement techniques, determining the portion of an observed change that is due to the activities under consideration, overcoming resource limitations and the lack of incentives, and dealing with inadequate commitment among stakeholders” (p. 29).

While there is wide variation in the types of performance measures that nonprofits use, there is a growing consensus that capturing outcomes is preferable to capturing outputs. In light of this, there have been a few industry-level initiatives to standardize metrics for assessing nonprofit program performance, such as the Urban Institutes’ Outcomes Indicator Project. Through this initiative, a team of evaluation experts developed a set of outcomes across 14 programs common among nonprofits, such as assisted living, health risk reduction, prisoner re-entry, job training, community organizing, and youth mentoring. The goal of this initiative was not only to offer nonprofits working in these program areas a ready-to-use resource to encourage internal performance measurement programs, but also to provide funders a way to compare outcomes among organizations providing similar services (Lampkin et al.,...
While this initiative showed promise for helping nonprofits overcome some of the expertise and capacity challenges that have limited nonprofits from implementing performance measurement programs, the initiative never expanded and many nonprofit leaders remain unaware of its existence. Other efforts have included toolkits by state nonprofit associations to help their member organizations develop outcome measures so as to more effectively message to the public about their impact, and enhance prospects for funding. However, these initiatives have been limited to only a handful of states such as Illinois, Minnesota, and New York. In short, the vast majority of nonprofits do not measure outcomes, restrict their measures only to outputs, and lack awareness of existing tools and resources to help build performance measurement capacity. At the same time, they remain very much aware of the pressure to capture outcome data to demonstrate their value to the public.

While there is consensus about the importance of capturing quantifiable outcomes, there is also agreement among scholars that perceptual aspects of performance, such as client satisfaction, are also important and should not be discounted. Scholars agree that nonprofit performance is multidimensional and no single outcome indicator can adequately capture the complexity of nonprofit performance (Thomson, 2010; MacIndoe and Barman, 2012; Sowa et al., 2004). Given the wide range of stakeholder perspectives and complexity of most nonprofit missions, many scholars have called for a multidimensional approach to measuring nonprofit performance. Sowa et al. (2004), for example, state that “hidden behind outcome measurements are complex and diverse dynamics that may vary across and within organizations and programs,” arguing that “both objective and perceptual measures are needed to fully capture the dimensions of effectiveness.” These authors advocate for a multidimensional and integrated model of nonprofit organizational effectiveness that should take into account managerial capacity, program outcomes, organizational financial health, and employee satisfaction. On the program dimension, the authors suggest client satisfaction, along with objective indicators of outcome measures related to a specific program consistent with the program’s theory of change, and managerial perceptions of program performance. Thomson (2010) conceptualizes multidimensionality in terms of the number, collection frequency, and complexity of outcome measures, while MacIndoe and Barman (2012) define it as a combined assessment from resource providers (for example, funding agencies such as government organizations and foundations), organizational networks (accrediting bodies, nonprofit associations, federations,) and internal stakeholders (for example, boards) who collectively influence the use and resources available for outcome measurement.

Hence, while outcome measures are important, they do not constitute a singular or “best way” of assessing nonprofit performance. The multidimension-
ality of nonprofit performance is driven in large part by the reality that these organizations must account to multiple stakeholders, and various stakeholder groups may define and value performance differently. And this brings us to the next major debate in the nonprofit performance literature, which is the question of who, or which groups, has a legitimate claim to defining and assessing nonprofit performance.

Who Defines and Shapes Nonprofit Performance?

Another major theme in the nonprofit performance literature relates to the question of who should be involved in the process of defining performance; which can include playing a role in the design of performance measures, capturing performance data, or processing its meaning through analysis or evaluation. Nonprofit stakeholders rely on different indicators to judge performance, but these indicators might produce conflicting evidence of “good” performance. For example, from the perspective of nonprofit board members, financial cuts that result in cost savings and a balanced budget for the organization may be viewed as an indicator of positive organizational performance, but if those cost savings come at the expense of staff benefits or reduced investment in staff training, it may undermine personnel performance and cause the organization to be perceived as a poorly performing employer by employee stakeholders. Similarly, funders of nonprofit organizations often impose measures of performance that conflict with the organization’s own definitions of performance. For example, a local public health authority that contracts with a nonprofit for community based mental health care may evaluate contractor performance by the number of days the nonprofit utilized inpatient psychiatric hospitals stays for its clients, a costly form of treatment making other alternatives preferable. This creates an incentive for the nonprofit’s leaders and frontline staff to restrict this form of treatment, even in scenarios where they believe it is clinically in the best interest of clients and their families.

Campbell and Lambright (2016) propose a multiple constituency theory of performance, arguing that reporting obligations of funders tend to disproportionately shape the ways nonprofit leaders define and measure performance. This focus on complying with funder-imposed performance requirements often means that nonprofit leaders overlook or minimize other important sources of performance feedback, such as input from their own frontline staff, or from service recipients. The multiple constituency approach is consistent with the spirit of the multidimensional approach; the former emphasizes input from multiple people or stakeholder groups, while the latter emphasizes reliance on multiple outcome measures, which can include input derived from different stakeholder groups. The multiple constituency theory and combined literature on multidimensionality prescribe four sets of actors that play critical
roles in defining, shaping and assessing nonprofit performance: funders, nonprofit leaders (including board members), frontline staff, and nonprofit clients/patrons.

As the hollow state metaphor suggests, funders play perhaps the most important role in dictating nonprofit performance, often by prescribing in contract language what gets measured and how often. Campbell and Lam bright (2016) argue that “because of resource dependencies, reporting obligations developed by funders play a large role in defining performance and determining the measures providers use.” And there is ample evidence to support this contention. In a national study of nonprofit human service providers, LeRoux and Wright (2010) found that greater reliance on institutional funding sources (government, foundations, intermediaries) increased the adoption and use of performance measures by nonprofits. In another survey of nonprofits’ evaluation activities, Carman and Fredericks (2008) found that 71 percent of nonprofits that engaged in some form of performance measurement produced reports for funders about financial expenditures related to program activities. Thomson (2010) studied the impact of government mandates on nonprofit propensity to measure and convey performance, and found that the threat of losing government funding was highly effective in motivating nonprofits to institute performance measurement programs. His study relayed how the city of Detroit, under pressure from the federal Housing and Urban Development agency (HUD), threatened to eliminate funding to nonprofits which failed to adopt performance measures and demonstrate outcomes. Thomson reports that the city’s nonprofit sector prior to the mandate was less successful in measuring actual outcomes than it was in identifying hoped-for outcomes. Unsurprisingly, he found that after the mandate, many more organizations began implementing performance measures, and by five years after the mandate, roughly half of all the nonprofits were able to identify at least one specific, observable outcome.

While mandates and requirements of funders might account for the single biggest factor in nonprofits’ performance measurement efforts, it may have the unintended consequence of diverting scarce administrative resources to capturing data that organizational leaders do not find useful. In an ideal scenario, nonprofit funders such as government organizations and foundation program officers would consult with the leaders of their nonprofit contract, and grant recipients to negotiate what gets measured, to ensure the types of data that are captured are mutually beneficial. Nonprofit leaders and board members also play a key role in defining and measuring performance. Using data from a national survey of nonprofit human service organizations, LeRoux and Langer (2016) find that boards play a role in establishing and measuring organizational performance, although executive directors generally would prefer even higher levels of board engagement in these activities. In another
study of government–nonprofit partnerships, Gazley (2010) found that “the strongest association to real performance improvement comes from the intensity of shared goals and the level of investment in the partnership.” She finds the strongest predictors of effectiveness to include the level of goal agreement between government and nonprofit actors, and age of the partnership, pointing to the importance of time for relationships to form. Yet all too often performance goals are dictated to nonprofit leaders by funding agencies without any negotiation or discussion (LeRoux and Wright, 2010).

Although considered with far less frequency in the nonprofit performance literature, frontline staff represent another critical constituency involved in performance measurement. Benjamin and Campbell (2014) have demonstrated that frontline workers in human service nonprofits play instrumental roles in nonprofit performance by supporting clients in achieving desired outcomes through a process they call “co-determination.” Similar in concept to co-production, co-determination relies on frontline workers to develop partnerships with clients, define desired outcomes and strategies collectively with clients, and support client capacity to take positive action for change. Highlighting the role that nonprofit clients and service users play in shaping performance outcomes, Benjamin and Campbell emphasize the role of client agency in the process: performance gains through co-determination cannot be achieved by frontline workers alone. They argue that “clients are active agents whose desires, attitudes, needs and situational constraints play key roles in the change process” (2014, p. 989). Their arguments and findings align with others, such as Corrigan (2006), who offers evidence from mental health and elder care studies revealing that consumer-led service models lead to improved recovery outcomes for clients.

While frontline workers can thus play important roles in shaping nonprofit performance, their understanding of organizational performance goals should not be assumed. In their national study of child welfare organizations, Jolles et al. (2017) found that fewer than half of managers reported that their frontline workers had a strong understanding of the agency’s performance measures. These authors conclude that managerial communication and board involvement in performance measurement are key predictors of frontline worker understanding of performance measures. Thus the burden falls on nonprofit leaders to create a performance culture, even when the measures are dictated from the outside. Nonprofit leaders can strive for such a culture through transparency, positive messaging about performance measurement, securing staff buy-in, and providing sufficient training and feedback to frontline workers so they fully understand their role in organizational and program performance measurement.

Finally, nonprofit clients, service users, and patrons also shape performance measurement practices. In addition to their roles in co-determination work,
nonprofit clients and service users can also play a role in shaping nonprofit performance efforts by serving on nonprofit governing boards, task forces, and advisory councils (LeRoux, 2009a). Moreover, the majority of nonprofits collect some form of client or customer satisfaction data. Morley et al. (2001) examined the use of client surveys by nonprofits and found that approximately 78 percent of nonprofits conducted client surveys designed to measure satisfaction, client outcomes to be used as performance measures, or both. These authors further found that about half of these surveys also collected information from clients on other aspects of service quality, such as timeliness of service provision and helpfulness of staff (Morley et al., 2001). Similarly, Carman and Fredericks (2008) found that 67 percent of nonprofits collect data on consumer or participant satisfaction on a regular basis. However, there is evidence to suggest that client feedback is generally not accorded the same value or given the same weight as that of other stakeholders. LeRoux and Wright (2010) found that among six different types of performance data collected by nonprofits, client satisfaction data was the only type not statistically linked to decision-making by nonprofit leaders. Similarly, LeRoux (2009b) found that when nonprofit clients are paying customers, nonprofits become more responsive to their input, but when nonprofits receive indirect payments for those clients, nonprofit leaders allocate more of their time and attention to ensuring responsiveness to those entities that do pay. These facts underscore that while nonprofit service users do play a role in defining performance, their input remains somewhat marginalized and is not a reliable predictor of how well the nonprofit might be performing on other measures. This also brings us to our final debate in the nonprofit performance literature, which is how performance data and information should be used.

**How Should Nonprofit Performance Information Be Used?**

The third critical debate in the nonprofit performance literature relates to the question of who benefits from capturing nonprofit performance data, along with whether and how this performance information should be used. In theory, managers develop and implement performance measurement systems to document how well they are doing, identify areas for improvement, and improve decision-making (Moynihan, 2005). From the vantage point of institutional funders such as government, performance information might help identify areas where capacity building and technical support are needed among seemingly lower-performing nonprofit contract partners. From the perspective of nonprofit leaders, it is presumed that performance information will be used to identify inefficiencies, optimize personnel resources, and make necessary adjustments to services and program implementation. Yet given that performance measures are often imposed by outside funders, it remains unclear to
what extent nonprofit organizational leaders can use the performance data they have collected in any meaningful way. On one hand, LeRoux and Wright (2010) found that the range of performance measures used by nonprofits was positively linked to improved strategic decision-making by nonprofit leaders. On the other hand, Carman and Fredericks (2008) found fully one-third of all nonprofit employees view evaluation of their programs and services as a “resource drain and distraction,” making it unlikely that managers will make use of data, even if forced to collect it. These authors also found that some nonprofit managers view performance measurement as a strategic planning tool, and others view it simply as a marketing and promotional tool (Carman and Fredericks, 2008).

Another central question that relates to the use of performance information is whether it should be used by funders in resource distribution or allocation decisions. Will performance data be used by funders to reward nonprofits that appear to meet all their pre-established performance goals, and punish those which do not? Will data be used by funding principals to compare performance across contractors in a provider network, and if so, to what end? Linking incentives to and penalties to performance is fraught with challenges, and yet performance-based contracting is a fairly common practice by government organizations. All too often these arrangements lead to problems of “cream skimming” or incentivizing nonprofits to admit/treat/take the easiest clients, leaving a gap in the market for the most difficult and problematic. When imperfect performance measures are tied to financial incentives, it can easily lead to “performance perversity” (Moynihan, 2014), a situation that incentivizes actors to “game the system.” For example, a private nonprofit charter school might tout its high graduation rate as a measure of performance in order to attract more voucher “customers,” but indiscriminately pass students in order to appear high-performing. Dias and Maynard-Moody (2006) described a similar challenge of performance-based pay arrangements in a study examining outcomes for job training recipients in a welfare-to-work program. With organizational financial incentives tied to the performance outcome of job placement for program recipients, workers were encouraged by managers to steer people into easy-to-find low-paying jobs, rather than pursue extended training or education that might have led to a higher-paying job and greater economic self-sufficiency in the long term. Dias and Maynard-Moody describe this as a “performance paradox”: the same efforts made to meet contractually imposed performance metrics ironically produce negative program practice and poorer client outcomes, “doing little to reduce long-term welfare use or diminish recipients’ poverty.” Hence, while use of performance information to distribute financial incentives to nonprofits by their funders may appear as a logical way to incentivize performance in theory, the unintended
consequences of these arrangements may outweigh any potential performance improvements.

NONPROFITS AND PUBLIC VALUE CREATION

Given the complexities and challenges associated with measuring nonprofit performance, another way to think about performance in the hollow state is through the lens of public value. Indeed, Bryson et al. (2014) suggest that public value can serve as a performance measurement and management framework, pointing out via Kalambokidis (2014) that an advantage of the public value idea is that there is no single bottom line. Thus, as a perspective for evaluating the performance of hollow state arrangements, it is consistent with the nonprofit performance literature reviewed earlier, which concluded there is no single best measure or metric, and it is informed by a variety of stakeholder perspectives. A robust body of scholarly work has emerged from public interest theory on organizational publicness (Bozeman and Moulton, 2011), and public value (Moore, 1995; Stoker, 2006; Bozeman, 2007). So extensive is this literature that it would be impossible to review it in the space of this chapter. For readers wishing to delve further into this broad expanse of work, we would direct them to Van der Wal et al.’s (2013) meta-analysis of public values research, based on 397 articles. Here, we will simply define what is meant by public value and consider some of the insights from the literature most relevant to public values in the context of the hollow state.

In tracing the history of public interest theory, Bozeman (2007) argues that public interest represents more of an ideal, while public values have “specific, identifiable content” (p. 12). He defines the public interest as “the outcomes best serving the long-run survival and well-being of a social collective construed as a ‘public’” (2007, p. 12), while he defines public values as “normative consensus about a) the rights, benefits, and prerogatives to which citizens and should (and should not) be entitled; b) the obligations of citizens to society, the state, and to one another; and c) the principles upon which governments and policies should be based” (2007, p. 13). Whereas Bozeman focuses on the society or policy level, Mark Moore (1995) focuses more on the role of the public manager in creating public value (Bryson et al., 2014). Moore’s (1995, 2014) basic premise is that if the primary objective of private firms is to produce shareholder value, then the central goal of public managers should commensurately be to create public value. His conception of public value encompasses both traditional values of efficiency and effectiveness of public bureaucracies, but also emphasizes procedural and substantive justice. Thus, assessing whether public value has been created from Moore’s perspective requires some combination of input, process, output, and outcome measures.
Moore (1995) conceives of public value as value creation, and role of managers is central to that process. He asserts that:

Public managers create public value. The problem is that they cannot know for sure what that is … It is not enough to say that public managers create results that are valued; they must be able to show that the results obtained are worth the cost of private consumption and unrestrained liberty forgone in producing the desirable results. Only then can we be sure that some public value has been created. (Moore, 1995, pp. 57, 29)

Moore offers a strategic triangle to be employed by managers in their calculus of public value creation. He argues that strategy: (1) must be aimed at achieving something substantively valuable (must have or create public value); (2) must be legitimate (that is, politically feasible); and (3) must be operationally and administratively feasible (Moore, 1995, pp. 22–23). Additionally, Moore (2014) proposes the concept of “public value accounting” for public managers to ascertain whether and how public value has been created. In the process of public value accounting, he suggests that managers need to look at costs and benefits as well as at less tangible aspects of value when they assess public value creation (Moore, 2014). Thus, while early foundational works on the topic of public values and public value accounting offered broad conceptualizations of these ideas, they lack specific direction for measuring public value creation, and as Bryson et al. (2014) point out, these early voices are largely silent on the role of nonprofits in public value creation. Scholars have expanded upon this work more recently (Moulton and Eckerd, 2012; Stoker, 2006; Aldridge and Stoker, 2002) to offer more specific indicators by which we might assess public value creation. We shall examine these in greater detail shortly.

Public Values in the Hollow State

Drawing on Bozeman and Moore’s work, Bryson et al. (2014), propose a new paradigm that they term “public value governance” for today’s era of managing contract and multi-sector networked arrangements. These authors contrast public value governance with traditional public administration and its successor, new public management (NPM), arguing that different values are accorded greater priority in each of these paradigms. They argue efficiency to be the key value characterizing traditional public administration; efficiency and effectiveness to be the main values underlying new public management; while “efficiency, effectiveness, and the full range of democratic and constitutional values” characterize this new era of public value governance (2014, p. 446). Accordingly, they see the need for public managers to demonstrate adherence to an increased range of accountability mechanisms, including not only the
“old” NPM type mandates for performance reporting, but also to an expanded range of interests including citizens, community values, political norms, and professional standards. Under this new paradigm of public value governance, public values are less determined by political goals and legislative priorities, but rather determined through a process of negotiation led by public agencies, involving deliberation and dialogue with citizens, community groups, private organizations, and other stakeholders. Thus the new public governance is both process and outcome oriented.

The notion of public value governance was derived in part from earlier work by Stoker (2006), who argued that “public value management” is the new paradigm for the contemporary hollow and networked state, replacing new public management and traditional public management. Stoker (2006) defines the public interest in the public value management paradigm as “individual and public preferences produced through a complex process of interaction that involves deliberative reflection over inputs and opportunity costs.” Stoker argues that public value management must answer three core questions about what it will deliver when it comes to efficiency, accountability, and equity, and how it will deliver benefits in these areas. His emphasis on negotiated goal setting and developing citizen capacity for the expression of individual rights and responsibilities is very consistent with those articulated by Bryson et al. (2014). Stoker’s position on the important role of managers in creating public value parallels those of Moore (1995), but Stoker situates his arguments more squarely in the context of the hollow state.

Stoker argued for the field to place a greater premium on a public service ethos rather than a public sector ethic. He envisions this public sector ethos as transcending sector distinctions, serving as a moral guide not only for public managers tasked with procurement of services, but also for those working in private for-profit or nonprofit organizations tasked with public service. He argues:

there is no ideological dimension to deciding who provides services and no particular moral virtue in people receiving their wages directly from the government. The assumption is that although direct provision from within the organization may be appropriate in some circumstances, in many others the advantages of private or voluntary sector provision will be greater. The private sector, for example, may be able to combine vitally needed investment with the responsibility for providing services. The voluntary sector, through its wider involvements, may be able to guarantee a more joined-up or seamless service for the public. Above all, what public value management expects is for a relational approach to service procurement. There should not be a great divide between client and contractor; both should see each other as partners looking to sustain relationship over the long run and should not be narrowly focused on any contract. (Stoker, 2006, p. 48)
In assessing public value creation, Stoker (2006) calls for pluralist stakeholder review of hollow state actors, and also calls for benchmarking and open competition in the procurement process.

Using slightly different terminology, Bozeman too pointed to “manifestations of market-based policies” such as privatization, contracting, use of vouchers, and government management of contractors as fundamentally changing the meaning of “publicness” (Bozeman, 2007). Building on this idea, Bozeman and Moulton (2011) argue that “just as organizations and policies can be evaluated by the extent to which they are constrained or enabled by political and economic authority, policies and organizations can also be evaluated by the extent to which they achieve public values”; a process described by Moulton (2009) as “realized publicness.” While the concepts of realized publicness, public value, achievement, and public value accounting remain somewhat diffuse, based on our rendering of the literature, we attempt in the next section to narrow in on some specific attributes we might look for in determining whether public value has been created by the hollow state.

**Specific, Identifiable Content of Public Values**

Bozeman argued that public values have “specific, identifiable content” (Bozeman, 2007, p. 12), and while he makes no mention of how we might extract this content from hollow state arrangements, we make some inferences from the literature throughout this book to guide our assessment of performance and public value in the hollow state. First, we would point to Aldridge and Stoker’s (2002) five elements comprising a common set of public values shared by those engaged in public sector service delivery: a performance culture, a commitment to accountability, a capacity to support universal access, responsible employment practices, and contribution to community well-being. When we see these attributes present, we consider this to be public value creation, and we balance this against traditional perspectives of performance. Relatedly, Moulton and Eckerd (2012) theorized that specific resource streams may shape nonprofits’ value systems, positing that greater reliance on government funding institutionalizes public value within nonprofits. They developed and tested a “nonprofit sector role index” consisting of six expressions of public value: service delivery, innovation, advocacy, individual expression, social capital creation, and citizen engagement. When we see these attributes present, we also consider this to be public value creation, and similarly consider these benefits alongside traditional perspectives of nonprofit performance. Finally, we consider some indicators of public value creation suggested by Bryson et al. (2014). When we observe in hollow state arrangements instances of citizens acting as problem-solvers or co-creators actively engaged in what is valued by the public, instances of civic learning...
and/or building civic capacity, and instances of citizens shaping policy through dialogue and deliberation, we consider this, too, to be public value creation. In our assessment of performance in the hollow state, we thus consider traditional perspectives on how performance is measured, but also whether some degree of public value is generated by these arrangements, using the “identifiable content” specified here.

OUTLINE FOR THIS BOOK

Toward the conclusion of their seminal work elaborating the concept of the hollow state, Milward and Provan pose the question: “Is the hollow state better or worse than the bureaucratic state? This is a question that at present cannot be answered” (Milward and Provan, 2000, p. 364). While we may never be able to definitively answer this question, we have more answers today than we did 20 years ago. The public management of today calls for a different question. The question is no longer one of whether “private provider or government is better” but whether, and to what extent, public value exists today under the conditions of hollow state. And looking to the future, how can public organizations best foster public value creation in the design of outsourced service delivery arrangements? In this book, we attempt to shed light on these questions.

To this end, we have assembled a set of empirical studies that allow us to take stock of performance and public value in the hollow state. While the majority of these studies are set in the context of the US, and one in Europe, the themes of this book are applicable to most countries in which the state relies heavily on the nonprofit sector and/or voluntary organizations to deliver public services. Each of the following chapters in this book examines a different local policy or public service area. These works were intentionally selected to showcase the breadth of nonprofit involvement in local public service delivery. While many people tend to associate the government–nonprofit partnership with social welfare provision, governments enlist the aid of nonprofits to carry out a variety of local services and policy objectives. We selected studies for this book that represent some of these lesser-known, yet highly common partnerships with nonprofits for services spanning K-12 (kindergarten to 12th grade) public education, parks and recreation services, public health, and community development, as well as social welfare. These chapters were also selected for their methodological diversity, creatively employing new and interesting data, both qualitative and quantitative. Each empirical chapter not only examines nonprofit performance in a different policy or service area, but also highlights a different policy tool or financing mechanism that governments use to carry out services through nonprofits, including vouchers, fee-for-service contracts, and government-supporting charitable foundations, a form of co-production.
Collectively, these studies help us to answer the question of whether and how nonprofits improve public service outcomes or otherwise create public value. Here we offer a brief overview of each chapter, in an effort to provide readers with a road map for the rest of the book and to highlight the service and/or policy areas that are the focus of study in each. Every chapter represents a self-contained study; it is not necessary to read them in sequence. While we hope that readers find value in the entire book, this overview is designed to provide an information shortcut of sorts to those who may wish to skip around or begin at the chapter or policy issue of greatest interest to them.

Chapter 2 presents the first of these empirical studies. In this chapter, Michael Ford examines the ability of nonprofit charter schools to effectively address the “wicked problem” of urban public education. This case is set in the city of Milwaukee, home to the oldest and largest school choice voucher program in the US, and where about 40 percent of publicly funded K-12 students attend a nonprofit voucher or charter school. The use of vouchers as a mechanism for distributing government funding is rooted in public choice logic. In theory, allowing consumer choice to shape the market for local public education should force education providers to compete in ways that improve service quality. Yet, the results of Ford’s study demonstrate the flaws in this logic, and highlight the difficulty of holding government-funded service providers to account when public funding is not conditioned upon clear performance expectations, but relies instead upon market forces. Absent any imposed performance or accountability measures, and given the highly market-driven nature of this public good, Ford uses school closures as the measure of accountability. The logic suggests that when schools exhibit poor performance, parents will send their children to another, better-performing school and eventually the low-performing schools will be selected out of the market. Using a panel dataset of publicly funded nonprofit schools spanning four years, Ford’s survival analysis reveals that while enrollment growth and other organizational factors can reduce closure rates, student outcomes on reading and math proficiency have no relationship to charter school failure. In other words, nonprofit schools failing in their educational mission face no real consequences for their low performance.

Chapter 3 presents the case of public health, examining whether hollow state health care actors perform in ways that improve health outcomes at the community level. In this chapter, Zhao and LeRoux offer results from a national study of service contracting by over a thousand local public health departments (LHDs) across the US. The consequences of contracting have been studied extensively, but as the authors point out, most of the research to date has been concerned with the efficiency question: that is, whether or to what extent contracting reduces costs. Zhao and LeRoux speak to a lesser-studied contracting question which is that of effectiveness: whether contracting yields better
service quality or improved outcomes. Working from theoretical perspectives of transaction costs, the authors theorize that outsourcing may improve service quality, but only under ideal conditions in which the services being contracted are characterized by low transaction costs. The authors examine contracting for four types of health services which vary in their degree of transaction costs: mental health prevention; treatment for HIV/AIDS and other sexually transmitted infections; pregnancy prevention; and regulation of tobacco retailers and enforcement of smoking ordinances. To examine the impact of contracting for each of these functions, the authors combine survey data capturing contracting decisions from the National Profile of Local Health Department Study collected by the National Association of County and City Health Officials, and pair it with outcome data from County Health Rankings and Roadmaps (CHRR), for four corresponding health outcomes at the county level: frequent mental distress, sexually transmitted infections, teen births, and adult smoking rates. The findings indicate that contracting out lower transaction cost services is more likely to reduce risks of contract failure and produce better health outcomes.

In Chapter 4, Nathaniel Wright and Tony Reames examine the role of community action agencies (CAAs) in contributing to the sustainability of distressed neighborhoods. Originating through the federal War on Poverty funds during the 1960s, CAAs were an important fixture in American communities long before the concept of a hollow state emerged. For nearly seven decades, CAAs have provided for basic needs such as housing, and stimulating economic investment and opportunity in impoverished communities that markets have largely failed to serve. However, the mission of these organizations has evolved over time to include a much wider array of services and public responsibilities. Community sustainability represents one area where CAAs have played an increasingly important role in recent years. CAAs have assumed responsibility for many aspects of sustainable development, including residential weatherization, energy conservation, transportation, and the provision of housing and shelter to urban and rural communities. In light of this, Wright and Reames set out to identify the factors that shape the effectiveness of CAAs in creating sustainable communities, through a study of these organizations situated in major US cities. They find that community engagement, human resource capacity, county/regional government collaboration, government funding, and revenue diversification are important predictors of CAAs’ community sustainability performance. These findings have important implications: government funding drives nonprofits to perform better, yet government does not have to bear the entire funding burden in order to have this effect. Their results further suggest that citizen participation, a fundamental public value, contributes to greater effectiveness when it is embraced and implemented by nonprofit leaders.
Authors Daniel Cheng, Yu Shi, and Simon Andrew take a somewhat different approach to the government–nonprofit partnership in Chapter 5, by examining how private philanthropic foundations help to raise funds for public parks and whether this philanthropic support enhances performance of the park system. The growth of private nonprofit foundations that raise funds for government organizations underfunded by tax revenues is a trend that has been studied in several contexts including national parks (Yandle et al., 2016), K-12 public schools (Nelson and Gazely, 2014), and public libraries (Schatteman and Bingle, 2015). Cheng and his colleagues examine park-supporting foundations, also known as “friends of the park” groups, as a form of co-production. Their study documents the growth of charitable support for public parks in the face of declining public revenues, and examines the link between this private funding and performance of public park systems. They measure performance by capturing park acreage, public access, and outputs (play structures and facilities). Studying the 68 largest city park systems in the US, they find that total spending by nonprofit groups largely has a positive effect on city park performance. Another important factor driving city park performance is education level of the population, suggesting that citizen demand-making is also important. However, these authors point to several equity implications of their findings. These authors also caution us that while philanthropic support might supplement public appropriations and user fees at the margins, it is not a substitute for tax revenue. If citizens want more and better parks, they will have to make demands of city government, and if nonprofit parks groups can help spur this mobilization, they are helping to create public value through increased citizen engagement.

In Chapter 6, Peter Raeymaeckers and Peter Cools examine an increasingly common dilemma confronted by government-funded social service providers, which is the pressure to demonstrate their performance through measurable outcomes as a condition of continued funding. Many nonprofits confronted with the scenario simply acquiesce to externally imposed performance metrics demanded by their government funders, even if leaders believe these measures to be inadequate for capturing the organization’s impact or, worse, a waste of managerial effort. The reality is that many nonprofits feel powerless to question their government funders, since their current and future funding may be threatened by challenging these measures. Yet, Raeymaeckers and Cools demonstrate how an organized, strategic effort among a collective of social service organizations tasked with combatting poverty successfully challenged a new mandate for funder-imposed performance measures, and instead negotiated a process with their subsidizing government to determine how performance should be measured. Their study is situated in the Belgian city of Antwerp and examines a network of social service organizations known as “associations where people in poverty lift up their voice.” Using qualita-
ative data from interviews and focus groups with agency coordinators, social workers, people in poverty who receive services, and representatives from the local government who fund the providers, Raeymaekers and Cool yield illustrative findings that depart from the conventional approach of acquiescence, and might serve as a guide for other nonprofits that strive for negotiated performance assessment with their funders. They reveal a range of tactics used by social service providers in this process, including using the power of their own professional expertise not only as a form of control, but also as a form of resistance and for promoting the values of a social service organization, especially the value of participatory practices and client inclusion. In doing so, they conclude that social service providers might achieve a more balanced approach to performance measurement that takes into account process, as well as more diverse range of stakeholder perspectives including those of clients.

Finally, in Chapter 7, we distill the lessons offered by these studies for governing the future of the hollow state and propose a new framework for integrating traditional performance measures with measurement of public value creation. We frame our discussion with the ultimate objective of aiding policymakers and public sector leaders in the design of their partnerships with non-governmental organizations in ways that yield the best deal for the public. We highlight consistencies in these studies that point to themes for enhancing the accountability, performance, effectiveness, and public value creation of third-party organizations involved in the production and delivery of public services. We also highlight consistencies in the evidence with regard to challenges faced by these organizations in achieving their goals, including capacity issues, funding issues, and government disinvestment, along with flaws in the design of current accountability and oversight mechanisms. We also discuss roles for the public and individual citizens, including specific ways they can engage with these organizations to help improve nonprofits’ accountability, performance, and public value.

NOTE

1. While the term “hollow state” diffused widely with the publication of Milward and Provan’s 2000 article, it was introduced earlier by the authors in their book chapter titled “The hollow state: Private provision of public services” (Milward and Provan, 1993). Others had introduced the concept even earlier, although using different terminology, such as the “shadow state” (Wolch, 1990).

REFERENCES


