WashEval DEEP DIVE

Steven Putansu, PhD

June 10, 2020

12:30 to 1:30pm

Find the Book HERE! (https://t.co/feWQ2PJe9u?amp=1)

20% off June 1-29 w/ promo code: **khQmNFPwTTAkSd3**

Politics and Policy Knowledge in Federal Education

Confronting the Evidence-Based Proverb

Steven Putansu

palgrave
macmillan

Politics and Policy Knowledge in Federal Education

"In this important book, Steven Putansu argues cogently that a simplistic adherence to the idea of 'evidence-based policy' has negatively impacted how we understand policy and politics. His careful analysis of US federal education policy over 50 years, provides a clear case for embracing 'evidence-based governance'—a broader and more realistic understanding of the interplay of policy knowledge, decision purposes and politics."

—Jenny M. Lewis, *Professor of Public Policy, The University of Melbourne, Australia*

"Politicians demand that governments produce policy knowledge but then struggle to put that information to good use. With a rare combination of a practical experience, in-depth knowledge of the history of federal programs, and on-point use of scholarly theory and cutting-edge research, Putansu offers a compelling assessment of this struggle, while signposting the factors that make a more realistic marriage of policy knowledge and policymaking possible."

—Donald Moynihan, *Professor, McCourt School of Public Policy, Georgetown University, USA*

"It is rare that scholarship in public administration or political science inform one another. It is rarer still that students of each are informed by the practical acumen of those in government. In *Politics and Policy Knowledge in Federal Education*, Putansu provides a refreshing exception to these norms, as he skillfully unpacks the complexities and politics of information production and consumption in the US federal government."

—William Resh, *Associate Professor, Sol Price School of Public Policy, University of Southern California, USA*

Contents

About the Author

Steven Putansu is a *pracademic* who splits his time as a public servant at the US Government Accountability Office (GAO); a professorial lecturer at American University; and a public management scholar. Steven studies the intersection of evidence and politics in decision making, strategic management practices to enhance evidence-based decisions, and applied work in a variety of policy areas. At GAO, Steven has contributed to products on best practices for evaluating government programs, audit criteria, and frameworks for assessing performance systems, interagency collaboration, human capital, risk management, duplication, and fraud in the federal government. He has applied these and other frameworks in GAO engagements in a range of policy areas, including education, defense, international affairs, nuclear energy, project management, wildfire response, immigration, climate change, veterans affairs, freight rail, and traffic safety. His methodological contributions have ensured rigorous technical standards in hundreds of GAO performance audits, provided policymakers with high-quality information for overseeing federal programs, and ensured the best use of taxpayer dollars, helping to contribute over $260 billion in financial benefits for the federal government. For these contributions, Steven received an Arthur S Flemming award for Leadership and Management in 2017.

CHAPTER 2

No Panacea: The Purposes, Uses, and Limitations of Policy Knowledge

Advocates for evidence-based governance have generally made broad calls for more policy knowledge, or have championed a narrow focus on specific data, information, or evidence. These calls for reform have also commonly featured overly broad promises that the resulting policy knowledge will be a panacea for decision making. This book argues that broad agreement on the importance of policy knowledge enables reformers to overlook important differences in the purposes, uses, and limitations of this knowledge in specific decision making contexts. Further, and counter to the intent of evidence-based reformers, this has actually hindered efforts to increase the role of data, information, and evidence in decision making by ignoring and excluding alternative sources. By emphasizing specific forms of policy knowledge and implicitly devaluing others, evidence-based reforms have contributed to recent retreats from political support for various forms of data, information, and evidence. This includes attempts to defund research grants, restrictions on certain types of data collection, and increasing reliance on "ordinary knowledge" to support decision making (Lindblom and Cohen 1979).

This chapter introduces a typology of policy knowledge rooted in the quality of data, information, and evidence. Then, it adds a dimension of purpose, which considers the different aspects of governance that can be focal points for policy knowledge. Finally, it considers the development of policy knowledge as a process, through which decisions about purpose, coverage, and quality shape the production of data, information, and evidence. These additional dimensions further demonstrate the need

S. Putansu, *Politics and Policy Knowledge in Federal Education*,
https://doi.org/10.1007/978-3-030-38395-4_2

for evidence-based reformers to be explicit about the qualities and limitations of the data, information, and evidence they produce, and the relevance of alternative types and sources for specific purposes. The chapter then considers the different ways that policy knowledge can impact decisions. Calls for evidence-based policy have promised a panacea for decision making, but this analysis shows that three limitations in the treatment of policy knowledge have consistently limited utility of the reforms. Specifically, reforms have relied on narrow definitions of policy knowledge, failed to consider different purposes it can support, and overlooked the different ways that data, information, and evidence are used. This discussion paves the way for more detailed consideration, in Chapter 3, of the role of policy knowledge, vis-a-vis political forces, in the decision making process.

Defining a Typology of Policy Knowledge

To overcome the evidence-based proverb, any examination of the role that policy knowledge plays in the political decision making process must rely on a more complete and nuanced model of data, information, and evidence. This chapter introduces a four-part typology of data, information, and evidence that allows for more complete cataloging, understanding, and examination of the different ways that policy knowledge is produced, considered by decision makers, and ultimately used. Rather than attempt to establish a singular hierarchy of policy knowledge rooted in methodological approaches, this typology is rooted in a consideration of the quality of policy knowledge, as defined by varying levels of coverage and rigor.[1] In this context, coverage refers to the representativeness of the data, information, or evidence being considered, and ranges from an individual experience, on the low end, to a complete set of all relevant observations on the high end. Rigor refers to the analytical sufficiency of the policy knowledge, in terms of whether it is valid and reliable for an intended purpose, which generally ranges from ad hoc on the low end, to more systematic on the high end. Rigor is used as a summative term throughout this text to represent the range of practices and methods used to improve the reliability and validity of policy knowledge, including but

[1] This typology borrows from, integrates, and updates a variety of definitions offered in Feldman (1989), Lindblom and Cohen (1979), Majone (1989), Shulock (1999), and Weiss (1977).

not limited to greater qualitative depth, more structured evidence gathering, inclusion of additional quantitative variables, or higher levels of experimental control. As shown in Fig. 2.1, dividing policy knowledge based on coverage and rigor creates four quadrants of data, information, and evidence.

Before turning to the implications of this model, an illustrative discussion of the elements of policy knowledge in each quadrant is instructive. The lower left quadrant illustrates a relatively ad hoc approach to generating policy knowledge, with a small and unrepresentative level of coverage. A single data point (i.e., an observation) may be that an agency hires a new employee. Information (i.e., detail) about that person may include characteristics of this employee, such as gender, race, and education. This becomes evidence when an association between the data and information (i.e., an assertion), such as "we hired her because she met the education requirements for the position," is made. The top left quadrant illustrates higher coverage. For example, this might include data on the last 20 hires (i.e., collection), with information (i.e., description) that presents the proportions of those in each of the demographic categories, and evidence (i.e., correlation) that shows whether education is associated with hiring decisions.

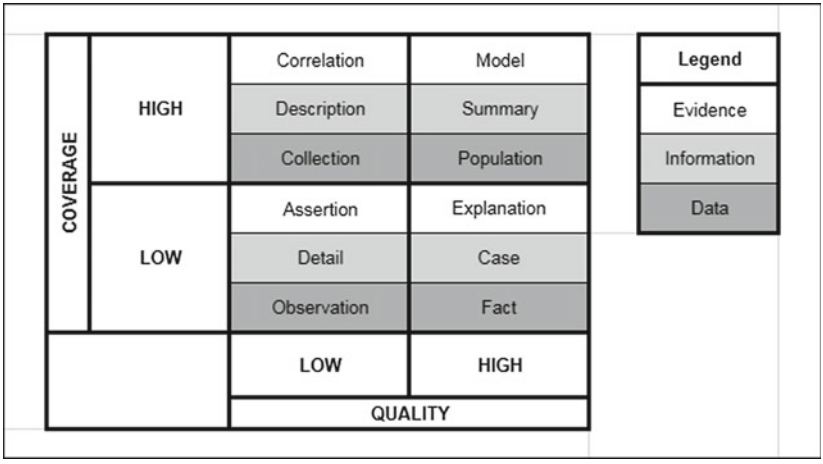

Fig. 2.1 Policy knowledge types, coverage, and quality

Using a systematic analytical approach introduces more rigor into the production of data, information, and evidence, which enables the data to be used for more complex purposes. The bottom right quadrant illustrates the addition of quality to the low coverage examples above. Specifically, if the agency assessed the new hire's education by looking at their resume, this would be of lower rigor than if they confirmed this with an official transcript (i.e., fact). Similarly, the agency may have details collected about gender, race, and education, but they could more rigorously review other relevant characteristics about the individual (i.e., case), including past work experience, references, and so on. Finally, by systematically assessing how each of these characteristics was valued in the hiring decisions, they may develop better evidence (i.e., explanation) of what drove the hiring decision.

The rigor dimension can also be added to the higher coverage scenarios, as illustrated in the top right quadrant. When a collection is representative of all observations (i.e., population), it provides a more complete picture of the data. Similarly, having information about the full population (i.e., summary) allows for more thorough descriptive analyses. Finally, when researchers systematically consider potential correlations (i.e., model), they enable an assessment of strength of those relationships while controlling for other factors. For example, the company may review data on all applicants and hires, may compare informational summaries of the demographic characteristics of each, and may develop an evidentiary model to determine the independent associations between those demographic characteristics and hiring decisions.

Similar to evidence-based reforms, this typology suggests variation in the relative value of different forms of policy knowledge (high coverage, high rigor policy knowledge should generally be trusted more than low coverage, low rigor). In practice, this means that an individual assertion of education as the reason an employee was hired should be overruled by a detailed explanation rooted in a full case study that offers contradictory evidence, by analysis that shows a negative correlation between education and hiring, or by a model that shows education has no impact, once other factors are controlled for. However, in contrast to evidence-based reforms, this typology does not presume that there is one best type of evidence for all decisions, but rather acknowledges that policy knowledge in all four quadrants could be sufficient and appropriate for specific decision making purposes.

The appropriate coverage of policy knowledge is rooted in the scope of the decision at hand. For example, government investigations into criminal activity most appropriately focus only on "persons of interest" related to the alleged crime, while the decennial census is required to consider all people living in the United States. Similarly, the sufficiency of rigor of policy knowledge depends on its intended purpose. For example, members of Congress who receive a high number of ad hoc complaints describing a problem may consider that sufficient for them to attempt to step in and serve their constituents through individual casework, while they may seek more systematic and rigorous summary information before pursuing more formal oversight actions. Throughout this text, the typology offered above is used both to comprehensively classify different types and sources of data, information, and evidence presented throughout the book, and to offer deeper insights into whether and how policy knowledge has been used to support specific governance decisions.

The Purposes and Creation of Policy Knowledge

Decision making can be understood and improved through detailed understanding of the various sources, types, and uses of policy knowledge (Weiss 1977). Policy and decision making literatures offer an abundance of factors that can influence whether and how individual decision makers receive or observe policy knowledge, whether and how they interpret it, and how this interpretation impacts their perspectives and strategic approach to decision making. This section explores the purposes of policy knowledge and link these purposes to the typology introduced in the previous section. It then shows how the generation of policy knowledge can identify and reinforce these links.

Evidence-based reforms are founded on the belief that data, information, and evidence can support decision makers and lead to better outcomes. However, these reforms often do little to detail the kinds of decisions they can support, instead promising a panacea for decision making. In the context of governmental decision making, policy knowledge needs to be provided to several different players in the decision making process: (1) managers and employees can use it to improve the way they function, (2) Congress and the President can use it to improve policy decisions, or (3) citizens can use it to decide their preferences for a particular mix of service provisions and mode of service delivery. For example, an agency

department head may be interested in program effectiveness, staff distribution, and cost comparisons, a bureau head may be more concerned with outcomes, and line managers may be focused on inputs and processes. These differences are copacetic with the long-accepted axiom that, in governance, where you stand is where you sit.

Philip Joyce (2005) noted that different actors may rely on different forms of policy knowledge about programs, as they are participating in different activities and working toward different final products. Donald Moynihan (2008) provides additional support for this finding, showing that individual roles impact the selection and use of data, information, and evidence. In line with these multiple perspectives, scholars and practitioners have identified a broad array of potential uses of policy knowledge in decision making that vary in both number and detail. They include forty-four potential uses offered by van Dooren (2004), eleven major uses discussed by Hatry (2006), eight managerial uses enumerated by Behn (2003), and four types of managing performance illustrated by Bouckaert and Halligan (2008), among others. In other words, policy knowledge can help improve various aspects of decision making, but only when it is aligned with the needs of decision makers. Further, others note various ways that policy knowledge can be misused, including in the commissioning of work, the process for developing it, or in manipulation of the results (Alkin and King 2017).

The choice of what to study has important implications for whether policy knowledge can provide a complete and accurate understanding of a policy or program. For example, if all data, information, and evidence about a program were to focus on economy—the total cost of the policy—then decision makers would not have a way to understand whether the policy effectively achieved goals or resulted in equitable outcomes, or whether these outcomes were being produced efficiently. All of these choices of selection and prioritization mean that the range of policy knowledge available for a given policy, its processes, and its results will be dependent on the interests, priorities, and independence of knowledge producers.

This examination of different uses of policy knowledge obviates the impossibility of applying a one-size-fits-all approach to evidence-based policy. No single source of policy knowledge is likely to produce data, information, and evidence reflecting a level of coverage and quality that fulfills the wide array of perspectives and purposes needed for decision

making. Instead, it shows that when decisions are being made, considering how data, information, and evidence contribute to specific purposes will be critical to ensuring that policy knowledge meets the needs of decision makers.

This chapter now turns to the process of creating policy knowledge, in order to show how purposes and needs can be integrated into the development of data, information, and evidence. There are four broad activities required to develop policy knowledge. First, there must be a specification of program goals. Next, relevant data must be collected. Then, data must be analyzed, described, and interpreted in order to turn it into information and evidence. Finally, the resulting policy knowledge must be communicated to decision makers.

Production of policy knowledge begins with decisions about what to measure. Producers of policy knowledge, in determining their scope and research questions, must decide where to invest limited resources in order to serve one or more decision making needs. This includes deciding which policies or programs to review, what processes to focus on, and which kinds of results to measure. In generating policy knowledge, any policy (or combination of policies) may be reviewed, and evaluations of those programs may center on clarification of objectives or measures of public opinion, inputs and program costs, implementation, outcomes, or impacts. Goals can be stated in general (strategic) or specific (actionable) terms. For example, either a broad goal of transparency or a more specific goal that records should be publicly available online may be specified. Goals can also have varied timelines, and may be short-term, medium-term, or long-term. In some cases, efforts to develop policy knowledge may consider all three timeframes, with intermediate outcomes identified at each stage. Finally, goals might be specified at different institutional levels. For example, Congress might set a goal of a balanced budget that impacts the entire federal government, an agency might set a goal of implementing new systems across the entire organization, and a program might set goals about the provision of its specific services.

Once the policy and relevant aspects are chosen, the review may proceed to analyze one or more kinds of results, including economy, effectiveness, efficiency, and equity, which may refer to distributional equity, outcome disparities, procedural fairness, and process equity (National Academy of Public Administration 2000). These approaches may consider broad, macro-level uses, such as goal setting, others focus on meso-level management issues, and others at micro-level details of implementation

(Roberts 2019). These multiple goals reflect the desires of multiple actors in the decision making process, as different actors have different theories about the programs, understandings of cause-effect relationships, and prioritization of goals.

The multitude of ways to specify goals, as described in Table 2.1, highlights the first potential challenge for evidence-based reforms. In order to sort through the wide array of different foci and analytical considerations, stakeholder buy-in is critical. Evidence-based reforms may profess to treat all goals equally, try to prioritize them independently, or seek to prioritize them through interaction with stakeholders, but if the relative importance of goals is disputed, all the resulting policy knowledge may be called into question (Newcomer 2007; Julnes and Holzer 2001; Posner and Fantone 2007; Pollitt 2006; Melkers and Willoughby 2005). In the legislative context, repeated findings suggest that evidence-based reforms are mistrusted, not valued, or completely ignored by decision makers, due in large part to their exclusion from the process of setting goals and metrics (Newcomer 2007; Pollitt 2006; Posner and Fantone 2007). This is true at the state level (Melkers and Willoughby 2005) as well as in a federal context (Posner and Fantone 2007). The enumeration of goals may also suffer from a lack of strategic focus (Posner and Fantone 2007; Joyce 2005), from insufficient specificity to be actionable (GAO 2005; Behn 2003), from inappropriate timelines (Newcomer 2007), or from a focus on the wrong institutional level (Behn 2002, 2003; Pollitt 2006; Talbot 2005; Barnow 1992). Finally, one or more stated goals may be in conflict with others (Wildavsky 1974; Davis et al. 1966). Beyond this, different

Table 2.1 Example dimensions of policy knowledge use

Dimension	*Aspects that policy knowledge might focus on*
What to study	Government-wide activities, entire policy areas, one or more programs, specific implementation practices
Policy levels	Macro, Micro, and Meso level goals and objectives
Program levels	Prioritization of goals, management, implementation, impacted populations
Operations	Inputs, outputs, outcomes, impacts
Analytical focus	History, public opinion/support, economy, effectiveness, efficiency, social equity
Users	Policy makers, managers, implementers, oversight bodies, interest groups, the general public

actors in the process are interested in different goals. The development of policy knowledge must consider which of these goals they intend to contribute to, and how they intend to manage the resulting ambiguity and conflict (Table 2.1).

These limitations of goal specification cast doubt on the ability of a single reform effort to effectively produce data, information, and evidence for a variety of decision making needs (Radin 2006). In other words, the reality and complexity of government programs and policies prohibit the use of a one-size-fits-all approach, and assure that no single effort will be a panacea for decision makers. Second, when evidence-based reforms specify goals, they may either prioritize competing goals or ignore certain goals. This occurs most often in cases where individuals who are involved with the program are not given a chance to contribute to the enumeration of goals and can result in a lack of stakeholder buy-in (Newcomer 2007; Posner and Fantone 2007; Melkers and Willoughby 2005; Gilmour and Lewis 2006a, b; Talbot 2005; Behn 2002). Finally, misalignment of goals at institutional levels can lead to a mismatch of incentives (Behn 2003; GAO 2005, 2012; Pollitt 2006). For example, employee goals may be misaligned with program goals, resulting in an organization culture that resists performance information (Durant 1999; Dubnick 2005; Newcomer 2007).

The second step in developing policy knowledge is the identification and collection of data. Data falls into four broad categories. The first of these is related to inputs, or the resources utilized by programs and other government activities. Second, data may be gathered about processes used to direct those inputs. Third, data may reflect outputs, or the products or services being produced. Finally, data may be gathered about outcomes, or the measures related to the overall goals of the activity.

The quality of policy knowledge is directly related to the rigor, accuracy, and validity of the methods used to identify, generate, and collect data (GAO 2009; Weiss 1977; Lindblom and Cohen 1979). The accuracy of data refers to the extent to which measurement or record of observing the data is correct. For example, if a program is producing widgets on an assembly line, and production amounts are estimated by multiplying a standard number of units produced per hour by the number of hours the production line was active, the resulting figure may not be accurate if, for example, one employee left the line and caused the production rate to fluctuate over the period. In this case, a count of widgets produced as they come off the line could be a more accurate source of data. The validity of

information refers to the extent to which the subject of measurement or observation is represented by the data collected. For example, knowing the number of students in one class is not likely a valid measure for the quality of their teacher. Additionally, multiple measures may arguably be valid in some circumstances. For example, one person may believe that the most valid measure of intelligence is a student's grade, whereas another may believe that standardized test scores are more valid. Some may utilize both measures, whereas others may prefer one over the other. Ultimately, however, if a measure is not valid, the data it produces will not be useful in guiding decisions.

The coverage of policy knowledge derives from the degree to which the data are complete and representative of the relevant population, program, or activity being studied. For example, if a decision maker desires information about the average emissions of vehicles in a given area, a sample of hybrid cars would not completely represent the population. To be complete, data would need to be collected on a census or a statistically representative sample of vehicles. A desire to know the national average of automobile emissions would require the entire national population of automobiles of all types to be represented. However, if a decision maker only wanted to compare hybrids and diesel engines, only those two types of automobile would need to be represented for the information to be considered complete. In some cases, multiple sources of data may be required in order to be sufficiently complete and representative.

The ability of data to reflect sufficient quality and appropriate coverage is largely dependent on the rigor of the methods used to collect them. Data collection methods refer to the approach used to elicit and capture observations and measures. Methodology, including identification of the relevant population, what instruments are used to record data, where and when data are gathered, and how data are observed, has a direct impact on the data produced. For example, if a statistical sample is developed using a population frame that is not representative of the population of interest, results may be biased. If a survey is administered to a group of individuals in person, it may affect whether respondents are willing to answer certain sensitive questions. If data are gathered during rush hour in Los Angeles, they cannot represent daily driving habits in California. For decision makers to utilize policy knowledge, they must believe that the methodological design is sufficient and appropriate to meet their needs.

Several limitations of evidence-based reforms emerge during data gathering and measurement activities. First and foremost, it is resource-intensive and technically challenging to gather data that are relevant to stated goals, sufficient and appropriate for understanding, and accurate, valid, complete, and reliable for interpretation of results (Cyert and March 1963). Limitations in data quality are often cited as reasons that data, information, and evidence are ignored (Dubnick 2005; ter Bogt 2004; Julnes and Holzer 2001). This can be exacerbated by reliance on administrative data, since these data are generally produced for other purposes, and therefore often do not meet the standards of data quality outlined above when applied to new areas of inquiry (Frederickson and Frederickson 2006). Similarly, when federal efforts rely on states for data, there are additional difficulties because states may not have standardized, comparable data that can be used with data from other states to produce nationally representative coverage (Majone 1989).

The relationships between policy knowledge—as defined by the coverage and rigor of data, information, and evidence—and the range of decision making purposes make it abundantly clear that a single source of data could not possibly serve every need of every decision maker for every type of decision. It is critical that evidence-based reforms consider these relationships as they produce policy knowledge, promote its use, and assess its impact. If they fail to do so, the seemingly ad hoc and adversarial use of policy knowledge is likely to continue.

After the enumeration of goals and collection of data, analysis and interpretation of the data are required to produce information and evidence about a given activity. In other words, the data that have been gathered must be processed into a form that directly relates to the goals that have been established. In this process, producers of policy knowledge need to ensure they are meeting their stated goals and define what it means to be "achieving" those goals. They can conduct analysis that reflects program activities and changes in measures of these goals, or they can establish evaluative criteria to generate and test evidence of a relationship among these measures (Radin 2006).

Analysis and interpretation of data create additional considerations for the quality and coverage of policy knowledge. The information and evidence produced in this step cannot be of higher quality or broader coverage than the underlying data, but problematic analysis may obfuscate this fact. For example, some producers of policy knowledge may make generalizations from a single case or non-representative collection of cases to

an entire population, such as identifying a malfunction in one vehicle as a cause for recalling an entire fleet. They may also be less drastic, including technical manipulations of statistical significance (referred to colloquially as "p-hacking"). In addition, evidence-based reforms may focus on priorities that are driven by political factors, which may result in incomplete, misleading, subjective, or biased results (Weiss 1977; Moynihan 2006; Moynihan 2008). Beyond these problematic analytical practices, producers of policy knowledge may also be subject to heuristic learning processes, where data, information, and evidence processed by an individual are subconsciously filtered through his or her existing knowledge and made to reflect his or her own subjective understandings (Simon 1947). As they produce information and evidence, it will necessarily include some degree of this subjective judgment. Others argue that such judgment becomes a critical part of turning policy knowledge into compelling and useful narratives (Stone 1997). The usefulness of these narratives may vary depending on existing heuristics and chosen points of comparison (Webeck and Nicholson-Crotty 2019), as well as views of the individual providing the policy knowledge (Perna et al. 2019).

After completing the first three steps, producers of policy knowledge need to distill their findings into a format that can be used by decision makers (Weiss 1977; Lindblom and Cohen 1979). The method of communicating policy knowledge impacts whether it will be used. If presented too simplistically, it may be rejected for a lack of demonstrated nuance and understanding. However, if presented in too technical a format, it may be rejected for being convoluted or hard to understand (Behn 2002; Burke and Costello 2005; Pollitt 2006). As such, tailoring the communication of the policy knowledge to the specific audience is a critical step in promoting its use.

An important component of this step is to explicitly and clearly communicate the limitations and subjective judgments that may occur in the first three steps, as decision makers may disagree about the relative weight of individual goals, the quality and coverage of data, the value of analytical techniques, the appropriateness of selected evaluative criteria, or the neutrality of the resulting policy knowledge (Newcomer 2007; Julnes and Holzer 2001; Posner and Fantone 2007; Pollitt 2006; Gilmour and Lewis 2006a; Talbot 2005; Melkers and Willoughby 2005). Clear communication of the range of decisions made to develop policy knowledge, and the related strengths and limitations of data, information, and evidence

can enable decision makers to better integrate this knowledge into their deliberations.

While producers of policy knowledge can work to communicate their data, information, and evidence in a way that is useful to decision makers, this knowledge often arrives indirectly. Specifically, direct communication of policy knowledge is supplemented by interpretation, combining sources, and communication among stakeholders and decision makers. In *The Dynamics of Performance Management*, Donald Moynihan (2008) explains that policy knowledge diffuses among decision makers and other stakeholders through a process of interactive dialog. This interactive dialog defines the use of data, information, and evidence as a process of presentation, exchange, and interpretation. Many scholars and advocates have undertaken efforts to promote the dissemination and use of policy knowledge through direct involvement in these dialogs. These efforts have sometimes coalesced into interest groups that advocate directly to decision makers and indirectly through media and other communications outlets for increased use of specific policy knowledge. This may lead to shared understanding, greater agreement, and coordinated action among stakeholders whose interpretations lead to consistent conclusions. However, it may also lead to little more than a restatement of already expressed conflicting opinions among competing preferences, with limited willingness to accept alternative explanations.

The considerations and limitations identified in the four steps outlined above have a major influence on whether and how the resulting data, information, and evidence are used, and the use of this knowledge is often a primary factor in whether evidence-based reforms are seen as successful. However, government decisions on what to do, how to do it, and how much it should cost are only partially dependent on policy knowledge (Shulock 1999). Information overload, misleading heuristics, and selective use of policy knowledge can prevent utilization by decision makers (Moynihan 2006, 2008; Tversky and Kahneman 1986; Van de Walle and Boivard 2007). To the extent that these challenges are overcome, there is also significant variation in the ways that decision makers use policy knowledge and the degree of impact they have on the decision outcome.

Direct and Indirect Uses of Data, Information, and Evidence

Direct and explicit links between policy knowledge and decisions are foundational to evidence-based reforms that promise to be a panacea for decision makers. However, it is often difficult to determine whether and how specific data, information, and evidence contributed to a decision. For example, information on the unemployment rate could represent a warning about economic troubles. When addressing that problem, such as by establishing a new jobs program, decision makers may not explicitly cite the unemployment rate, or any other policy knowledge. This represents a second limitation of evidence-based reforms: even when policy knowledge influences a decision, it may not be a clear and deterministic factor. Further, Moynihan (2008) has demonstrated that multiple and conflicting goals among decision makers may result in compromises and trade-offs that further obscure whether and how data, information, and evidence are used. The policy dialog enables them to consider and prioritize multiple sources and types of policy knowledge before making a decision, but also makes tracking this use more difficult.

Consistent with this insight, Majone (1989) has argued that information should not be viewed as deterministic to decisions, but as part of a persuasion that is necessary to achieve desired policy outcomes. Instead, government action is driven by a series of decisions, each of which is the result of the data, information, and evidence available to and utilized by decision makers. This process reflects the personal preferences of decision makers and the preferences of other actors that influence decisions, understanding of decision making rules, expectations about available resources, and specific substantive information about a particular choice (Lindblom and Cohen 1979). Reflective of these varied degrees of persuasion and influence, evidence-based governance reflects variation in how much policy knowledge is considered, and how central it is to decision outcomes. Specifically, decision makers may eschew it all together, may rely on a single source or selection of relevant sources, or may comprehensively consider the full range of relevant policy knowledge before acting. Similarly, the data, information, and evidence they use may only provide context for the decision, may provide insight into the deliberation of decision options, or may be fully deterministic of the outcome.

Program Design and the Production of Policy Knowledge

James Q. Wilson's (1989) typology of bureaucratic organizations provides relevant insights into factors that can influence production and use of policy knowledge. Specifically, producing certain forms of data, information, and evidence can be more or less difficult, depending on the characteristics of the specific program being assessed. In particular, Wilson identifies differences in the ability to measure program activities and outputs related to government efforts. Where both can be easily measured, the bureaucratic type is production; where only action or only outcomes can be measured, the types are procedural and craft, respectively; when neither activity nor outcome can be measured, the type is coping. These four types will be described in more detail below, along with some expectations for the production of policy knowledge and the purposes it may serve.

Production organizations are defined as those where both process and outcomes can be measured. In these organizations, Wilson argues, managers can potentially design a system of rules that will lead to an optimal outcome. However, one impact of this may be that those processes and outcomes that are easily measured could lead program managers to de-emphasize work on processes and outcomes that are more difficult to measure. When examining the production of policy knowledge, we might expect to see production organizations whose primary goals are clear and well defined and whose activities and outcomes are observable and measurable. In production organizations, secondary goals may be split into two categories. The first of these is goals whose activities and outcomes are also easily measurable, while the second is those with activities or outcomes that are harder to measure. Production organizations may elevate the importance of measurable secondary goals over those of secondary goals that are difficult to measure. In production organizations, data on inputs, outputs, and outcomes are expected to be readily available and easily converted into relevant information and evidence. Similarly, the tendency for production organizations to limit their focus to these areas suggests that they will be readily able to use this policy knowledge to guide outcomes.

Procedural organizations are those where process of the primary goal can be observed but outcomes cannot. In these organizations, managerial focus is on professional expertise. The result of this is that standard operating procedures and rule-following become more important than

results. In a decision making context, we would expect policy knowledge to be developed around process goals and data collection to focus primarily on inputs, secondarily on any measurable outputs, and generally not on outcomes. When required to produce outcome data, procedural organizations may attempt to focus on the outcomes of secondary goals, which may be easier to measure. This focus on process is expected to result in a focus on framing policy issues, with limited use to influence decision outcomes.

The opposite of this occurs in craft organizations, where the process needed to reach the primary goal cannot be measured but outcomes can. The defining management style of craft organizations is discretion for operators to perform according to professional norms. In a decision making context, craft organizations may be able to produce output and outcome data related to their primary goal, but may have difficulty providing data on inputs and related measures of efficiency. In public organizations, Wilson argues, craft organizations tend toward establishing measures of process that are not directly related to the primary outcome as a means of limiting the opportunity for abuse of discretion. Policy knowledge primarily focused on outcomes is expected to be used to influence bargaining among stakeholders.

The last bureaucratic type, coping organizations, can measure neither the process of their operators nor the outcome related to their primary goal. Managers in these organizations must attempt to recruit staff with professional expertise, though they cannot be sure which expertise is appropriate, and must try to establish procedures that will lead to good outcomes without knowing what those outcomes should be. Aside from this, they must intervene during crises, even though they do not know whether the crises are real or persistent. Cohen et al. (1972), in developing their garbage can model of decision making, called this an "organized anarchy." In a decision making context, we do not expect to see valid and reliable measures of inputs or the primary outcome, but instead to see output measures of secondary goals, potentially viewed as intermediate outcomes related to the primary goals. In this system, the program may alternate its focus among secondary goals over time, as none are ultimately able to fully present progress toward the final outcome. Given these difficulties, any related policy knowledge is expected to be used to frame decisions and influence bargaining among decision makers.

Wilson's typology of bureaucratic organizations can be used to understand the relationship between program design and policy knowledge. In

order to examine this relationship, the two cases in this book reflect the two ends of this model. Specifically, student loans are most similar to a production program, where measuring is relatively easy, while Title I most closely resembles a coping program, where both process and outcome are difficult to measure.

This chapter has shown that evidence-based reforms have often achieved uneven and limited success as a result of overpromising that the policy knowledge they produce will be a panacea for decision making. Chapter 3 shows that these reforms have also been limited by lack of attention to decision making processes and political factors. Evidence-based reforms, in addition to their simplified treatment of policy knowledge, have tended to assume that decisions are either political or rational, treating policy knowledge as a substitute for, rather than supplement to, politics in decision making. Chapter 3 considers how the literature on political decision making has reinforced acceptance of this proverb and offers an alternative model for considering the competing and complementary roles of policy knowledge and political factors in the decision making process.

References

Alkin, M. C., & King, J. A. (2017). Definitions of evaluation use and misuse, evaluation influence, and factors affecting use. *American Journal of Evaluation, 38*(3), 434–450.

Barnow, B. (1992). The effects of performance standards on state and local programs. In C. Manski & I. Garfinkel (Eds.), *Evaluating welfare and training programs* (pp. 277–309). Cambridge, MA: Harvard University Press.

Behn, R. D. (2002). The psychological barriers to performance management: Or why isn't everyone jumping on the performance-management bandwagon? *Public Performance & Management Review, 26*(1), 5–25.

Behn, R. D. (2003). Why measure performance? Different purposes require different measures. *Public Administration Review, 63*(5), 586–606.

Bouckaert, G., & Halligan, J. (2008). *Managing performance: International comparisons.* London: Routledge.

Burke, B. F., & Costello, B. C. (2005). The human side of managing for results. *American Review of Public Administration, 35*(3), 270–286.

Cohen, M. D., March, J. G., & Olsen, J. P. (1972). A garbage can model of organizational choice. *Administrative Science Quarterly, 17*(1), 1–25.

Cyert, R. M., & March, J. G. (1963). *A behavioral theory of the firm*. Englewood Cliffs, NJ: Prentice-Hall.

Davis, O. A., Dempster, M. A. H., & Wildavsky, A. (1966). A theory of the budgetary process. *American Political Science Review, 60*(3), 529–547.

Dubnick, M. (2005). Accountability and the promise of performance: In search of the mechanisms. *Public Performance & Management Review, 28*(3), 376–417.

Durant, R. F. (1999). The political economy of results-oriented management in the "neoadministrative state": Lessons from the MCDHHS experience. *American Review of Public Administration, 29*(4), 307–331.

Feldman, M. S. (1989). *Order without design: Information production and policy making*. Stanford, CA: Stanford University Press.

Frederickson, D. G., & Frederickson, H. G. (2006). *Measuring the performance of the hollow state*. Washington, DC: Georgetown University Press.

Gilmour, J. B., & Lewis, D. E. (2006a). Does performance budgeting work? An examination of OMB's PART scores. *Public Administration Review, 66*(5), 742–752.

Gilmour, J. B., & Lewis, D. E. (2006b). Assessing performance assessment for budgeting: The influence of politics, performance, and program size in FY 2005. *Journal of Public Administration Research and Theory, 16*(2), 169–186.

Government Accountability Office (GAO). (2005). *Performance budgeting: States' experiences can inform federal efforts, GAO-05–215*. Washington, DC: GAO.

Government Accountability Office (GAO). (2009). *Assessing the reliability of computer-processed data, GAO-09-680G*. Washington, DC: GAO.

Government Accountability Office (GAO). (2012). *A guide for using the GPRA modernization act to help inform congressional decision making, GAO-12-621SP*. Washington, DC: GAO.

Hatry, H. P. (2006). *Performance measurement: Getting results* (2nd ed.). Washington, DC: Urban Institute.

Joyce, P. G. (2005). Linking performance and budgeting: Opportunities in the federal budget process. In J. M. Kamensky & A. Morales (Eds.), *Managing for results 2005* (pp. 83–140). Rowman & Littlefield Publishers, Incorporated.

Julnes, P. D. L., & Holzer, M. (2001). Promoting the utilization of performance measures in public organizations: An empirical study of factors affecting adoption and implementation. *Public Administration Review, 61*(6), 693–708.

Lindblom, C. E., & Cohen, D. K. (1979). *Usable knowledge: Social science and social problem solving* (Vol. 21). New Haven, CT: Yale University Press.

Majone, G. (1989). *Evidence, argument, and persuasion in the policy process*. New Haven, CT: Yale University Press.

Melkers, J., & Willoughby, K. (2005). Models of performance-measurement use in local governments: Understanding budgeting, communication, and lasting effects. *Public Administration Review, 65*(2), 180–190.

Moynihan, D. P. (2006). Managing for results in state government: Evaluating a decade of reform. *Public Administration Review, 66*(1), 77–89.

Moynihan, D. P. (2008). *The dynamics of performance management: Constructing information and reform*. Washington, DC: Georgetown University Press.

National Academy of Public Administrators Social Equity Committee. (2000, November). *Standing panel on social equity in governance: Issue paper and work plan*. Washington, DC: NAPA.

Newcomer, K. E. (2007). How does program performance assessment affect program management in the federal government? *Public Performance and Management Review, 30*(3), 332–350.

Perna, L. W., Orosz, K., & Kent, D. C. (2019). The role and contribution of academic researchers in congressional hearings: A critical discourse analysis. *American Educational Research Journal, 56*(1), 111–145.

Pollitt, C. (2006). Performance management in practice: A comparative study of executive agencies. *Journal of Public Administration Research and Theory, 16*(1), 25–44.

Posner, P. L., & Fantone, D. M. (2007). Assessing federal program performance: Observations on the US Office of Management and Budget's program assessment rating tool and its use in the budget process. *Public Performance & Management Review, 30*(3), 351–368.

Radin, B. A. (2006). *Challenging the performance movement*. Washington, DC: Georgetown University Press.

Roberts, A. (2019). Bridging levels of public administration: How macro shapes meso and micro. *Administration & Society*. https://doi.org/10.1177/0095399719877160.

Shulock, N. (1999). The paradox of policy analysis: If it is not used, why do we produce so much of it? *Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management, 18*(2), 226–244.

Simon, H. A. (1947). *Administrative behavior*. New York: Free Press.

Stone, D. A. (1997). *Policy paradox: The art of political decision making* (Vol. 13). New York: W. W. Norton.

Talbot, C. (2005). Performance management. In E. Ferlie, L. Lynn Jr., & C. Pollitt (Eds.), *The oxford handbook of public management* (pp. 491–517). Oxford: Oxford University Press.

ter Bogt, H. (2004). Politicians in search of performance information? Survey research on Dutch Aldermen's use of performance information. *Financial Accountability and Management, 20*(3), 221–252.

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *The Journal of Business, 59*, 251–278.

Van de Walle, S., & Boivard, T. (2007). *Making better use of information to drive improvements in local public services. A report for the Audit Commission*. Birmingham: Institute of Local Government Studies, University of Birmingham.

van Dooren, W. (2004). Supply and demand of policy indicators: A cross-sectoral comparison. *Public Management Review, 6*(4), 511–530.

Webeck, S., & Nicholson-Crotty, S. (2019). How historical and social comparisons influence interpretations of performance information. *International Public Management Journal*, 1–24.

Weiss, C. H. (Ed.). (1977). *Using social research in public policy making* (Vol. 11). Lexington, MA: Lexington Books.

Wildavsky, A. B. (1974). *Budgeting: A comparative theory of budgetary processes.* Boston: Little, Brown.

Wilson, J. Q. (1989). *Bureaucracy: What government agencies do and why they do it* (2nd ed.). New York: Basic Books.