

Mapping the Institutionalization of Evaluation in the U.S. Federal Government

A Dissertation submitted in partial fulfillment of the requirements for the degree of  
Doctor of Philosophy at George Mason University

by

Esther C. Nolton  
Master of Education  
University of Virginia, 2011  
Bachelor of Science  
George Mason University, 2010

Directors: Marvin G. Powell, Assistant Professor & Rodney K. Hopson, Professor  
College of Education and Human Development

Summer Semester 2020  
George Mason University  
Fairfax, VA



THIS WORK IS LICENSED UNDER A CREATIVE COMMONS  
ATTRIBUTION-NODERIVS 3.0 UNPORTED LICENSE.

## **Dedication**

This dissertation is dedicated to my father, Robert S. Chou, who shared with me his love for studying and documenting history through ethnography and instilled in me a commitment to service and scholarship.

## Acknowledgements

First and foremost, I would like to thank my friends and family for extending me an enormous amount of patience and grace during this long journey. Thank you for encouraging me, praying for me, and supporting me. Most of all, thank you for not abandoning me while I was less present in your lives.

Brian, thank you for loving me unconditionally through this phase of my career and for never letting me quit. None of this would have been possible without your help in managing our lives and effort to respect my time. Words cannot express how much you mean to me. I love you and cannot wait for the next chapter of our life!

Thank you to my amazing dissertation committee members, Drs. Marvin Powell, Rodney Hopson, Divya Varier, and Valerie Caracelli. All of you spent countless hours caring deeply for both my scholarship and my well-being. Each of you have been an incredible blessing and have made this process as pleasant as it could possibly be. I strive to be as good a mentor to my future students as you have been to me.

Thank you to Dr. Cynthia Phillips, my beloved GEDI mentor, who was the original spark that led me to this remarkable topic. Without the experience you provided at such an opportune time, this topic would never have come to fruition. Thank you for pushing me and always reminding me not to overthink this “long paper”.

Thank you to my other thought partners, especially Dr. Janine Wedel, who taught, inspired, and challenged me at various stages throughout my studies. I have encountered so many incredible people through this journey, who motivated me to do this topic justice. Thank you to fellow doctoral students for holding me accountable and for lifting me up throughout this process.

Last, but certainly not least, thank you to Mommy, Daddy, and Denny for supporting me through every stage of my life. Thank you for all of the lessons, beliefs, and habits you instilled in me. Thank you for everything you all have done to make sure I could have every joy in life.

To God be ALL of the glory for all of the goodness that I have received.

## Table of Contents

	Page
List of Tables.....	viii
List of Figures .....	ix
List of Abbreviations .....	x
Abstract.....	xi
Background.....	1
Introducing a Conceptual Framework and Method of Mapping Influences on the Process of Institutionalizing Evaluation in the U.S. Federal Government.....	9
Evaluation in the U.S. Federal Government.....	10
Building professional legitimacy.....	11
Evaluation for evidence-based policymaking. ....	14
The Foundations for Evidence-Based Policymaking Act.....	15
Conceptualizing the Diffusion of Value in Evaluation.....	19
The manifestation of value in evaluation.....	19
Developing the conceptual framework.....	21
Complexity of the U.S. Federal government. ....	22
Comparison of existing frameworks.....	23
Proposing a new conceptual framework.....	27
The Mapping Method.....	29
Existing methods of mapping.....	29
A social anthropological approach. ....	31
Possible limitations.....	37
Conclusion.....	38
References .....	40
The Diffusion of Value in Evaluation by the U.S. Federal Government Through the <i>Analytical Perspectives</i> of the U.S. President’s Budget.....	60
Increasing Value in Evaluation.....	60
Conceptualizing the process.....	60
Mapping the process.....	64
Methods.....	66

Analyzing content and discourse.....	66
Document selection. ....	67
Coding and analytic procedures. ....	70
Coding for prevalence. ....	70
Coding for evidence of institutionalization. ....	72
Establishing rigor and validity. ....	73
Results and Interpretations.....	79
Document metadata. ....	79
Prevalent words. ....	83
Evidence of institutionalization.....	92
Discussion and Implications.....	98
Realizing the process. ....	98
Conveying persuasive change messages.....	101
Emphasizing executive buy-in. ....	102
Creating principles and standards.....	103
Establishing centralized evaluation units.....	104
Limitations and future research. ....	105
Conclusion.....	107
References.....	109
Mapping Influences on Institutionalizing Evaluation in the U.S. Federal Government .	121
Current understanding of the process. ....	122
What lies beyond, between, and beneath? .....	123
Methods.....	125
Study design. ....	125
Participant selection.....	126
Interview procedures. ....	127
Coding and analytic procedures. ....	128
Coding for themes. ....	129
Coding for chronology. ....	131
Code cleaning and sorting. ....	132
Establishing rigor and validity. ....	135
Results and Interpretations.....	138

Actors, agendas, and assumptions. ....	138
Origins of federal evaluation. ....	141
Early setbacks and missteps for evaluation. ....	142
Demand for credible evidence. ....	144
Proliferation of robust evaluation methods.....	145
Prioritization of experimentation. ....	146
Addressing skepticism about evaluation. ....	147
Diffusing value in evaluation.....	149
Context, circumstances, and facilitators of institutionalization. ....	151
Institutional ecosystem and process. ....	154
Discussion and Implications.....	162
Cultivating a culture of evaluation. ....	162
Fostering leaders and advocates. ....	164
Equipping evaluators and evaluation units. ....	166
Establishing an evidence infrastructure. ....	169
Limitations and future research.....	171
Conclusion.....	173
References .....	175
Appendix A.....	192
Appendix B .....	194
References.....	199

## List of Tables

Table	Page
<i>Table 1.</i> Features and Relevant Elements of Existing Conceptual Frameworks .....	26
<i>Table 2.</i> Comparison Among Various Mapping Methods.....	35
<i>Table 3.</i> Metadata for the Analytical Perspectives Documents .....	69
<i>Table 4.</i> Counts of Word Prevalence for Searched Terms .....	71
<i>Table 5.</i> Selected Grounded Theory Coding Examples for Institutionalization .....	77
<i>Table 6.</i> Metadata for Analytical Perspectives Evaluation and Evidence Chapters .....	82
<i>Table 7.</i> Mostly Highly Prevalent Words in the Evaluation Chapters .....	88
<i>Table 8.</i> Aggregated Counts of Most Highly Prevalent Words .....	92
<i>Table 9.</i> Mapping Data to Levels of Institutionalization .....	94
<i>Table 10.</i> Demographic Background of Interview Participants.....	127



## List of Figures

Figure	Page
<i>Figure 1.</i> Relationship Among Evaluation Use, Influence, and Capacity.....	20
<i>Figure 2.</i> Conceptual Framework for Diffusing Value in Evaluation.....	28
<i>Figure 3.</i> Revisiting the Conceptual Framework for Diffusing Value in Evaluation .....	62
<i>Figure 4.</i> Stages of Institutionalization Framework.....	64
<i>Figure 5.</i> Example of Memo Documented in ATLAS.ti.....	76
<i>Figure 6.</i> Line Graph of Word Prevalence for Searched Terms .....	85
<i>Figure 7.</i> Snapshot of Word Cloud Function in ATLAS.ti .....	86
<i>Figure 8.</i> Snapshot of Word Cloud Produced by ATLAS.ti.....	87
<i>Figure 9.</i> Word Cloud of Aggregated Most Prevalent Words .....	91
<i>Figure 10.</i> Snapshot of Code-Documents Table Function in ATLAS.ti .....	93
<i>Figure 11.</i> Modified Institutionalization Framework Aligned to Types of Influences ...	100
<i>Figure 12.</i> Modified Institutionalization Framework Aligned to Types of Influences ...	124
<i>Figure 13.</i> Code Grouping and Sorting Procedure.....	131
<i>Figure 14.</i> Example of Code Sorting and Bucketing Process.....	134
<i>Figure 15.</i> Mapping a Chronology of Significant Events .....	140
<i>Figure 16.</i> Common Barriers and Facilitators of Institutionalizing Evaluation .....	153
<i>Figure 17.</i> Mapping the Process of Influence on Diffusing Value in Evaluation.....	156
<i>Figure 18.</i> Manifestations of Influences Toward Institutionalization.....	158

## List of Abbreviations

American Evaluation Association .....	AEA
Analytical Perspectives .....	AP
Chief Evaluation Office(r) .....	CEO
Chief Financial Office(r) .....	CFO
Central Policy Review Staff.....	CPRS
Commission on Evidence-Based Policymaking .....	CEP
U.S. Department of Education .....	ED
U.S. Department of Labor.....	DoL
Evaluation Capacity Building .....	ECB
Evaluation Policy Task Force .....	EPTF
Fiscal Year .....	FY
U.S. General Accounting Office/Government Accountability Office .....	GAO
Government Performance and Results Act.....	GPRA
Institute of Education Science.....	IES
National Academies of Sciences, Engineering, and Medicine .....	NAS
U.S. Office of Management and Budget.....	OMB
U.S. Office of Personnel Management .....	OPM
Performance Assessment Rating Tool.....	PART
Paperwork Reduction Act.....	PRA
Qualitative Data Analysis/Analyses .....	QDA

## Abstract

### MAPPING THE INSTITUTIONALIZATION OF EVALUATION IN THE U.S. FEDERAL GOVERNMENT

Esther C. Nolton, Ph.D.

George Mason University, 2020

Dissertation Directors: Drs. Marvin G. Powell, Assistant Professor & Rodney K. Hopson, Professor

Despite marked growth in evaluation use and capacity in the U.S. Federal Government over the last 50 years, the levers that have driven institutional change have not previously been identified. Knowledge of effective and ineffective mechanisms guides efficient approaches to complex organizational change. This dissertation adopts a novel mapping method to study the process and factors that have led to the diffusion of value in evaluation in the U.S. Federal Government. It is an ethnographic study, grounded in social anthropological techniques, that combines interview data from 15 evaluation experts with content and discourse analyses of the U.S. President's Budget *Analytical Perspectives* from fiscal years 1996-2020 to capture shifts in the prioritization of evaluation over time. Thematic analyses of interview data revealed several actors, events, or policies of note that influenced the (de)institutionalization of evaluation. A content and discourse analysis of the *Analytical Perspectives* demonstrated a progression toward

institutionalization by analyzing the prevalence of related key terms and shifts in language. It was discovered that beliefs and actions toward institutionalization are manifested through a progression from cognitive, motivational, behavioral, and structural influences, and that the types of factors at play can impact processes across any level of a nested ecosystem. This study offers unique insight to the trajectory of evaluation in the U.S. Federal Government. Findings may be instrumental to policymakers and organizational leaders who continue to work toward institutionalizing evaluation in the federal government or other complex organizations.

## **Background**

Evaluation is a relatively new field of systematic inquiry that uses formal methodologies to make judgements regarding the values, merit, and worth of social programs and policies (Cousins et al., 2014; Rossi et al., 2019; Scriven, 1967; Trochim, 1998). Specifically, since the 1960s, federal (government) evaluation plays a large role in demonstrating the effectiveness of government investments in these programs and policies to impact outcomes (Lemire, Fierro, et al., 2018; Maynard et al., 2016). As such, our society perpetuates the belief that taxpayers are owed evidence that government officials are making responsible policy decisions and are held accountable to the consequences of those decisions (Clarke, 2005). Evaluations generate the necessary evidence that can be instrumental in informing program improvement and policymaking decisions (Grob, 2007; Maynard et al., 2016).

The federal evaluation market has grown dramatically in the last 60 years, which has led to increasing use and capacity for evaluation (Lemire, Nielsen, et al., 2018; Mark et al., 2009). However, the uptake of evaluation has been disproportionately distributed and there have been varying mechanisms for building value in evaluation across the government. The evolution of evaluation utilization has not been linear (Johnson, 1998)—in fact, the realistic complexity of building and sustaining organizational evaluation capacity includes a myriad of factors that promote or resist valuing evidence

(Preskill & Boyle, 2008). As power is wielded by the hands of various actors in the policy arena, the levers of influence on institutionalizing evaluation continue to ebb and flow in the U.S. Federal Government and our understanding of them remains unclear (Stone, 2008; Wedel, 2009). Studying and documenting the effective and ineffective drivers of progress can be useful in informing future policies and strategies for institutional change. Specifically, we need to investigate the parties that are involved, their motivations, the breadth and depth of the ecosystem, the surrounding powers or agendas, the barriers and facilitators, and the institutional structure (Wedel, 2019).

This dissertation—collectively titled, *Mapping the Institutionalization of Evaluation in the U.S. Federal Government*—is a collection of three manuscripts which aim to map the process and influential factors for institutionalizing evaluation in the U.S. Federal Government. There are many methods available that could be suitable in studying the breadth of factors at play over time which influence the institutionalization of evaluation. Beyond the method of inquiry, it is necessary to conceptualize the context and process of institutionalizing organizational change to fully understand the complexity of this framework. Therefore, the aim of the first manuscript, *Introducing a Conceptual Framework and Method of Mapping Influences on the Process of Institutionalizing Evaluation in the U.S. Federal Government*, was to: 1) provide an overview of the landscape of evaluation in the Federal government; 2) present a conceptual framework for building institutional value in evaluation; and 3) introduce a novel ethnographic method of studying the process of institutionalization.

First, a distinction is made between the value *of* evaluation (i.e., a statement of worth) and value *in* evaluation (i.e., belief and investment in that worth). The first manuscript then proposes a model to conceptualize the process of diffusing value *in* evaluation. This framework accounts for the complexity of the Federal government as a nested ecosystem with internal and external layers that influence its organizational structure and culture. Central to this process are many different factors that might influence how value in evaluation is generated and distributed. Borrowed from the Mark and Henry (2004) model of evaluation influence, these factors also have different types of influences: cognitive (thinking and feeling), motivational (valuing), behavioral (doing), and structural (being). Although it is assumed that this framework appropriately portrays the process of diffusing value in evaluation, it is not clear which mechanisms have been used by the U.S. Federal Government to achieve this process.

Colyvas and Jonsson (2011) separates diffusion (i.e., ubiquity) and institutionalization (i.e., legitimacy) whereby, “diffusion emphasizes the pace and pattern of the object that spreads [and] institutionalization underscores depth and durability,” (p. 45). This process occurs by transforming an exterior concept through microlevel actions toward an alignment with subjective understandings and changes in macrolevel structures (Li, 2017; Zucker, 1977). Two other models mimic the proposed cumulative process. Armenakis and Harris (2002) describe institutional change as a three-part model beginning with a *readiness* for the uptake of a new idea followed by *adoption* and *institutionalization* of a new idea. This model relates stages of institutionalization in organizational maturity and capacity for institutional change. Furthermore, Meyer et al.

(2018) explained that text could be used to express varying degrees of institutionalization for which they defined as eight levels across three discrete phases: pre-institutionalization (i.e., similar to Armenakis & Harris' stage of readiness), semi-institutionalization (i.e., adoption), and full institutionalization.

One way in which the government demonstrates value in rigorous evidence is by investing resources toward greater capacity for evaluation activities (Lemire, Nielsen, et al., 2018). In preparation for each upcoming fiscal year (FY), the U.S. Office of Management and Budget (OMB) publishes the President's Budget which projects funding allocations across the entire Federal government. Research shows an increase in support for evaluation by observing Federal spending on evaluation over time (Kinarsky, 2018; Lemire, Fierro, et al., 2018; Rizvi, 2004); however, there is more to the story behind the numbers. Each President's Budget includes supplementary documents that append elaborations, tables, and a detailed list of appropriations. The narratives are compiled in a separate document titled *Analytical Perspective*, which further explicate the rationale and analyses that informed the changes to the proposed budget. Therefore, the second manuscript of this dissertation analyzed the content and discourse of *The Diffusion of Value in Evaluation by the U.S. Federal Government Through the Analytical Perspectives of the U.S. President's Budget*.

In summary, the nature and content of the *Analytical Perspectives* greatly changed over the last 25 years. The inaugural chapter dedicated to evaluation in FY 2011 led to a pivotal change in an increase in the overall prevalence of evaluation. In addition to becoming more prevalent, the chapters grew in content and gradually elevated in chapter



order over time. Some shifts in organization and content could be attributed to the transition of Presidential Administration; implementation of new initiatives and policies; or shifting focus between evaluation, performance, evidence, and administrative data. The language in the evaluation chapters of the *Analytical Perspectives* suggest that the Federal government has moved through stages of diffusing value in evaluation toward fully institutionalizing evaluation.

However, the language in these documents are limited to what the government wanted to convey in those narratives. Consequently, the *Analytical Perspectives* reflect the desired institutional change through political rhetoric, but it is not clear if their efforts to diffuse value in evaluation are effectively translated into reality and what other internal and external factors were effective or ineffective in influencing the institutionalization of evaluation in the Federal government. Despite its role as a central government document that can mold and align belief systems, there are other factors at play that must be acknowledged. These additional factors can be identified through learning about the lived experiences of individuals who observed changes in evaluation from their positions in and around the Federal government.

Therefore, the full mapping method based was employed in the third manuscript titled, *Mapping Influences on Institutionalizing Evaluation in the U.S. Federal Government*. In addition to analyzing the *Analytical Perspectives* documents, a total of 15 semi-structured interviews were conducted with key informants who could account for the changes they observed in federal evaluation over time to ground developing theories

of this process (Charmaz et al., 2018; Patton, 2015b; Zilber, 2002). The following six questions guided the empirical methods:

1. Who have been the **parties** (i.e., players, social networks, and organizations) involved, both formally and informally, in federal evaluation?
2. What have been the parties' respective interests, **agendas**, incentives, goals, motivations, operating assumptions, and expectations with regard to federal evaluation?
3. Who have been the parties' **sponsors** (i.e., funders or providers of other indispensable resources) and what are their interests, agendas, incentives, goals, motivations, operating assumptions, and expectations with regards to federal evaluation activities?
4. What has been the larger context(s)/circumstances of **power** and **resources** in which the parties are embedded? How have features of the larger context(s)/circumstances constrain, enable, or otherwise affect the parties involved in federal evaluation efforts?
5. To what extent has the **structure** that facilitates federal evaluation activities or efforts institutionalized (or becoming so) even if it remains informal/unincorporated and the players change over time?
6. What has been the **ecosystem** in which the various parties operate and how does it affect federal evaluation efforts? What has been the "culture(s)" and mindsets of the milieus, organization(s), or venues in which the parties operate (i.e., norms and mindsets of a community of practice, norms and ethics of a profession) and how do these affect federal evaluation efforts?

Informants were selected based on their expertise in one or more of the following areas:

federal evaluation, evaluation use/influence/capacity, or the field of evaluation. The interviews and documents were analyzed by identifying trends in themes, barriers and facilitators, and types of influence on institutionalization. Coding and analytic methods grounded an understanding of actors, agendas, and assumptions; context, circumstances, and facilitators; and institutional ecosystem and process of diffusing value in evaluation.

Although the progress toward institutionalizing evaluation has seemed to ebb and flow in the Federal government over the last 60 years, findings suggest there are trends that can be used to trace the trajectory of institutional change. The process for institutionalization begins at the micro-levels and influences structural changes at the macro-levels (Li, 2017). A variety of factors can have an effect on any level of influence or layer of the ecosystem. These factors were strategically engaged to influence cognition, motivation, behavior, or structure depending on the culture and capacity of an organization. The U.S. Federal Government has used the *Analytical Perspectives* to codify a system-wide values chain that prioritizes evaluation as an essential function for a more efficient and effective government (van Leeuwen, 2007). Interviews with experts in evaluation practice informed the mapping of influential factors on institutionalizing evaluation. Many positive and negative factors were identified that have cognitive, motivational, behavioral, or structural influences on institutionalizing evaluation, but the pace and pattern of institutional change occurs at different magnitudes, rates, and frequencies within the U.S. Federal Government.

Influences on institutional change manifest in different ways, ranging from a belief system of value in evaluation to building a robust infrastructure to do and use more evaluation. Building an organizational culture of evaluation is at the core of institutionalization and the levels of influence are progressive, which suggest earlier stages must be established to sustain higher levels of influence. In a complex organization like the Federal government, this means incremental changes within individual agencies can catalyze the process toward full institutionalization of evaluation across the entire

government. Therefore, cultivating a healthy culture of evaluation, fostering leaders and advocates, equipping evaluators and evaluation units, and establishing a robust evidence infrastructure are essential to institutionalizing evaluation in the U.S. Federal Government (Armenakis et al., 1993; Shah et al., 2017). This dissertation study offers unique insight in the trajectory of evaluation in the U.S. Federal Government over time. Its findings may be instrumental to policymakers and organizational leaders who continue to work toward institutionalizing evaluation in the Federal government or other complex organizations.

## **Introducing a Conceptual Framework and Method of Mapping Influences on the Process of Institutionalizing Evaluation in the U.S. Federal Government**

Evaluation is a relatively new field of systematic inquiry that uses formal methodologies to make judgements regarding the values, merit, and worth of social programs and policies (Cousins et al., 2014; Rossi et al., 2019; Scriven, 1967; Trochim, 1998). Specifically, since the 1960s, U.S. Federal (government) evaluation has played a large role in demonstrating the effectiveness of government investments in these programs and policies to impact outcomes (Lemire, Fierro, et al., 2018; Maynard et al., 2016). As such, our society perpetuates the belief that taxpayers are owed evidence that suggest government officials are making responsible policy decisions and are held accountable to the consequences of those decisions (Clarke, 2005). Evaluations generate the necessary evidence that can be instrumental in informing program improvement and policymaking decisions (Grob, 2007; Maynard et al., 2016).

The federal evaluation market has grown dramatically in the last 60 years, which has led to increasing use and capacity for evaluation (Lemire, Nielsen, et al., 2018; Mark et al., 2009). However, the uptake of evaluation has been disproportionately distributed and there have been varying mechanisms for building value in evaluation across the government. The evolution of evaluation utilization has not been linear (Johnson, 1998)—in fact, the realistic complexity of building and sustaining organizational evaluation capacity includes a myriad of factors that promote or resist valuing evidence (Preskill & Boyle, 2008). As power is wielded by the hands of various actors in the policy arena, the levers of influence on institutionalizing evaluation continue to ebb and

flow in the U.S. Federal Government and our understanding of them remains unclear (Stone, 2008; Wedel, 2009). Studying and documenting the effective and ineffective drivers of progress can be useful in informing future policies and strategies for institutional change. Specifically, we need to investigate the parties that are involved, their motivations, the breadth and depth of the ecosystem, the surrounding powers or agendas, the barriers and facilitators, and the institutional structure (Wedel, 2019).

There are many methods available that could be suitable in studying the breadth of factors at play over time which influence the institutionalization of evaluation. Beyond the method of inquiry, it is necessary to conceptualize the context and process of institutionalizing organizational change to fully understand the complexity of this framework. This is the first of a series of three manuscripts that will study the institutionalization of evaluation in the U.S. Federal Government. A study of this magnitude and approach has not yet been found in the literature and it is necessary to consider the theoretical and methodological intricacies of this topic before engaging in an empirical study. Therefore, the aim of this first manuscript is to: 1) provide an overview of the landscape of evaluation in the Federal government; 2) present a conceptual framework for building institutional value in evaluation; and 3) introduce a novel ethnographic method of studying the process of institutionalization.

### **Evaluation in the U.S. Federal Government**

“Occupations have no real power of their own. They have only their knowledge and skill. Knowledge and skill can influence others, but only if others believe in their value. In order to gain a privileged position in the marketplace an occupation must persuade those who do have the power to grant and sustain it.”

—Eliot Freidson, 1989, p. 427

**Building professional legitimacy.** McDavid and Huse (2015) called evaluation a proto-profession, acknowledging its relative infancy as a field of science. Evaluators were initially employed to measure the effectiveness of programs—particularly community-based programs that used social interventions to promote change—but evaluations are now used in a variety of settings (Centers for Disease Control and Prevention, n.d.; Rossi et al., 2019). As the evaluation market continues rapid and steady growth, there may be a greater need to credential evaluators to qualify for high-stakes, lucrative positions (Lemire, Nielsen, et al., 2018; Love, 1994). Professional credentialing was derived out of the need to limit occupational practice to those who could perform the necessary duties precisely and safely, distinguishing practitioners between professional and amateur (Freidson, 1986, 2001). Since 1985, scholars have discussed and strategized the path toward and implications of professionalizing evaluation (Merwin & Wiener, 1985). Professional leaders have struggled with how and when to pursue professionalization for many reasons and they warn of the dangers we risk by establishing evaluation as a profession too quickly or not quickly enough (Altschuld, 1999; Bickman, 1999; Worthen, 1999).

The desire to professionalize evaluation is founded on the belief that professions have the power to frame societal issues and define (or privilege) “norms, standards, principles, and benchmarks” (Muzio, Brock, & Suddaby, 2013, p. 706) of the problems within their professional jurisdiction (Abbott, 1988; Loewenstein, 2014). For example, car owners often rely on mechanics to diagnose their vehicle for necessary repairs. For consumers who do not share the esoteric knowledge that certified mechanics possess,

they must trust these professionals to make fair and accurate judgments and the cost of this service is priced at the discretion of the professional (i.e., somewhat limited by the industry standard). Evaluators seek a similar elevation that certifies their specialization and empowers them to control the market standards.

Although the appeal of certification is justified, the process in qualifying, training, and measuring evaluation competency should not be undertaken lightly. In effort to maximize construct validity, certification exam developers would first require a clear understanding of what *is* evaluation practice and how to verify relevant knowledge, skills, and abilities through measurable tasks (LaDuca, 1994; Love, 1994; Mertens, 1994; Schwandt, 2018a). Professional credentialing requires its own workforce development to continuously build, run, analyze, and maintain certification integrity, and most certifying bodies would then also require some form of continuing education to maintain skills over time. Two decades later, evaluation continues to walk a fine line between its fear of losing legitimacy as a reputable field of practice and the burden of determining fair and accurate criteria to certify the legitimacy of current or future practitioners (Worthen, 1999). Despite the complexity and risks, there have been efforts to move toward defining evaluation practice and competency.

In 2003, the American Evaluation Association (AEA) Ethics Committee created the Guiding Principles for Evaluators with five principles: 1) systematic inquiry, 2) competence, 3) integrity/honesty, 4) respect for people, and 5) responsibilities for general and public welfare (American Evaluation Association, 2004). Parameters for systematic inquiry emerge from extant and thriving literature on evaluation theories and



multidisciplinary methods. The last three principles define a moral compass that are intended to drive evaluators to be good stewards of evidence generation and use. However, the second principle—competence—was subject to interpretation, which led to the creation of an AEA Task Force in 2015 that developed what are now called the Evaluator Competencies. The five domains of the AEA Evaluator Competencies include: 1) professional practice, 2) methodology, 3) context, 4) planning & management, and 5) interpersonal (American Evaluation Association, 2018). They serve to define the hard and soft skills of a competent evaluation practitioner and have the potential to operationalize formal evaluation policy, hiring criteria, or curricula (Christie & Lemire, 2019; King & Stevahn, 2015; Mark et al., 2009; Schwandt, 2017; Trochim, 2009).

To date, there are not well-established evaluation policies across the entire U.S. Federal government, and this sector lacks a job series dedicated to hiring “evaluators”. Despite the lack of a formal structure that includes evaluation in the government, there have been avenues to hire evaluators in areas where evaluation is valued. For example, the Office of Personnel Management (OPM) issued a direct-hire authority for Scientific, Technical, Engineering and Mathematics (STEM) positions (GW-007, 2018). Although evaluators are not a specified job title listed under the qualifying STEM positions, some agencies have used this authority to hire evaluators as “Economists” or “Statisticians” because these titles are most closely related to evaluation.

Valuing evaluation is built on a culture of trust and capacity for improvement, which grounds organizational learning objectives (Cousins et al., 2014; Fierro & Christie, 2016; Labin et al., 2012; Taylor-Powell & Boyd, 2008; Thames & Webster, 2009).

Unfortunately, modern society is plagued by “fake news” or reasonable doubt regarding the believability of evidence presented to the public, which leads to a greater desire for more trustworthy sources of evidence (Clarke, 2005). Not having clear standards or principles for conducting evaluations—internally or externally—or a clear identity for evaluators makes evidence generation inconsistent and inefficient (Commission on Evidence-Based Policymaking, 2017). Government inefficiencies lead to higher costs, increased public burden, and decreased effectiveness of public policies and programs (Gamoran, 2018; Liebman, 2018; Maynard, 2018; Olejniczak et al., 2016). This is a poor investment of government resources and diminishes the impact of these efforts in influencing social change (Pawson et al., 2011; Tabuga, 2017). Therefore, prioritizing evidence quality and efficiency causes the role of evaluators, evaluation practice, and evaluation use to become essential priorities for evidence-based policy decisions (Bogenschneider & Corbett, 2010; Schwandt, 2017).

**Evaluation for evidence-based policymaking.** Evidence-based policymaking is the process by which a critical evaluation of available evidence is done to ethically inform policy decisions (Bond, 2018; Greenhalgh & Russell, 2009; Haskins, 2018; McKay et al., 2015; Pawson et al., 2011). Both explicit and implicit (or tacit) forms of knowledge are viable in communicating useful evidence depending on the mechanism and purpose of the exchange (Blake & Ottoson, 2009; D'Eredita & Barreto, 2006; Stone, 2012). For example, Patton (1998) discovered that simply the process of doing evaluation can spark organizational learning as much as the findings from an evaluation. Christie (2015) adds,

Information is used conceptually, influencing how people think about an issue and assumes that the stakeholders hold beliefs, values, assumptions, and tacit knowledge, which are independent from the information generated from an evaluation study. (p. 2)

This type of ‘process use’ can be an intentional form of utilization that organizations can leverage to support evaluation capacity building (ECB) efforts (Alkin & King, 2016; Amo & Cousins, 2007). However, evidence from process use do not exist in isolation, but within a larger ecosystem of utilization. Building capacity to do evaluations is not simply the addition of discrete professional skills, it also requires organizational support structures (Cousins et al., 2014; Lawrenz et al., 2018; Mertens, 1994; Taylor-Powell & Boyd, 2008). A recently enacted policy establishes evaluation as an essential function and has the potential to change the culture and capacity of evaluation in the U.S. Federal Government.

**The Foundations for Evidence-Based Policymaking Act.** Since its emergence in the late 1960s, evaluation theories and methods have been investigated among academicians, while evaluation policies have been developed and implemented to guide evaluation practice (Alkin, 2013; Christie & Lemire, 2019; Mark et al., 2009; Shadish, 1998). What the government has really needed—in order to legitimize policies that dictate practice—is a mandate that could whet an appetite for more evaluation. This is precisely what the Foundations for Evidence-Based Policymaking Act ([P.L. 115-435](#)) has done. On January 14th, 2019, the Foundations for Evidence-Based Policymaking Act, commonly referred to as the Evidence Act, was enacted. The Act includes: Title I—

Federal Evidence-Building Activities; Title II—Open Government Data Act; and Title III—Confidential Information Protection and Statistical Efficiency Act (CIPSEA). The Evidence Act is expected to systematically revolutionize how evidence is generated, managed, and used in the public sector, and legitimizes the institutionalization of federal evaluation and statistical systems (Hanberger, 2001; O'Grady, 2010). While Title I has the most direct impact on federal evaluation, all three Titles influence evaluation practitioners to some degree.

Title I focuses on building agency capacity for evidence generation, use, and influence. Pursuant to §312a, agencies must develop an agency-wide strategic plan (i.e., an agency learning agenda), which should include: 1) policy-relevant questions, 2) a list of data required for policymaking, 3) a list of methods and analytical approaches for policymaking, and 4) a list of challenges to accessing relevant data (Duignan, 2003). This learning agenda must be reviewed and revised every four years. Furthermore, §312b mandates that each agency should then also create an annual evidence plan describing the agency's evaluation activities, which should include key evaluation questions and a data collection and acquisition plan. Strategies to engage key stakeholders should be outlined in the evidence plan (Duignan, 2003), but this mandate assumes agencies have a clear understanding of current practices, questions, and data sources that are employed. As such, it assumes that evaluation is already valued in each of the agencies and that there is a healthy culture for strengthening existing practices. Before creating learning agendas and evidence plans, agencies need to assess their current approaches for organizational learning to ensure that their data and methods are adequate to inform and sustain these

processes. Initial work to equip the agency with a robust infrastructure for evidence generation and management may be necessary before learning strategies can be developed and implemented.

Pursuant to §313, each head of agency must designate a senior official to serve as the Chief Evaluation Officer (CEO) of the agency to oversee and coordinate activities described in §312. In addition to mandating the designation, this section also describes the qualifications and function of the CEO and the Office of Management and Budget (OMB) has issued guidance (M-19-23) on Phase 1 of the implementation of the Act including information on the qualifications and duties of the CEO (U.S. Office of Management and Budget, 2019). By establishing evaluation with executive leadership, the goal is that the field can build the social capital and cultural infrastructure necessary to improve evaluation capacity (Bogenschneider & Corbett, 2010; Espedal, 2017; Morrill, 2008; Muzio et al., 2013; Thames & Webster, 2009). Whether to bolster an established culture and trust for evaluation or to foster a developing environment, the CEO should lead these efforts and use the aforementioned strategic activities to guide the agency. Likewise, §314 requires the head of agency to also appoint a Chief Statistician to advise on statistical policy, techniques, and procedures of the agency.

Both the CEO and the Chief Statistician are eligible to serve as representatives on the U.S. Advisory Committee on Data for Evidence Building, which will convene regularly to review, analyze, and make recommendations on how to promote the use of Federal data for evidence generation (Pasachoff, 2017). As these are new positions to the public sector, there is a greater need to equip these executives with tools that can inform

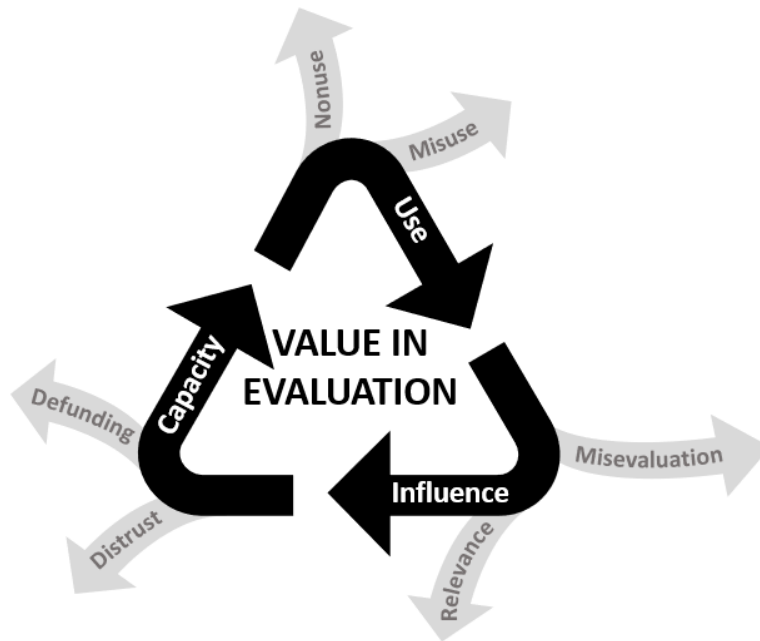
their decision-making processes on agency-wide evidence plans. However, the caveat is that the Evidence Act is also an unfunded mandate and agencies must rely on reallocating existing resources to support implementation. As such, agencies may be disproportionately equipped to build a greater culture and capacity for evaluation. These changes will need to be supported by two levels of capacity: *general capacity* of overarching entities like the OMB and the General Services Administration (GSA) to support widespread implementation and *specialized capacity* to manage and support agency-wide capacity, which would allow agency leaders be responsive to their unique organizational culture (Hall & Hord, 2015; Ting, 2009). The CEO and Chief Statistician need cost-effective approaches for a rapid and effective response to the Act.

Chelimsky (2015) recently gave a personal account of internal organizational factors that *impede* evaluation use based on her notable career with the U.S. Government Accountability Office (GAO). She suggests that the factors can be broken down into a careful balance between three major categories: 1) political factors, 2) systemic factors, and 3) evaluation-related factors. She asserts that “to achieve strong evaluation practice with appropriate use of findings, you need to have so many elements functioning simultaneously at the highest possible level that a miracle would be needed for it all to come together successfully” (Chelimsky, 2015, p. 88). These factors provide a substantial foundation for understanding the threats to evaluation use, but there are more pieces of this multi-factorial puzzle to be uncovered—particularly, a greater understanding of how value in evaluation is distributed or stifled in the government.

## Conceptualizing the Diffusion of Value in Evaluation

**The manifestation of value in evaluation.** Early trailblazers wrote many publications and reports that emphasized the value of evaluation in the government (Chelimsky, 1977; Datta, 2003; Grob, 2007; Richardson, 1992). Making a case for the value *of* evaluation (i.e., a statement of worth) inspires a value *in* evaluation (i.e., an indication of belief). For example, when shopping for a new cellular phone, a salesperson might state the price of a device and attempt to justify its worth based on its unique features. The buyer could express interest in purchasing a cheaper device with seemingly similar qualities, but the retailer must then convince the buyer how and why the first device is special and deserving of consideration. If the buyer believes in the worth of the device, they will invest in purchasing it and indicate their value in its unique functionality. Therefore, although many have expressed the legitimacy of evaluation, there has been far less documented in the literature regarding how (i.e., the factors and decisions made) value in evaluation has been demonstrated and diffused.

Scholars maintain that the value of evaluation is manifested in evaluation use, influence, and capacity (Cousins et al., 2014; King & Alkin, 2019; Kirkhart, 2000; Leviton, 2014). Therefore, there is a cyclic relationship among evaluation use, influence, and capacity that are built upon valuing evaluation as a fundamental principle (Figure 1). Although each facet is simplified in this model, this figure also illustrates offshoots in which organizations may diverge from a healthy culture of evaluation resulting in devaluing evaluation in an organization (Oliver, 1992). As a result, less evidence may be generated to support critical decisions for improvement and learning.



*Figure 1.* Relationship Among Evaluation Use, Influence, and Capacity

*Note.* This model shows that value in evaluation requires a careful balance of evaluation use, influence, and capacity whereby increase in one facet leads to an increase in another. Likewise, decreased or inappropriate applications of one facet can cause an offshoot from the cycle and result in a decrease in value in evaluation.

The consequence of conducting evaluations of poor methodological rigor (i.e., not using systematic and robust methods which would threaten validity) or relevance is that the use and influence of evaluation on decision-making would be threatened (Alkin & King, 2017; Patton, 2015a). The threat of unreliable evidence—evidence that is unfounded and is not credible—may lead to a culture of distrust and demotivates funding evaluation, as those activities would not be a valuable contribution to decision-making processes (King & Stevahn, 2013; Labin et al., 2012). Alternatively, responsible conduction and consumption of evaluation can lead to the prioritization of evidence and



fosters a greater capacity to do and use evaluation (Cousins et al., 2014). Influence, then, is the applicability of evaluation to elicit social change (Kirkhart, 2000). Although it is unclear how much evaluation has grown in the Federal government since its inception, it is clear the field has greatly increased in maturity and demand (Lemire, Fierro, et al., 2018; Maynard et al., 2016). This growth demonstrates increasing value in evaluation, but it is not certain that the belief in its worth indicates that evaluation has been institutionalized in the government. We should first conceptualize the process by which beliefs and attitudes are distributed in an organization to influence widespread change.

**Developing the conceptual framework.** There have been many theoretical frameworks or models used to describe evaluation use in the literature. Each offers different combinations of dimensionality to frame factors that influence evaluation use in different ways. However, institutionalization extends beyond the varying methods of doing evaluation or how the findings are utilized. The way in which episodic changes become systemic is by establishing significance, amplification/diffusion, and lamination of new ideas (Gray et al., 2015; Li, 2017; Strang & Meyer, 1993). Likewise, the institutionalization of new markets in organizations is reliant on the diffusion of a concept across all levels and contexts (Colyvas & Jonsson, 2011; Meyer et al., 2018; Strang & Soule, 1998; Zucker, 1977).

The process in which a novel idea becomes diffuse in an organization is through the combination of many factors that promote institutionalization. Therefore, there needs to be a framework and method that can guide the study of such a complex interplay of factors that may be used to study the conditions that allowed or prohibited the diffusion

of evaluation in the U.S. Federal Government (Shore & Wright, 2011; Strang & Meyer, 1993; Wedel, 2019). Despite not being designed for this exact purpose, due to the relationship between use and value, there is potential to adapt existing models of evaluation use or influence to conceptualize the diffusion of value in evaluation in the U.S. Federal Government.

**Complexity of the U.S. Federal government.** The U.S. Federal Government is, as a whole, a complex organization that is comprised of three branches: legislative, executive, and judicial (usa.gov, 2019). Congress—comprised of the Senate and House of Representatives—is responsible for drafting legislation and has a number of agencies that support its function. The Supreme Court and other Federal courts in the judicial branch interpret and uphold the laws that are passed by Congress. The executive branch includes the executive offices of the U.S. President, 15 departments (each with layers of sub-agencies and bureaus), and several independent agencies. There are an estimated 2.1 million civilian workers across the Federal government (Jennings & Nagel, 2019). Including the President, Vice President, Senators, and Representatives, there are a total of 542 elected officials (fvap.gov, n.d.). Once elected the President appoints Directors, Secretaries (and Deputy Secretaries), and Administrators for the executive departments and agencies, many of which then sit on the President’s Cabinet (The White House, n.d.).

The three branches of government are interconnected by design to hold each other accountable of their duties. The organizational structure is not only large, but the relationships among the various individuals and offices are complicated. Although term limits and elections foster innovation and new perspectives, it nearly guarantees that the

organizational culture is always changing. In addition to the internal complexities that influence institutional change in the Federal government, there are external factors to consider as well (Hall & Hord, 2015). External to the Federal government are also layers of domestic and international factors that can impact the organizational behavior of the Federal government. Therefore, the framework used to conceptualize the process of diffusing system-wide value in evaluation must adequately recognize the complexity of the entire ecosystem.

**Comparison of existing frameworks.** A collection of works by Marvin Alkin and Jean King documented the evolution of theories on evaluation use and the frameworks developed by different theories over time (Alkin & King, 2016, 2017; King & Alkin, 2019). Among those described by Alkin and King, are several models that could be used to inspire a new framework to conceptualize the process of diffusing value in evaluation. Lincoln and Guba (2004) provided an early adaption of Kaplan's pattern theory to map the "constellation" of an organizational ecosystem (Alkin & King, 2017; Kaplan, 1964). Lincoln and Guba's model appropriately accounts for the players, networks, and behaviors, but does not include any external factors that might be acting *on* them (as shown in Alkin & King, 2017).

Mark and Henry (2004) shared a model of alternative mechanisms for evaluation influence, which included general, cognitive/affective, motivational, and behavioral factors. General influence processes refer to the structural changes that might trigger changes in other areas. Cognitive and affective influences are ones that positively or negatively impact the thoughts, feelings, and attitudes of the players. Motivational

influences have an effect on the goals or aspirations of the players by punishing or incentivizing behavior. Behavioral influences, then, are direct changes in actions or practices that can occur at the individual- or organizational-level. These influences speak to how people think about, justify doing, and conduct evaluations. In their proposed influence framework, Mark and Henry (2004) add levels of analysis which allow the categories of influences to be associated with a unit of agency (i.e., individual, interpersonal, or collective). While their framework offers more dimensionality than Lincoln and Guba's model, the levels of analysis limit the greater ecosystem to a "collective", but there are other distinct levels of internal and external players affected by or acting on changes in evaluation use (Hall & Hord, 2015; Sroufe, 2017).

Alternatively, King (1988) and Johnson (1998) both conceptualized their categories of evaluation utilization—user, evaluator, evaluation, and organizational—to provide greater detail of the internal layers, but continue to omit layers external to the organization (Alkin & King, 2016; King & Alkin, 2019). Lastly, Ottoson and Martinez (2010) created an ecological model of evaluation use whereby nested layers of the ecosystem had different threads, leveraging, and types of use. This model approaches a more accurate structural representation of the environment in which federal evaluation exists because it appreciates external influences (i.e., from the community, field, and society) on valuing evaluation. The layers in Ottoson and Martinez's model demonstrate the interactions among various levels and acknowledge the potential for factors to be self-contained within an inner level or reverberate to other levels (Greve & Teh, 2018). However, the smallest unit in their model is the program, with the immediate next layer

in the ecosystem as the community. Systems thinking for complex organizations like the Federal government requires an intermediary layer between the programs and community which include organizational-level factors.

Kirkhart (2000) took a very different approach to conceptualizing the influential factors of both evaluation use and influence. She posited that evaluation results had instrumental, symbolic, or conceptual uses, and that there were many factors that influenced how findings were used. She also constructed a three-dimensional model of evaluation influence, which she situated intention, source, and time as related elements. In contrast to Mark and Henry's model, cognitive and affective are separated as two different reasons for evaluation use. These and other motivations for evaluation use are combined with other variables such as frequency, limitations, and mechanisms of use. However, this model only maps the possible motivations and applications for evaluation use or misuse. Kirkhart's model does not include actors or effectively position players as users, generators, or influencers of evaluation. As such, the existing models cannot be used to study the process of diffusing value in evaluation because they all do not adequately account for the actors involved in this process, their relationships to each other, and the depth or complexity of the ecosystem in which they coexist (Table 1).

Table 1.

Features and Relevant Elements of Existing Conceptual Frameworks

<b>Theorist(s)</b>	<b>Features of the Existing Model</b>	<b>Relevant Elements for the New Framework</b>
Kirkhart (2000)	Circular model with instrumental, symbolic, and conceptual uses of evaluation in the middle surrounded by other factors. Only includes motivations and mechanisms for utilization.	Circular model and some of the motivations redistributed as either types of influence or factors that influence diffusion.
King (1988) and Johnson (1998)	Feedback loop including background, interactional, and utilization variables across user, evaluator, evaluation, and organizational layers.	Nested internal layers of an organization and acknowledgement of a relationship between the types of influences acting on these layers.
Lincoln and Guba (2004)	Pattern theory constellation including players, networks, and behaviors. Does not include the nested nature of the Federal government.	The interconnectedness of players, networks, and their behaviors.
Mark and Henry (2004)	4x3 table with type of process as rows (i.e., general, cognitive and affective, motivational, and behavioral) and levels of analysis as columns (i.e., individual, interpersonal, and collective).	Types of process adapted to four levels of influence: cognitive, motivational, behavioral, and structural.
Ottoson and Martinez (2010)	Circular model with nested layers (i.e., program, community, field, and society) with leveraged, threads, and types of use spanning across all layers.	Circular, nested layers with facets of the concept that travel through all layers of the ecosystem.

**Proposing a new conceptual framework.** A new model to be employed as a conceptual framework of the factors involved in diffusion of valuing evaluation is displayed as Figure 2. This proposed integrated conceptual framework is equipped to capture the intricacies of influence on shifts in federal evaluation. The elements borrowed from existing frameworks are shown in Table 1. King's and Johnson's categories of use are adjusted and represented as the levels of analysis with nods to factors within an organization (i.e., the microsystems; agencies within the Federal government), across an organization (i.e., the mesosystem; the entire Federal government), and in the larger national or global context (i.e., layers of the macrosystem; domestic and international arenas).

Lincoln and Guba inspired thoughtfulness about the interconnectedness of actors, networks, and their behaviors. As such, the layers in the proposed model are also not rigid. They represent an exchange of information and techniques to permeate as actors communicate with each other, publicly disseminate their work, and collaborate with other players. These contextual layers are nested, as Ottoson and Martinez also proposed, to illustrate how they build on and relate to each other. In the U.S. Federal Government, this captures the nuances within Federal agencies and between agencies across the macrosystem, but also the events, actions, and policies in a broader context that may also have downstream effects at the sub-levels. Field-configuring events may be experienced and perceived in different ways by actors across different levels with other factors at play in their atmosphere (Hardy & Maguire, 2010; Zilber, 2002).

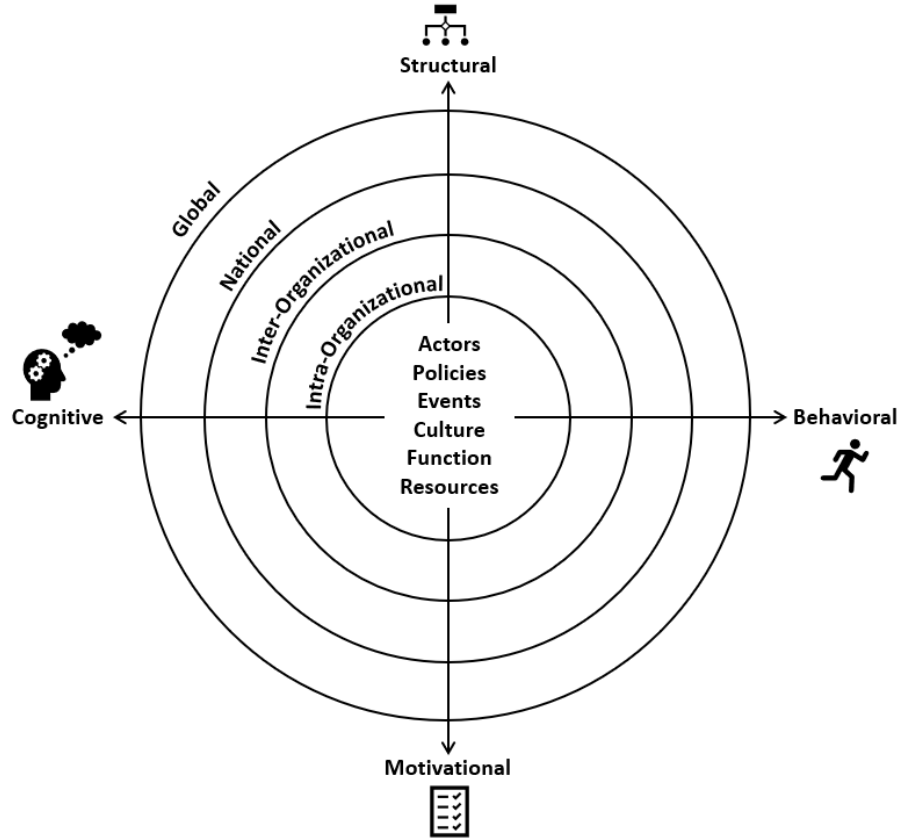


Figure 2. Conceptual Framework for Diffusing Value in Evaluation

*Note.* This framework illustrates the ecological system of diffusing value in evaluation. It captures the complexity of the U.S. Federal Government and breadth of factors that impact diffusion, which may influence cognition, motivation, behavior, or structure.

In this new framework, Mark and Henry’s mechanisms of influence now flow across all contextual levels, but we rename “general” to “structural” to more clearly reflect the types of influences that might fit this category in the organizational (infra)structure. Although it is theorized that the levels of influence have an effect on each other, empirical data will reveal how they are related. In the center of this framework are common factors that may or may not play a role in influencing the



diffusion of value in evaluation across all levels, which reflect Kirkhart's recognition of the varying factors that facilitate or impede evaluation use. In contrast, the creation of new roles (i.e., hiring of new employees), policies, or practices may also be byproducts (i.e., outputs or outcomes) or evidence of organizational change (Hogan & Coote, 2014). These factors will be discussed in greater detail within the next section because they may be potential data sources to study the process. This theoretical framework can be used to guide coding methods during the mapping process but may also have other applications for research on evaluation use in complex organizations. The following section presents an ethnographic mapping methodology as a viable approach to studying the institutionalization of evaluation in the U.S. Federal Government.

### **The Mapping Method**

**Existing methods of mapping.** Popularized by the field of psychology, some forms of idea, cognitive, or mind mapping have been used introspectively as a graphical projection of one's cognition (Behrens et al., 2018; Tolman et al., 1946). These spatial mapping techniques can help researchers and practitioners navigate neuronal pathways using a coding scheme to organize connections between thoughts (Behrens et al., 2018; O'Keefe & Nadel, 1978). These mind models map brain activity and relationships between pathways created by a thought process or behavior. The study of spatial behavior is a way to uncover vectors of cognition (O'Keefe & Nadel, 1978; Tolman et al., 1946), but require a high-level of expertise in neuroscience and do not necessarily influence other people or activities. Instead of mapping internal cognitive processes, however, there is a greater need to map concepts, events, and actors to understand the entire landscape.

Methods of conceptual cartography are not new to social science research, and certainly not to evaluation (Trochim, 1989). Concept mapping has been used by evaluators as a structured approach to understanding different concepts and their relationships to one another (Trochim, 1989). Kane and Trochim (2007a) describe a six-step process that aims to cluster stakeholders' ratings regarding the importance of an issue at hand. It incorporates both an exploratory qualitative process to create an appropriate tool for rating importance and a quantitative approach to statistically model the resultant ratings into nested clusters (Trochim & McLinden, 2017). Trusting that the qualitative discovery of key concepts has been comprehensive, the rating can be very robust and allow an understanding of subjective rankings for important concepts.

This concept mapping method has been employed by researchers to surface tacit knowledge or implicit community partnerships (Brennan et al., 2012; Wutzke et al., 2017). However, findings assume the rating tool abides by the methodological rules of engagement for construct validity because participants only rate concepts that emerged from the qualitative interviews (Kane & Trochim, 2007b; Wedel, 2019). Although there have been documented cases in which concept mapping has been influential to evaluation, Trochim (2017) recently published a critique noting the technical and mechanical limitations of his own method. The primary concerns were with the complexity of the method and esoteric demands on practitioners. This method requires specialized software and skills in using statistical programs to conduct the analyses. Some of the tools built to streamline this method are proprietary and would be expensive

to acquire. Lastly, Trochim criticizes the unchanging nature of this method, which has failed to change with the advancements of big data and technology.

Social network analyses (SNA) appreciate networks as having deep social capital because they are multifaceted and complex (Kadushin, 2012; Martin, 2015; Tichy et al., 1979). In most SNA approaches, the units of analyses are similar (i.e., individuals or groups), but their relationships to each other vary and are mapped based on the strength and proximity of their relationships (Hanneman & Riddle, 2005; Kadushin, 2012). However, some social phenomena, similar to the process in which evaluation became institutionalized in the Federal government, do not simply involve interactions between actors and organizations (i.e., groups of actors). A method that would adequately capture the intricacies of this process would need to include other influential policies, events, cultures, or contextual factors—in addition to the players involved—that have also effectively changed how evaluation has been valued in the government. Wedel (2019) offers the foundation of an approach that addresses these concerns.

**A social anthropological approach.** What is missing from the other methods of mapping is the recognition of relationships and culture that are deeply embedded in why or how people may perceive importance. Social anthropologists have a knack for appreciating the milieus and cultural logics that others may overlook or even intentionally parse out as confounding factors (Però, 2011; Shore & Wright, 2011; Wedel, 2004, 2019). In fact, Wedel (2019) first used this mapping method to study policy and governance processes and found a need to understand a vast, but hidden, social network between various power sources—not only ones codified in documents and newsprint. The

factors involved were primarily actors, but also policies, events, documents, and other influences that acted on actors and actions (Hogan & Coote, 2014; Zilber, 2002). In addition to concrete factors, there were also intangible influences that emerged such as political agendas, power, and culture. This mapping method was birthed from the demand for a comprehensive approach that could be adaptable to real-time developments in field research. Among the various contexts where Wedel employed this mapping method, one relevant example is a study she conducted to map the players and processes in and around the U.S. Federal Government that led to the 2008 financial crisis in the United States (Wedel, 2014, 2019). Wedel's accounts of that study will be embedded throughout the following description of the methodology to provide further context.

Wedel describes this ethnographic approach as an effort to study “through” an issue in a non-linear manner instead of top-down like many other methods (Burawoy, 2009; Peck & Theodore, 2015; Wedel, 2001, 2004; Wright & Reinhold, 2011). Però (2011) suggests a bottom-up approach when studying the ways in which policies have affected the governed—and not just starting with the governing institution—by writing,

In order to recognise the agency of the governed, we need to adopt a broader and more comprehensive definition of policy change, one that conceives of change as not just the result of top-down initiatives of powerful institutions but also of the bottom-up engagements of subaltern individuals. (p. 260)

As such, Wedel advises to start by observing the organizational milieus and speaking to informants who experienced the effects of the processes, not only to those who created or

executed them. Although those institutional players will also be included in the study, those affected may be easier to access and able to speak to reality rather than intent.

It may also be natural to seek artifacts (i.e., tangible documents and records) related to the topic of interest, but undocumented lived experiences can describe policy and government processes beyond what is codified in artifacts. Especially when studying polarizing or sensitive issues, the lack of explicit artifacts may be intentional and information that can be uncovered might be a facade. However, there may be actors in the shadows or other processes worth investigating that may be in tandem or parallel to the issue(s) of interest. For example, in the U.S. financial crisis study, Wedel found that a “power clique” was involved in deregulating exotic derivatives according to their personal agenda (Wedel, 2014). Their decisions to game America’s economy is not likely a story that members of this clique would accurately tell themselves, but it is still important to hear how they perceive their actions, their motivations for them, and the effect they believe their actions had on others. It is for these reasons that this mapping method does not have a single or dedicated entry point for data collection (Wedel, 2019).

Led by a series of open-ended guiding questions, the method, while systematic, is a flexible and interactive approach to making unknowns known about social processes (Pawson et al., 2011; Wedel, 2019). Thus, although the questions are listed below in number order, there is no particular order in which they should be answered when collecting data. The trajectory of the study, the order in which the questions are answered, and the number of questions that are able to be answered depend on how the data unfold as the study progresses. The methodology is truly developmental, as it allows

the ethnographer to respond to the dynamic process of uncovering complex truths (Wedel, 2004). Table 2 displays a comparison of each of the aforementioned methods across the reason or context for use, data sources required, and the advantages and disadvantages for each. Wedel suggests the following six questions, which are tailored here to study the process of institutionalizing evaluation in the U.S. Federal Government, as guiding questions that would drive empirical methods:

1. Who have been the **parties** (i.e., players, social networks, and organizations) involved, both formally and informally, involved in the institutionalization of federal evaluation?
2. What have been the parties' respective interests, **agendas**, incentives, goals, motivations, operating assumptions, and expectations with regard to institutionalizing federal evaluation?
3. Who have been the parties' **sponsors** (i.e., funders or providers of other indispensable resources) and what are their interests, agendas, incentives, goals, motivations, operating assumptions, and expectations with regards to federal evaluation activities?
4. What has been the larger context(s)/circumstances of **power** and **resources** in which the parties are embedded? How have features of the larger context(s)/circumstances constrain, enable, or otherwise affect the parties involved in federal evaluation efforts?
5. To what extent has the **structure** that facilitates federal evaluation activities or efforts institutionalized (or becoming so) even if it remains informal/unincorporated and the players change over time?
6. What has been the **ecosystem** in which the various parties operate and how does it affect federal evaluation efforts? What has been the "culture(s)" and mindsets of the milieu, organization(s), or venues in which the parties operate (i.e., norms and mindsets of a community of practice, norms and ethics of a profession) and how do these affect federal evaluation efforts?

Table 2.

Comparison Among Various Mapping Methods

Mapping Method	Context for Use	Data Sources	Advantages	Disadvantages
Cognitive Mapping <b>(Tolman et al., 1946)</b>	Mapping neural pathways for thought processes	Brain Imaging Cognitive Interviews	Can help researchers understand how people think and feel about issues.	Expensive and requires expert knowledge regarding neuroscience. Not applicable for mapping organizational behavior.
Concept Mapping <b>(Trochim, 1989)</b>	Clustering concepts by importance as rated by actors in an organization	Cognitive Interviews Questionnaire	Can help researchers identify groupings of concepts and their relative importance to an issue or organization.	Can be time consuming. Requires specialized software and skills in statistical programs. Ratings of importance are reliant on and assume construct validity.
Social Network Analysis <b>(Tichy et al., 1979)</b>	Mapping the actors in a social network and their relationships to one another	Interviews Observations	Can help researchers understand relational aspects of a social network or infrastructure.	Typically only focused on one level of analysis (i.e., the actors) and does not include the relationships with other factors involved.
Wedel’s Mapping Method <b>(Wedel, 2019)</b>	Mapping any and all factors that influence a process	Any Available	Can help researchers to study “through” an issue or process, accounting for varying factors and utilizing any and all methods necessary. Skills and tools do not have to be specialized for this method.	Can be time consuming. Requires expert and diverse skills in research methods, and access to key informants and documents.

The guiding questions frame the aims of a mapping study and may be chosen as applicable to the topic, interests, or available evidence. They also happen to advantageously reflect our curiosities about institutionalizing evaluation, as expressed in the introduction of this manuscript. Therefore, data are collected “from as many means and sources as are available” (Wedel, 2019, p. 3). After all, the goal of the method is not necessarily to discover emerging themes or to extract direct quotes; but to map a complex process to the extent of the breadth and depth of what could be captured through data collection. Unlike other methods that require access to and proficiency in advanced analytic software, this method can be done completely by hand. In fact, when Wedel was unable to record sensitive conversations with informants, she transcribed them from memory and relied on field notes to capture details that were heard and observed. She cross-checked and triangulated her findings with her informants and other resources (Burawoy, 2009; Wedel, 2004, 2019; Wright & Reinhold, 2011).

Wedel (2019) instructs that researchers should begin with the (known) players affiliated with the organization. Affiliations may be loosely tied, as she emphasizes the need to include those who are impacted by the actions and decisions of the organization or those who simply have privileged knowledge about the system (Wedel, 2004). These actors can speak to if/how institutionalizing a new idea has been socialized in and around an organization from their vantage point (Zucker, 1977). It is a reflexive method that traces policies and connections across space and time, which means researchers must be amenable to following unpredictable leads as they develop and must also be resourceful in locating other sources of data as the study develops (Burawoy, 2009; Peck &



Theodore, 2015; Shore & Wright, 2011; Zucker, 1977). The research questions allow the researcher to simultaneously focus on micro- and macro-level systemic factors of influence by seamlessly traveling between layers of the ecosystem (Hall & Hord, 2015; Zucker, 1977).

**Possible limitations.** The limitations of this method are that it may require a great deal of time to follow sources to the furthest edges of a complex issue—in fact, the vastness could be indefinite, growing, or changing throughout the study. Due to the nature of various sources of data and methods of collection, this approach also requires the researcher to be well-versed in many different research methods. For example, if the researcher discovers documents that inform the mapping method, they may be required to perform a content analysis of the text. When informants are identified and agree to participate, the researcher will need to conduct robust interviews and analyze them as a part of the study. These sources may emerge during the course of the study, which would require the researcher to already be agile and highly proficient at these skills to respond appropriately. Limitations in the researcher’s ability to adapt their methods to match any source of data might lead to omitting pertinent data, which may compromise findings; unnecessary delays, if the time to learn data collection or analytic skills needs to be taken; or inconvenient expenses, if other experts need to be consulted for portions of the study. All of these obstacles severely threaten the validity of the findings or the completion of the study in a timely manner (i.e., risking the relevance of the findings as the institutional climate and culture rapidly changes).

For sensitive topics—because the goal is to capture the good, the bad, and the

wicked truths of an issue—the work may also be hazardous. For example, Wedel has used this mapping method to study the drivers of corruption in different regions of the world (Wedel, 2019). When she studied the 2008 financial crisis, she identified potentially corrupt players, processes, and policies that ultimately drove our nation into a recession. This work is important in helping government leaders and policymakers be more attentive to warning signs and seek measures to mitigate another recession in the future. This study, and others Wedel conducted in Africa, Europe, and Asia, were in high-stakes environments identifying powerful players in corruptive processes. Although this method is flexible and responsive, if the researcher is not cautious, methodical, and astute they could stumble into risky situations and make formidable enemies. Therefore, it is the responsibility of the researcher to be committed to truth seeking in a manner that is above reproach in rigor and ethics.

## **Conclusion**

Large-scale organizational change has been, will be, and is already happening across all agencies of the government to increase the use and influence of evaluation. Although advancements like the digital (or information) age caused a monumental shift in our ability and efficiency to do and use evaluations, there must first be a culture and capacity to do and use evaluations (Caracelli, 2000; King & Alkin, 2019; Wedel, 2009). Building culture and capacity for evaluation is a complex process and the most effective or ineffective levers of change remain unknown. It is beneficial to the profession of evaluation to understand factors that facilitate or impede the institutionalization of evaluation in complex systems (Caracelli, 2006; Mark, 2001; U.S. General Accounting

Office, 2003a; Warrick, 2017). As Stone (2008) states, “reflection on success and failure potentially promotes efficiency, innovation, and learning in policy” (p. 28).

Mapping the influential factors for institutionalizing evaluation in the Federal government requires an understanding of how evaluation is valued and the process that would be most effective in diffusing that belief given the unique micro-, meso-, and macrolevel factors at play (Nigam & Ocasio, 2010; Schlaufer et al., 2018; U.S. General Accounting Office, 2003a). This paper introduces a framework to conceptualize the diffusion of value in evaluation and an empirical method that can adequately map the full breadth of influential factors for institutionalizing evaluation in the U.S. Federal Government. If time constraints, research proficiency, and sensitive topics can be weathered, there is not a more robust method of inquiry to map complex processes. Wedel’s success in using this method to study complex and sensitive topics in many different regions of the world—but especially her study of economic policies in the U.S. Federal Government—offer insights to the application and potential of this method for the current topic of interest. Similar to how previous studies have informed policy and government processes in other contexts, this novel approach to organizational learning will inform strategies that have and can influence the institutionalization of evaluation in a manner that is responsive to the unique culture and complexity of the U.S. Federal Government.

## References

- Abbott, A. (1988). *The system of professions: An essay on the division of expert labor*.  
The University of Chicago Press.
- Alkin, M. C. (Ed.). (2013). *Evaluation roots: A wider perspective of theorists' views and influences* (2nd ed.). SAGE Publishers, Inc.
- Alkin, M. C., & King, J. A. (2016). The historical development of evaluation use.  
*American Journal of Evaluation*, 37(4), 568-579.  
<https://doi.org/10.1177/1098214016665164>
- 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. <https://doi.org/10.1177/1098214017717015>
- Altschuld, J. W. (1999). The certification of evaluators: Highlights from a report submitted to the Board of Directors of the American Evaluation Association.  
*American Journal of Evaluation*, 20(3), 481-493.
- American Evaluation Association. (2004). *AEA Guiding Principles for Evaluators*.  
Washington, DC Retrieved from <http://www.>  
<https://www.eval.org/p/cm/ld/fid=51>.
- American Evaluation Association. (2018). *AEA Evaluator Competencies*. Washington, DC Retrieved from <https://www.eval.org/page/competencies>.

Amo, C., & Cousins, J. B. (2007). Going through the process: An examination of the operationalization of process use in empirical research on evaluation. In J. B. Cousins (Ed.), *Process use in theory, research, and practice. New Directions for Evaluation*, 116, 5-26. <https://doi.org/10.1002/ev.240>

Behrens, T. E. J., Muller, T. H., Whittington, J. C. R., Mark, S., Baram, A. B., Stachenfeld, K. L., & Kurth-Nelson, Z. (2018). What is a cognitive map? Organizing knowledge for flexible behavior. *Neuron*, 100, 490-509. <https://doi.org/10.1016/j.neuron.2018.10.002>

Bickman, L. (1999). AEA, bold or timid? *American Journal of Evaluation*, 20(3), 519-520.

Blake, S. C., & Ottoson, J. M. (2009). Knowledge utilization: Implications for evaluation. In J. M. Ottoson & P. Hawe (Eds.), *Knowledge utilization, diffusion, implementation, transfer, and translation: Implications for evaluation. New Directions for Evaluation*, 124, 21-34. <https://doi.org/10.1002/ev.311>

Bogensneider, K., & Corbett, T. J. (2010). *Evidence-based policymaking: Insights from policy-minded researchers and research-minded policymakers*. Taylor and Francis Group, LLC.

Bond, G. R. (2018). Evidence-based policy strategies: A typology. *Clinical Psychology Science and Practice*, 25, 1-14. <https://doi.org/10.1111/cpsp.12267>

Brennan, L. K., Brownson, R. C., Kelly, C., Ivey, M. K., & Leviton, L. C. (2012).

Concept mapping: Priority community strategies to create changes to support active living. *American Journal of Preventative Medicine*, 43(5S4), S337-S350.

<https://doi.org/10.1016/j.amepre.2012.07.015>

Burawoy, M. (2009). The extended case method: Race and class in postcolonial Africa.

In *The extended case method: Four countries, four decades, four great transformations, and one theoretical tradition* (pp. 19-72). University of California Press.

Caracelli, V. J. (2000). Evaluation use at the threshold of the twenty-first century. In V. J.

Caracelli & H. Preskill (Eds.), *The expanding scope of evaluation use. New Directions for Evaluation*, 88, 99-111. <https://doi.org/10.1002/ev.1194>

Caracelli, V. J. (2006). Enhancing the policy process through the use of ethnography and other study frameworks: A mixed-method strategy. *Research in the Schools*, 3(1), 84-92.

Centers for Disease Control and Prevention. (n.d.). *Brief 1: Overview of policy evaluation*. <https://www.cdc.gov/injury/pdfs/policy/Brief%201-a.pdf>

Chelimsky, E. (1977). Improving the use of evaluation: An agency perspective. *The Prison Journal*, 57(1), 13-18. <https://doi.org/10.1177/003288557705700103>

- Chelimsky, E. (2015). A strategy for improving the use of evaluation findings in policy. In C. A. Christie & A. T. Vo (Eds.), *Evaluation use and decision making in society: A tribute to Marvin C. Alkin* (pp. 73-90). Information Age Publishing Inc.
- Christie, C. A. (2015). Setting the state for understanding evaluation use and decision making. In C. A. Christie & A. T. Vo (Eds.), *Evaluation use and decision making in society: A tribute to Marvin C. Alkin* (pp. 1-10). Information Age Publishing, Inc.
- Christie, C. A., & Lemire, S. T. (2019). Why evaluation theory should be used to inform evaluation policy. *American Journal of Evaluation*, 40(4), 490-508.  
<https://doi.org/10.1177/1098214018824045>
- Clarke, J. (2005). Performing for the public: Doubt, desire, and the evaluation of public services. In P. du Gay (Ed.), *Values of Bureaucracy* (pp. 211-232). Oxford University Press.
- Colyvas, J. A., & Jonsson, S. (2011). Ubiquity and legitimacy: Disentangling diffusion and institutionalization. *Sociological Theory*, 29(1), 27-53.
- Commission on Evidence-Based Policymaking. (2017). *The promise of evidence-based policymaking: Report of the commission on evidence-based policymaking*. Washington, DC Retrieved from <https://www.cep.gov/report/cep-final-report.pdf>.
- Cousins, J. B., Goh, S. C., Elliott, C. J., & Bourgeois, I. (2014). Framing the capacity to do and use evaluation. In J. B. Cousins & I. Bourgeois (Eds.), *Organizational*

*capacity to do and use evaluation. New Directions for Evaluation, 141, 7-23.*

<https://doi.org/10.1002/ev.20076>

D'Eredita, M. A., & Barreto, C. (2006). How does tacit knowledge proliferate? An evidence-based perspective. *Organization Studies, 27*, 1821-1841.

<https://doi.org/10.1177/017084060606067666>

Datta, L. (2003). The evaluation profession and the government. In T. Kellaghan & D. L. Stufflebeam (Eds.), *International Handbook of Educational Evaluation* (pp. 345-360). Dordrecht: Kluwer Academic Publishers.

Duignan, P. (2003). Mainstreaming evaluation or building evaluation capability? Three key elements. In J. J. Barnette & J. R. Sanders (Eds.), *The mainstreaming of evaluation. New Directions for Evaluation, 99*, 7-21. <https://doi.org/10.1002/ev.90>

Espedal, B. (2017). Understanding how balancing autonomy and power might occur in leading organizational change. *European Management Journal, 35*, 155-163.

<https://doi.org/10.1016/j.emj.2016.08.005>

Fierro, L. A., & Christie, C. A. (2016). Evaluator and program manager perceptions of evaluation capacity and evaluation practice. *American Journal of Evaluation, 38*,

376-392. <https://doi.org/10.1177/1098214016667581>

Freidson, E. (1986). *Professional powers: A study of the institutionalization of formal knowledge*. University of Chicago Press.



Freidson, E. (1989). Theory and the professions. *Indiana Law Journal*, 64, 423-432.

Freidson, E. (2001). *Professionalism: The third logic*. University of Chicago Press.

fvap.gov. (n.d.). *About Elections*. Federal Voting Assistance Program.

<https://www.fvap.gov/info/about-absentee-voting/elections>

Gamoran, A. (2018). Evidence-based policy in the real world: A cautionary view. *Annals of the American Academy of Political and Social Sciences*, 678, 180-191.

<https://doi.org/10.1177/0002716218770138>

Gray, B., Purdy, J. M., & Ansari, S. (2015). From interactions to institutions: microprocesses of framing and mechanisms for the structuring of institutional fields. *Academy of Management Review*, 40(2), 115-143.

<https://doi.org/10.5465/amr.2013.0299>

Greenhalgh, T., & Russell, J. (2009). Evidence-based policymaking: A critique. *Perspectives in Biology and Medicine*, 52(2), 304-318.

Greve, H. R., & Teh, D. (2018). Goal selection internally and externally: A behavioral theory of institutionalization. *International Journal of Management Review*, 20, S19-S38. <https://doi.org/10.1111/ijmr.12138>

Grob, G. F. (2007). The evaluator's role in policy development. In R. Mohan & K. Sullivan (Eds.), *Promoting the use of government evaluations in policymaking*. *New Directions for Evaluation*, 112, 99-108. <https://doi.org/10.1002/ev.211>

Governmentwide Direct-Hire Authority for Scientific, Technical, Engineering and Mathematics (STEM) Positions, (2018). <https://www.opm.gov/policy-data-oversight/hiring-information/direct-hire-authority/#url=Governmentwide-Authority>

Hall, G. E., & Hord, S. M. (2015). *Implementing change: Patterns, principles, and potholes* (4th ed.). Pearson Education, Inc.

Hanberger, A. (2001). Policy and program evaluation, civil, social, and democracy. *American Journal of Evaluation*, 22, 211-228.

Hanneman, R. A., & Riddle, M. (2005). *Introduction to social network methods*. University of California, Riverside.

Hardy, C., & Maguire, S. (2010). Discourse, field-configuring events, and change in organizations and institutional fields: Narratives of DDT and the Stockholm Convention. *Academy of Management Journal*, 53(6), 1365-1392.

Haskins, R. (2018). Evidence-based policy: The movement, the goals, the issues, the promise. *Annals of the American Academy of Political and Social Sciences*, 678, 8-37. <https://doi.org/10.1177/0002716218770642>

Hogan, S. J., & Coote, L. V. (2014). Organizational culture, innovation, and performance: A test of Schein's model. *Journal of Business Research*, 67, 1609-1621. <https://doi.org/10.1016/j.busres.2013.09.007>

- Jennings, J., & Nagel, J. C. (2019). *Federal workforce statistics sources: OPM and OMB* (R43590). Congressional Research Service.  
<https://fas.org/sgp/crs/misc/R43590.pdf>
- Johnson, J. B. (1998). Toward a theoretical model of evaluation utilization. *Evaluation and Program Planning, 21*, 93-110.
- Kadushin, C. (2012). *Understanding social networks: Theories, concepts, and findings*. Oxford University Press.
- Kane, M., & Trochim, W. M. K. (2007a). An introduction to concept mapping. In *Concept mapping for planning and evaluation* (pp. 1-26). SAGE Publications, Inc. <https://doi.org/10.4135/9781412983730>
- Kane, M., & Trochim, W. M. K. (2007b). Using concept mapping in evaluation. In *Concept mapping for planning and evaluation* (pp. 157-174). SAGE Publications, Inc. <https://doi.org/10.4135/9781412983730>
- Kaplan, A. (1964). *The conduct of inquiry: Methodology for behavioral science*. Harvard University Press.
- King, J. A. (1988). Research on evaluation use and its implications for evaluation research and practice. *Studies in Educational Evaluation, 14*, 285-299.

- King, J. A., & Alkin, M. C. (2019). The centrality of use: Theories of evaluation use and influence and thoughts on the first 50 years of use research. *American Journal of Evaluation*, 40(3), 431-458. <https://doi.org/10.1177/1098214018796328>
- King, J. A., & Stevahn, L. (2013). *Interactive evaluation practice: Mastering the interpersonal dynamics of program evaluation*. SAGE Publications, Inc.
- King, J. A., & Stevahn, L. (2015). Competencies for program evaluators in the light of adaptive action: What? So what? Now what? In J. W. Altschuld & M. Engle (Eds.), *Accreditation, certification, and credentialing: Relevant concerns for U.S. evaluators*. *New Directions for Evaluation*, 145, 21-37. <https://doi.org/10.1002/ev.20109>
- Kirkhart, K. (2000). Reconceptualizing evaluation use: An integrated theory of influence. In V. J. Caracelli & H. Preskill (Eds.), *The expanding scope of evaluation use*. *New Directions for Evaluation*, 88, 5-24. <https://doi.org/10.1002/ev.1188>
- Labin, S. N., Duffy, J. L., Meyers, D. C., Wandersman, A., & Lesesne, C. A. (2012). A research synthesis of the evaluation capacity building literature. *American Journal of Evaluation*, 33, 307-338. <https://doi.org/10.1177/1098214011434608>
- LaDuca, A. (1994). Validation of professional licensure examinations. *Evaluation & The Health Professions*, 17, 178-197.
- Lawrenz, F., Kunz Kollmann, E., King, J. A., Bequette, M., Pattison, S., Grack Nelson, A., Cohn, S., Cardiel, C. L. B., Iacovelli, S., Ostgaard Eliou, G., J., G., Causey,

- L., Sinkey, A., Beyer, M., & Francisco, M. (2018). Promoting evaluation capacity building in a complex adaptive system. *Evaluation and Program Planning*, 69, 53-60. <https://doi.org/10.1016/j.evalprogplan.2018.04.005>
- Lemire, S., Fierro, L. A., Kinarsky, A. R., Fujita-Conrads, E., & Christie, C. A. (2018). The U.S. federal evaluation market. In S. B. Nielsen, S. Lemire, & C. A. Christie (Eds.), *The evaluation marketplace: Exploring the evaluation industry*. *New Directions for Evaluation*, 160, 63-80. <https://doi.org/10.1002/ev.20343>
- Lemire, S., Nielsen, S. B., & Christie, C. A. (2018). Toward understanding the evaluation market and its industry—Advancing a research agenda. In S. B. Nielsen, S. Lemire, & C. A. Christie (Eds.), *The evaluation marketplace: Exploring the evaluation industry*. *New Directions for Evaluation*, 160, 145-163. <https://doi.org/10.1002/ev.20339>
- Leviton, L. C. (2014). Some underexamined aspects of evaluation capacity building. *American Journal of Evaluation*, 35(1), 90-94. <https://doi.org/10.1177/1098214013502844>
- Li, Y. (2017). A semiotic theory of institutionalization. *Academy of Management Review*, 42(3), 520-547. <https://doi.org/10.5465/amr.2014.0274>
- Liebman, J. B. (2018). Using data to more rapidly address difficult U.S. social problems. *Annals of the American Academy of Political and Social Sciences*, 675, 166-181. <https://doi.org/10.1177/0002716217745812>

- Lincoln, Y. S., & Guba, E. (2004). The roots of fourth generation evaluation: Theoretical and methodological origins. In M. C. Alkin (Ed.), *Evaluation roots: Tracing theorists' views and influences* (pp. 225-242). SAGE Publications, Inc.
- Loewenstein, J. (2014). Take my word for it: How professional vocabularies foster organizing. *Journal of Professions and Organizations*, *1*, 65-83.  
<https://doi.org/10.1093/jpo.jot004>
- Love, A. J. (1994). Should evaluators be certified? In J. W. Altschuld & M. Engle (Eds.), *The preparation of professional evaluators: Issues, perspectives, and programs*. *New Directions for Evaluation*, *62*, 29-40.
- Mark, M. M. (2001). Evaluation's future: Furor, futile, or fertile? *American Journal of Evaluation*, *22*(3), 457-479.
- Mark, M. M., Cooksy, L. J., & Trochim, W. M. K. (2009). Evaluation policy: An introduction and overview. In W. M. K. Trochim, M. M. Mark, & L. J. Cooksy (Eds.), *Evaluation policy and evaluation practice*. *New Directions for Evaluation*, *123*, 3-11. <https://doi.org/10.1002/ev.302>
- Mark, M. M., & Henry, G. T. (2004). The mechanisms and outcomes of evaluation influence. *Evaluation*, *10*(1), 35-57. <https://doi.org/10.1177/1356389004042326>
- Martin, S. A. (2015). *Social capital at the Capitol: A social network analysis of interest group influence in the 111th Congress* [University of Kentucky]. Lexington, KY.

- Maynard, R. (2018). The role of federal agencies in creating and administering evidence-based policies. *Annals of the American Academy of Political and Social Sciences*, 678, 134-144. <https://doi.org/10.1177/000271628768742>
- Maynard, R., Goldstein, N., & Smith Nightingale, D. (2016). Program and policy evaluations in practice: Highlights from the federal perspective. In L. R. Peck (Ed.), *Social experiments in practice: The what, why, when, where, and how of experimental design & analysis*. *New Directions for Evaluation*, 152, 109-135. <https://doi.org/10.1002/ev.20209>
- McDavid, J. C., & Huse, I. (2015). How does accreditation fit into the picture? In J. W. Altschuld & M. Engle (Eds.), *Accreditation, certification, and credentialing: Relevant concerns for U.S. evaluators*. *New Directions for Evaluation*, 145, 53-69. <https://doi.org/10.1002/ev.20111>
- McKay, S., Murray, M., MacIntyre, S., & Kashyap, A. (2015). Evidence-based policymaking and the public interest: Lessons in legitimacy. *Town Planning Review*, 86(2), 133-154.
- Mertens, D. M. (1994). Training evaluators: Unique skills and knowledge. In J. W. Altschuld & M. Engle (Eds.), *The preparation of professional evaluators: Issues, perspectives, and programs*. *New Directions for Evaluation*, 62, 17-27.
- Merwin, J. C., & Wiener, P. H. (1985). Evaluation: A profession? *Educational Evaluation and Policy Analysis*, 7, 253-259.

- Meyer, R. E., Jancsary, D., Höllerer, M. A., & Boxenbaum, E. (2018). The role of verbal and visual text in the process of institutionalization. *Academy of Management Review*, 43(3), 329-418. <https://doi.org/10.5465/amr.2014.0301>
- Morrill, C. (2008). Culture and organization theory. *Annals of the American Academy of Political and Social Sciences*, 619, 15-40.  
<https://doi.org/10.1177/0002716208320241>
- Muzio, D., Brock, D. M., & Suddaby, R. (2013). Professions and institutional change: Towards an institutionalist sociology of the professions. *Journal of Management Studies*, 50, 699-721. <https://doi.org/10.1111/joms.12030>
- Nigam, A., & Ocasio, W. (2010). Event attention, environmental sensemaking, and change in institutional logics: An inductive analysis of the effects of public attention to Clinton's health care reform initiative. *Organization Science*, 21(4), 823-841. <https://doi.org/10.1287/orsc.1090.0490>
- O'Grady, M. J. (2010). The role of the federal statistical system in evidence-based policymaking or how to make the statistical system essential. *Annals of the American Academy of Political and Social Sciences*, 631, 180-188.  
<https://doi.org/10.1177/0002716210374154>
- O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford University Press.



- Olejniczak, K., Raimondo, E., & Kupiec, T. (2016). Evaluation units as knowledge brokers: Testing and calibrating an innovative framework. *Evaluation*, 22(2), 168-189. <https://doi.org/10.1177/1356389016638752>
- Oliver, C. (1992). The antecedents of deinstitutionalization. *Organization Studies*, 13(4), 563-588.
- Ottoson, J. M., & Martinez, D. (2010). *An ecological understanding of evaluation use*. <https://www.rwjf.org/content/dam/web-assets/2010/10/an-ecological-understanding-of-evaluation-use>
- Pasachoff, E. (2017). Two cheers for evidence: Law, research, and values in education policymaking and beyond. *Columbia Law Reviews*, 117, 1933-1972.
- Patton, M. Q. (1998). Discovering process use. *Evaluation*, 4, 225-233.
- Patton, M. Q. (2015). Misuse: The shadow side of use. In C. A. Christie & A. T. Vo (Eds.), *Evaluation use and decision making in society: A tribute to Marvin C. Alkin* (pp. 131-147). Information Age.
- Pawson, R., Wong, G., & Owen, L. (2011). Known knowns, known unknowns, unknown unknowns: The predicament of evidence-based policy. *American Journal of Evaluation*, 32(4), 518-546. <https://doi.org/10.1177/1098214011403831>
- Peck, J., & Theodore, N. (2015). *Fast policy: Experimental statecraft at the thresholds of neoliberalism*. University of Minnesota Press.

- Però, D. (2011). Migrants' practices of citizenship and policy change. In C. Shore, D. W. Wright, & D. Però (Eds.), *Policy worlds: Anthropology and the analysis of contemporary power* (pp. 244-263). Bergahn Books.
- Preskill, H., & Boyle, S. (2008). A multidisciplinary model of evaluation capacity building. *American Journal of Evaluation*, 29(4), 443-459.  
<https://doi.org/10.1177/1098214008324182>
- Richardson, E. L. (1992). The value of evaluation. *New Directions for Program Evaluation*, 1992(55), 15-20. <https://doi.org/10.1002/ev.1614>
- Rossi, P. H., Lipsey, M. W., & Henry, G. T. (Eds.). (2019). *Evaluation: A systematic approach* (8th ed.). SAGE Publications, Inc.
- Schlauffer, C., Stucki, I., & Sager, F. (2018). The political use of evidence and its contribution to democratic discourse. *Public Administration Review*, 78, 645-649.  
<https://doi.org/10.1111/puar.12923>
- Schwandt, T. A. (2017). Professionalization, ethics, and fidelity to an evaluation ethos. *American Journal of Evaluation*, 38, 546-553.  
<https://doi.org/10.1177/1098214017728578>
- Schwandt, T. A. (2018). Acting together in determining value: A professional ethical responsibility of evaluators. *Evaluation*, 24, 306-317.  
<https://doi.org/10.1177/1356389018781362>

- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (Vol. 39-83). Rand McNally.
- Shadish, W. (1998). Evaluation theory is who we are. *American Journal of Evaluation*, 19(1), 1-19.
- Shore, C., & Wright, S. (2011). Introduction. Conceptualising policy: Technologies of governance and the politics of visibility. In S. Shore, S. Wright, & D. Però (Eds.), *Policy worlds: Anthropology and the analysis of contemporary power* (pp. 1-25). Bergahn Books.
- Sroufe, R. (2017). Integration and organizational change towards sustainability. *Journal of Cleaner Production*, 162, 315-329.  
<https://doi.org/10.1016/j.clepro.2017.05.180>
- Stone, D. (2008). Global public policy, transnational policy communities, and their networks. *Policy Studies Journal*, 36(1), 19-38. <https://doi.org/10.1111/j.1541-0072-2007.00251.x>
- Stone, D. (2012). *Policy paradox: The art of political decision making* (3rd ed.). W. W. Norton & Company, Inc.
- Strang, D., & Meyer, J. W. (1993). Institutional conditions for diffusion. *Theory and Society*, 22, 487-511.

- Strang, D., & Soule, S. A. (1998). Diffusion in organizations and social movements: From hybrid corn to poison pills. *Annual Review of Sociology*, 24, 265-290.
- Tabuga, A. D. (2017). Knowledge utilization in policymaking: Evidence from congressional debates in the Philippines. *Journal of Asian Public Policy*, 10(3), 302-317. <https://doi.org/10.1080/17516234.2016.1219144>
- Taylor-Powell, E., & Boyd, H. H. (2008). Evaluation capacity building in complex organizations. In M. T. Braverman, M. Engle, M. E. Arnold, & R. A. Rennekamp (Eds.), *Program evaluation in a complex organizational system: Lessons from cooperative extension. New Directions for Evaluation*, 120, 55-69. <https://doi.org/10.1002/ev.276>
- Thames, R. C., & Webster, D. W. (2009). *Chasing change: Building organizational capacity in a turbulent environment*. John Wiley & Sons, Inc.
- The White House. (n.d.). *The Cabinet*. <https://www.whitehouse.gov/the-trump-administration/the-cabinet/>
- Tichy, N. M., Tushman, M. L., & Fombrun, C. (1979). Social network analysis for organizations. *The academy of Management Review*, 4(4), 507-519.
- Ting, M. M. (2009). Organizational capacity. *The Journal of Law, Economics, & Organization*, 27, 245-270. <https://doi.org/10.1093/jleo/ewp021>

- Tolman, E. C., Ritchie, B. F., & Kalish, D. (1946). Studies in spatial learning. *Journal of Experimental Psychology*, 36, 13-24.
- Trochim, W. M. K. (1989). An introduction to concept mapping for planning and evaluation. *Evaluation and Program Planning*, 12, 1-6.
- Trochim, W. M. K. (1998). An evaluation of Michael Scriven's "Minimalist theory: The least theory that practice requires". *American Journal of Evaluation*, 19(2), 243-249.
- Trochim, W. M. K. (2009). Evaluation policy and evaluation practice. In W. M. K. Trochim, M. M. Mark, & L. J. Cooksy (Eds.), *Evaluation policy and evaluation practice. New Directions for Evaluation*, 123, 13-32.  
<https://doi.org/10.1002/ev.303>
- Trochim, W. M. K. (2017). Hindsight is 20/20: Reflections on the evolution of concept mapping. *Evaluation and Program Planning*, 60, 176-185.  
<https://doi.org/10.1016/j.evalprogplan.2016.08.009>
- Trochim, W. M. K., & McLinden, D. (2017). Introduction to a special issue on concept mapping. *Evaluation and Program Planning*, 60, 166-175.  
<https://doi.org/10.1016/j.evalprogplan.2016.10.006>
- U.S. General Accounting Office. (2003). *Ethnographic studies can inform agencies' actions*. (GAO-03-455). Washington, DC: U.S. General Accounting Office.

- U.S. Office of Management and Budget. (2019). *Phase 1 implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Learning agendas, personnel, and planning guidance*. (M-19-23). Washington, DC: Executive Office of the President.
- usa.gov. (2019). *Branches of the U.S. Government*. U.S. Government.  
<https://www.usa.gov/branches-of-government>
- Warrick, D. D. (2017). What leaders need to know about organizational culture. *Business Horizons*, 60, 395-404. <https://doi.org/10.1016/j.bushor.2017.01.011>
- Wedel, J. R. (2001). Some enchanted era. In *Collision and collusion: The strange case of western aid to Eastern Europe* (pp. 1-13). Palgrave.
- Wedel, J. R. (2004). Studying through a globalizing world: Building method through aidnographies. In J. Gould & H. Secher Marcussem (Eds.), *Ethnographies of aid-exploring development texts and encounters*. Roskilde University.
- Wedel, J. R. (2009). When privatization meets truthiness. In *Shadow elite* (pp. 23-45). Basic Books.
- Wedel, J. R. (2014). *Unaccountable: How the establishment corrupted our finances, freedom, and politics and created an outsider class*. Pegasus.
- Wedel, J. R. (2019). *The mapping method: A guide to charting corruption and influence processes* [Working paper forthcoming]. Global Integrity.

Worthen, B. R. (1999). Critical challenges confronting certification of evaluators.

*American Journal of Evaluation*, 20(3), 533-555.

Wright, S., & Reinhold, S. (2011). "Studying through": A strategy for studying political transformations. Or sex, lies and British politics. In C. Shore, S. Wright, & D.

Però (Eds.), *Policy worlds: Anthropology and the analysis of contemporary power* (pp. 86-104). Bergahn Books.

Wutzke, S., Roberts, N., Willis, C., Best, A., Wilson, A., & Trochim, W. M. K. (2017).

Setting strategy for system change: using concept mapping to prioritise national action for chronic disease prevention. *Health Research Policy and Systems*, 15(69). <https://doi.org/10.1186/s12961-017-0231-7>

Zilber, T. B. (2002). Institutionalization as an interplay between actions, meanings, and

actors: The case of a rape crisis center in Israel. *Academy of Management Journal*, 45(1), 234-254.

Zucker, L. G. (1977). The role of institutionalization in cultural persistence. *American*

*Sociological Association*, 42(5), 726-743.

## **The Diffusion of Value in Evaluation by the U.S. Federal Government Through the *Analytical Perspectives of the U.S. President's Budget***

When evaluation was first introduced in the U.S. Federal Government in the 1960s, it was very much a novel idea but has since undergone several shifts in practice and utilization to contribute to government efficiency and effectiveness (Lemire, Fierro, et al., 2018). Most recently, the Foundations for Evidence-Based Policymaking Act (P.L. 115-435)—commonly referred to as the Evidence Act—was enacted in January 2019, which now mandates evaluations and the appointment of a Chief Evaluation Officer (CEO) to lead a centralized office within each federal agency to coordinate evaluation activities. Although the Evidence Act codifies value in federal evaluation, it is not clear how this value chain was promoted or which levers of change were effective in driving institutional change in the government over time. Therefore, the purpose of the present study was to identify the pace and patterns in the process of diffusing value in evaluation in the U.S. Federal Government. This is the second in a series of three manuscripts investigating the longitudinal process of institutionalizing evaluation in the U.S. Federal Government.

### **Increasing Value in Evaluation**

**Conceptualizing the process.** Institutionalizing a novel idea in complex organizations involves a complicated process of first diffusing the idea widely and then establishing its permanency as a cultural norm (Colyvas & Jonsson, 2011). More simply, Colyvas and Jonsson (2011) separates diffusion (i.e., ubiquity) and institutionalization (i.e., legitimacy) whereby, “diffusion emphasizes the pace and pattern of the object that



spreads [and] institutionalization underscores depth and durability,” (p. 45). This process occurs by transforming an exterior concept through microlevel actions toward an alignment with subjective understandings and changes in macrolevel structures (Li, 2017; Zucker, 1977). The careful interplay between actions, meanings, and actors requires intentionality and coordination (Zilber, 2002), but organizations choose to pursue these complicated processes because the birth of new markets to diversify services is essential to maximizing their efficiency and effectiveness (Haveman, 1993).

Evaluations play a critical role in providing evidence to support complex policymaking decisions—the process known as evidence-based policymaking (Haskins, 2018; Manson, 2016; Maynard et al., 2016; Pasachoff, 2017; Pawson, 2006; Schlauffer et al., 2018; Weiss, 1988). It is, therefore, not surprising that evaluation was mandated in an Act intended to build a foundation in evidence-based policymaking practices. Although “evidence-based policymaking” is now a buzz word that is commonly heard across all governmental frequencies, the spotlight on evidence (or on evaluation as an integral part of the policymaking structure) has not always been the norm. The evolution of that priority has been a result of establishing a value of and in evaluation. The first manuscript in this series distinguished between value *of* evaluation as a statement of worth and value *in* evaluation as an indication of belief or investment in that worth (Nolton et al., 2020b). Therefore, the process of making evaluation ubiquitous in the Federal Government depends on the establishment of its worth and a pattern of investment in evaluation (conceptualized in Figure 3).

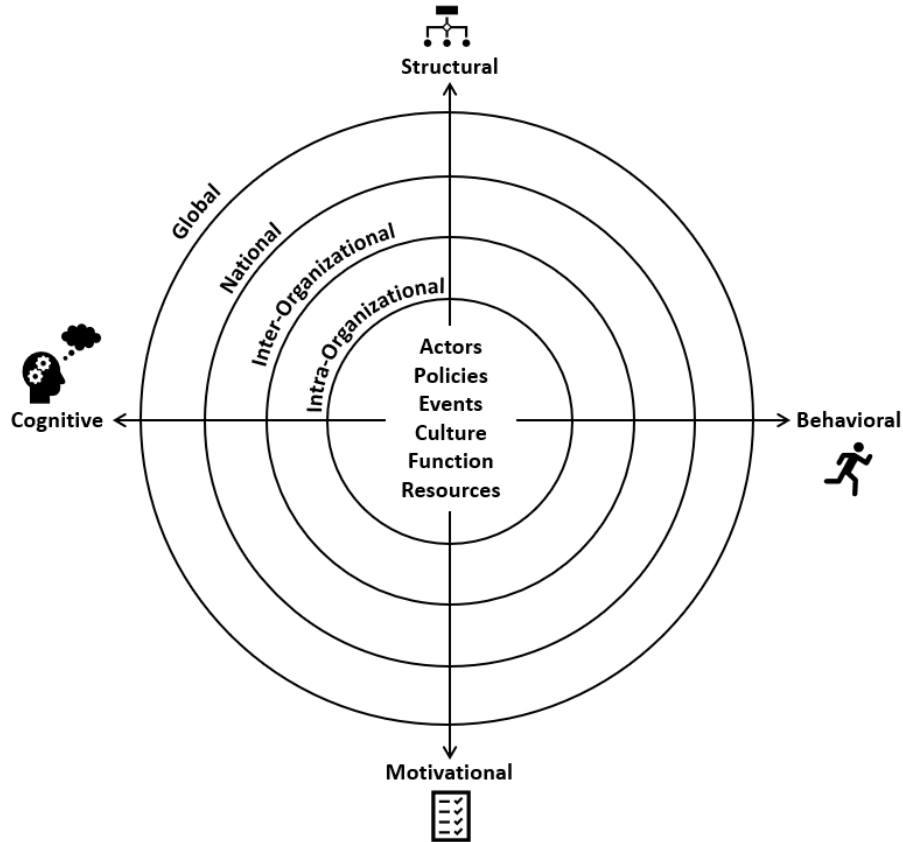


Figure 3. Revisiting the Conceptual Framework for Diffusing Value in Evaluation

*Note.* This framework illustrates the ecological system of diffusing value in evaluation. It captures the complexity of the U.S. Federal Government and breadth of factors that impact diffusion, which may influence cognition, motivation, behavior, or structure.

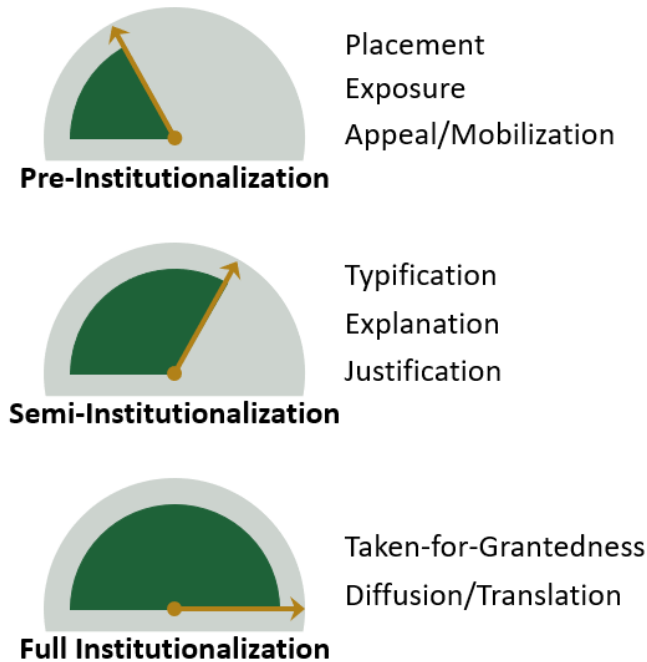
The proposed model in Figure 3 is used to conceptualize the process of diffusing value in evaluation. This framework accounts for the complexity of the Federal Government as a nested ecosystem with internal and external layers that influence organizational structure and culture. Central to this process are many different factors that might influence how value in evaluation is generated and distributed. Adapted from the Mark and Henry (2004) model of evaluation influence, these factors also have different

types of influences: cognitive (thinking and feeling), motivational (valuing), behavioral (doing), and structural (being). Although it is assumed that this framework appropriately portrays the process of diffusing value in evaluation, it is not clear which mechanisms have been used by the U.S. Federal Government to achieve this process.

Theorists who study institutional change agree that diffusion is a precursor to institutionalization and that the process can be further compartmentalized into different categories, but there are many iterations of how those categories are named and ordered (Armenakis & Harris, 2002; Brown et al., 2012; Colyvas & Jonsson, 2011; Green, 2004; Strang & Meyer, 1993). Berger and Luckman (1967) presented a three-step model which described the movement of a novel idea from a shared understanding that is constructed by symbols (i.e., externalization), its legitimacy as a shared reality (i.e., objectification), to socialization of the idea (i.e., internalization) through shifts in cognition (as cited by Colyvas & Johnson, 2011, p. 40). Whereas this model ends with an influence on cognition, other models suggest that changing how organizational actors think and feel about a new idea is the initial phase of institutionalization.

The model that most closely resembles the types of influences proposed for this study is a three-part model presented by Armenakis and Harris (2002) whereby institutional change begins with a readiness for the uptake of a new idea followed by adoption and then institutionalization of the new idea. This model relates stages of institutionalization to organizational maturity and capacity for institutional change. Furthermore, Meyer et al. (2018) explained that text could be used to express varying degrees of institutionalization for which they defined as eight levels across three discrete

stages (Figure 4): pre-institutionalization (i.e., similar to Armenakis & Harris' stage of readiness), semi-institutionalization (i.e., adoption), and full institutionalization.



*Figure 4.* Stages of Institutionalization Framework

*Note.* This figure depicts a framework for how stages of institutionalization may be conveyed through text, as presented by Meyer et al. (2018).

**Mapping the process.** Also introduced in the first paper of this series was an ethnographic mapping method to study policy and government processes (Nolton et al., 2020b; Wedel, 2019). Among the many potential data sources that inform this mapping method (i.e., observations, interviews, documents, policies, etc.), artifacts (i.e., tangible documents and records) are emphasized in this paper as essential resources to understand government processes (Nolton et al., 2020b). Artifacts are instrumental in transforming

new institutional logics—organizing principles that are socially constructed to institutionalize practices—from tacit to explicit knowledge (D'Eredita & Barreto, 2006; Friedland & Alford, 1991; Nigam & Ocasio, 2010). Green (2004) described the use of rhetoric in diffusion as a means to “gather information, develop shared understandings of the world, and persuade individuals to contribute to collective purposes, such as the adoption and implementation of new practices,” (p. 654). Therefore, language used to proliferate knowledge can be a powerful tool used to influence perceptions of new logics (Brown et al., 2012; Colyvas & Jonsson, 2011; Suddaby & Greenwood, 1995).

Although the full mapping method is a comprehensive approach to understanding government processes, this manuscript will only highlight the analysis of government documents in effort to answer the following guiding questions:

1. Who have been the **parties** (i.e., players, social networks, and organizations) involved, both formally and informally, involved in the institutionalization of federal evaluation?
2. What have been the parties' respective interests, **agendas**, incentives, goals, motivations, operating assumptions, and expectations with regard to institutionalizing federal evaluation?
3. Who have been the parties' **sponsors** (i.e., funders or providers of other indispensable resources) and what are their interests, agendas, incentives, goals, motivations, operating assumptions, and expectations with regards to federal evaluation activities?
4. What has been the larger context(s)/circumstances of **power** and **resources** in which the parties are embedded? How have features of the larger context(s)/circumstances constrain, enable, or otherwise affect the parties involved in federal evaluation efforts?
5. To what extent has the **structure** that facilitates federal evaluation activities or efforts institutionalized (or becoming so) even if it remains informal/unincorporated and the players change over time?

6. What has been the **ecosystem** in which the various parties operate and how does it affect federal evaluation efforts? What has been the “culture(s)” and mindsets of the milieus, organization(s), or venues in which the parties operate (i.e., norms and mindsets of a community of practice, norms and ethics of a profession) and how do these affect federal evaluation efforts?

This present manuscript will demonstrate the rigor and benefit of these analyses in providing a glimpse at the process of diffusing value in evaluation. However, limitations will also be acknowledged, which justify the need for additional sources of data to inform mapping influential factors, which will be explored in the third manuscript of this series.

## **Methods**

**Analyzing content and discourse.** Centralized government documents shape how policymakers, private citizens, and other stakeholders interact with social issues by using linguistic features to build meaning (i.e., signs or tasks) around institutional logics (Cabatoff, 2000; Gee, 2014; Li, 2017; Loewenstein, 2014). These features may include non-verbal, gestural, grammatical, or syntactical nuances that are hidden or excluded, but implied, to create or dissemble realities (Gee, 2011, 2014; Krippendorff, 2019; Peräkylä & Ruusuvuori, 2018; Rapley, 2007). Therefore, language (i.e., either descriptive or normative) is pivotal to conveying past, present, and future ideologies about evaluation practice (Mark & Henry, 2004; Patton, 2000). Analyzing the content and discourse of central government documents can uncover the presence, persuasion, and progression of value in evaluation.

Methodologically, content analyses can be done quantitatively where frequencies reflect content prevalence (Maier, 2017; Maxwell, 2010; Neuendorf, 2017). Additionally, longitudinal document metadata (i.e., length, titles, order, author, etc.) can inform content

decisions. Qualitatively, content analyses can produce thematic patterns in the text, while discourse analyses investigate discursive and non-discursive linguistic features that construct the information, action, and identity of institutional logics (Gee, 2011; Krippendorff, 2019; Phillips et al., 2004; Rapley, 2007). Pairing these two types of analyses together provides an approach to observe what, how, and why language was or was not used to diffuse value in evaluation (Rapley, 2007). The following sections describe the rationale for selecting documents for analysis and the analytic methods used to identify trends in the language of the selected documents.

**Document selection.** One way in which the government demonstrates value in rigorous evidence is by investing resources toward greater capacity for evaluation activities (Lemire, Nielsen, et al., 2018). In preparation for each upcoming fiscal year (FY), the U.S. Office of Management and Budget (OMB) publishes the President's Budget which projects funding allocations across the entire Federal government. These allocations reflect the priorities and agenda of that Administration. As such, the OMB plays a key role in directing evidence-based policymaking efforts in the Federal government (Stack, 2018). Research shows an increase in support for evaluation by observing federal spending on evaluation over time (Kinarsky, 2018; Lemire, Fierro, et al., 2018; Rizvi, 2004); however, shifts in the .

Each President's Budget includes supplementary documents that append elaborations, tables, and a detailed list of appropriations. The narratives are compiled in a separate document titled *Analytical Perspective*, which further explicate the rationale and analyses that informed the changes to the proposed budget. These narratives are structural

artifacts that codify trends toward isomorphic beliefs, trust, and attitudes about evaluation (Shore & Wright, 2011). Exploring “the relationships between policy texts and their historical, political, social, and cultural contexts” (Taylor, 2004, p. 435) helps to understand what triggered and motivated shifts in government discourse (Alkin & King, 2016; Nolton et al., 2020b; Shore & Wright, 2011; Wedel, 2019).

For the present study, the President’s Budget for fiscal years (FY) 1996-2020 were publicly available on the OMB website<sup>1</sup>. Although the *Analytical Perspectives* have supplemental materials that describe budget allocations for individual agencies, the majority of the document each year explains policy decisions made for the entire Federal government. In their effort to justify budget decisions, the OMB effectively utilizes persuasive language to communicate and influence value systems (Gee, 2014; van Leeuwen, 2007). The full *Analytical Perspectives* documents were downloaded from the OMB website in August 2019 as portable document format (PDF) files from FY 1996 through FY 2020. Table 3 displays pertinent metadata for the *Analytical Perspectives* including the fiscal year; the U.S. Presidential Administration at the time the Budget was published; the number of chapters of the document; and total number of pages.

Each year, the documents consisted 24-33 chapters dedicated to covering economic assumptions, areas of investments, receipts and expenditures, technical analyses, and other special topics arranged to explicate the reasoning for the proposed budget. Beginning in FY 2011, the *Analytical Perspectives* featured a distinct chapter on evaluation. These individual chapters were separately downloaded to explore how

---

<sup>1</sup> Office of Management and Budget. Washington DC. <https://www.whitehouse.gov/omb/budget/>



language was used to describe and normalize evaluation. Furthermore, an analysis of document metadata was conducted to observe presence, titling, position, and length of the chapters featuring evaluation to observe other characteristics of the text that may also influence the content. The next section provides a full account of the coding and analytic procedures and frameworks used to analyze these documents.

*Table 3.*

Metadata for the Analytical Perspectives Documents

<b>Fiscal Year</b>	<b>Administration</b>	<b>No. of Chapters</b>	<b>No. of Pages</b>
1996	Bill Clinton	27	472
1997	Bill Clinton	27	518
1998	Bill Clinton	26	559
1999	Bill Clinton	26	590
2000	Bill Clinton	25	592
2001	Bill Clinton	25	664
2002	George W. Bush	26	664
2003	George W. Bush	27	706
2004	George W. Bush	26	745
2005	George W. Bush	28	408
2006	George W. Bush	29	442
2007	George W. Bush	28	410
2008	George W. Bush	28	424
2009	George W. Bush	28	424
2010	Barack Obama	25	430
2011	Barack Obama	31	478
2012	Barack Obama	31	498
2013	Barack Obama	33	510
2014	Barack Obama	30	508
2015	Barack Obama	27	430
2016	Barack Obama	27	406
2017	Barack Obama	27	418
2018	Donald Trump	24	292
2019	Donald Trump	25	338
2020	Donald Trump	28	360

*Note.* This table displays the metadata for the *Analytical Perspectives* documents of the U.S. President’s Budget for each fiscal year (FY) between 1996 to 2020. FY 2011 was the first document to include a separate chapter dedicated to evaluation.

**Coding and analytic procedures.** All documents were imported and analyzed in ATLAS.ti 8.4.24.0 for Windows (ATLAS.ti Scientific Software Development GmbH; Berlin, Germany, 2020). ATLAS.ti is a qualitative data analysis (QDA) instrument, which allows data to be coded, linked, and visualized. Using a robust software for QDA is helpful to analyze multiple sources of data of varying formats (i.e., text, video, audio, image files) and/or large quantities of data. As shown in Table 3, the *Analytical Perspectives* are each several hundreds of pages, totaling over 12,000 pages of content in the last 25 years. The study utilized an automated coding technique to search for key terms and an open coding approach to ground and refine the proposed theory of diffusing value in evaluation. These coding methods are described in the following sections.

**Coding for prevalence.** A content analysis of the full documents was conducted to capture trends in word prevalence over time. The auto-coding function in ATLAS.ti was used to search for select terms across the *Analytical Perspectives*. The words that were searched were "evaluat\*", "analy\*", "statistic\*", "evidence", and "evidence-based". These words were selected because they were identified in the literature as key terms related to the use and practice of evaluation (Maynard et al., 2016; National Academies of Sciences, 2017; U.S. General Accounting Office, 1990). These texts repeatedly classify evaluation as an analytic activity performed to inform evidence-based policy decisions (Maynard et al., 2016). However, it is one of many empirical analytic activities conducted in the Federal government, including statistical analyses (U.S. General Accounting Office, 1990). Statistical principles and practices have long been established

in the Federal government and were used as a template to define similar principles and standards of evaluation practice. (National Research Council, 1992).

*Table 4.*

Counts of Word Prevalence for Searched Terms

<b>Fiscal Year</b>	<b>Evaluat*</b>	<b>Analy*</b>	<b>Statistic*</b>	<b>Evidence</b>	<b>Evidence-Based</b>	<b>Total</b>
1996	56	66	27	3	0	152
1997	60	94	49	9	0	212
1998	54	93	77	3	0	227
1999	67	115	86	6	0	274
2000	66	111	88	4	0	269
2001	72	112	98	3	0	285
2002	86	117	85	5	0	293
2003	88	128	101	7	0	324
2004	95	123	99	9	0	326
2005	60	110	106	12	0	288
2006	63	123	111	10	0	307
2007	68	123	121	7	0	319
2008	59	116	113	6	1	295
2009	61	124	120	4	1	310
2010	45	94	83	2	3	227
2011	117	148	119	29	6	419
2012	129	175	114	29	7	454
2013	96	237	124	43	7	507
2014	125	200	135	71	14	545
2015	133	137	142	82	19	513
2016	74	127	219	47	8	475
2017	190	120	171	110	8	599
2018	73	84	152	60	3	372
2019	117	92	172	73	13	467
2020	169	109	188	71	11	548

*Note.* \* denotes a wild card indicator used in the search process to allow for variants of root words to be searched in the *Analytical Perspectives* documents.

Asterisks were used in search terms as wild-cards to include variants of the words (e.g., evaluate, evaluation, evaluator; analysis, analyst, analytical; statistics, statistician, statistical). The researcher allowed ATLAS.ti to apply codes directly, but then manually

cleaned codes to exclude inappropriate data. For example, the word "ANALYTICAL", when written in all caps, was typically on the document title page (i.e., "ANALYTICAL PERSPECTIVES") and headers of each page. Also, any codes including a slash (i.e., "/") was indicative of a web address, which was often a footnote for reference. For example, <http://www.whitehouse.gov/omb/evidence> was listed as a reference to provide readers with additional information. The final counts of these terms were then filtered by document, which provided a table displaying frequencies of the words by year (Table 4).

*Coding for evidence of institutionalization.* A grounded theory approach was used to identify evidence of diffusing value in evaluation (Charmaz et al., 2018; Patton, 2015b). This methodological approach provides an opportunity to revisit and revise previously formulated theories of the process using empirical data. The evaluation chapters of the *Analytical Perspectives* were open coded (i.e., the initial phase of coding used in a grounded theory approach) to capture emergent themes (Charmaz et al., 2018; Patton, 2015b; Peräkylä & Ruusuvuori, 2018). In order to observe the mechanisms and manifestations of institutionalization in the text, the Meyer et al. (2018) framework (shown in Figure 4) was adopted as a coding scheme to identify the position of emergent codes in the process of institutionalization. Aligning the codes to this framework provides a method of quantifying qualitative data to visualize trends in progression (Patton, 2015b). The following are the definitions used for each of the categories (adapted from Meyer et al., 2018, p.401):

- **Placement** – aligning evaluation relationally to established field meanings
- **Exposure** – making evaluation accessible to the target audience
- **Appeal/Mobilization** – making evaluation attractive and relevant
- **Typification** – generalizing the meaning and applicability of evaluation

- **Explanation** – creating a link between evaluation and a relevant problem
- **Justification** – creating legitimacy for evaluation
- **Taken-for-Grantedness** – establishing and maintaining evaluation as an essential function
- **Diffusion/Translation** – making evaluation communicable across all disciplines

*Establishing rigor and validity.* There were many different techniques employed in this study to ensure methods were rigorous and credible (Morse, 2018; Shenton, 2004). The use of a robust QDA platform increased internal validity (i.e., credibility) of the coding process compared to manually coding by hand. The study was designed by a researcher who: has expert knowledge about the topic and is without strong political leanings; built her qualitative research skills in conducting these analyses; carefully selecting documents for inclusion; grounded her coding technique to an existing theoretical framework and revisited theories by grounding them in empirical data; thoroughly documented her methods and interpretations; and consulted peers to scrutinize the process. The following sections further describe these efforts to establish validity.

Prior to conducting this study, the author had worked in an evaluation unit of a Federal agency and studied research methods, evaluation, and public policy. Academic and professional training in these areas allowed the researcher to design a robust qualitative study with subject matter expertise about the intricacies of the field. Through these experiences, the researcher recognized the importance of the U.S. President's Budget and *Analytical Perspectives* in demonstrating value in evaluation (i.e., investment in the worth of evaluation). As such, in addition to being charged with providing guidance for the Evidence Act, OMB has consistently played a large role in driving past, present, and future use of evaluation in the Federal government (Stack, 2018).

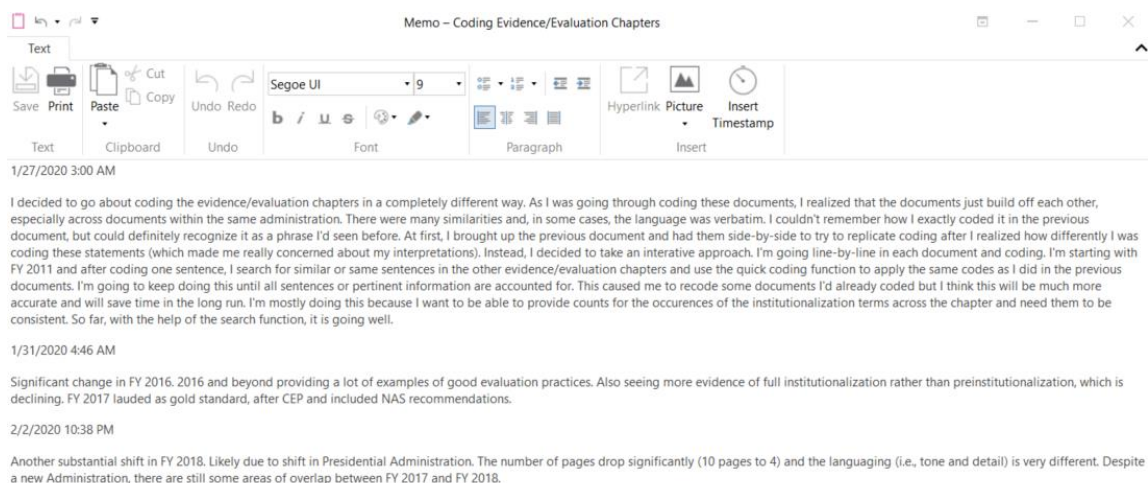
Other researchers had done early work to conduct budget analyses for funding evaluation, which inspired further analyses of the corresponding *Analytical Perspectives* to understand the rationale and motivation behind budget decisions. The current researcher was also intentional about including all available *Analytical Perspectives* (i.e., 25 years), which afforded an opportunity to observe longitudinal trends. Analyses of the text in the evaluation chapters, however, minimized irrelevance of other sections of the full documents to allow for focused analyses of text related to evaluation. The use of the Meyer et al. (2018) framework to analyze evidence on institutionalization in text grounded the coding technique in theory. The relevance of this theory was compared to other frameworks proposed by different researchers and was deemed most appropriate for this study because of the unique detail in categorizing stages of institutionalization and its deliberate intention for use to analyze text. The open coding for thematic analyses provided a method for the researcher to revisit existing theories and ground them with empirical data. This inspired a modification of the framework used as a guide to understand the reality of the institutionalization process in the U.S. Federal Government.

Given the exploratory coding method used to assign organic and open codes, it would have been unlikely for another coder to be able to replicate the codes to calculate intercoder reliability. Therefore, validity of the coding methods was established through peer debriefing and consulting critical friends throughout the coding and analytic process (Patton, 2015b). This was done with a qualitative research methodologist and a peer scholar who are both evaluators, knowledgeable of the topic, highly experienced in qualitative research, and proficient in using ATLAS.ti for QDA (Barber & Walczak,

2009; Patton, 2015b; Spall, 1998). The researcher also presented the project at varying stages to other colleagues who provided additional feedback (Shenton, 2004).

Meetings with the qualitative methodologist and peer debriefer (i.e., four meetings each), were conducted separately and approximately every three weeks during the coding and analytic process. The researcher recorded notes about methods, thought processing, and developing interpretations using memos in ATLAS.ti as a researcher journal (Figure 5). Whereas hand-written notes were taken in the field during design and data collection phases, these memos in ATLAS.ti could be linked within the system to features within the platform. Additional comments for the documents, quotations, codes, and code groups were also documented in ATLAS.ti to explicate decision-making processes. For example, after coding the first three years of the evaluation chapters, the coder determined that the way in which coding was done needed to be changed.

As shown in Figure 5, documents were originally coded one at a time, chronologically, but it was discovered that there were many similarities—some verbatim—between documents. Upon further review, there were several chunks of text (i.e., whole paragraphs or pages) that were duplicated, which suggested that the evaluation chapters were often revisions of preceding *Analytical Perspectives* and not derived from scratch each year, especially within the same Administration. In order to preserve the consistency of coding, the coder switched to a method of coding one sentence (or a string of related sentences) at a time then reapplied codes to the same blocks of text found in subsequent chapters. The search and quick coding functions in ATLAS.ti allowed the coder to do this.



*Figure 5. Example of Memo Documented in ATLAS.ti*

*Note.* Memos were used as a researcher journal to document code development and refinement, the analytic process, and preliminary findings or interpretations.

During peer debriefing meetings, the coder would discuss her methods, findings, and interpretations as they were developing. Most frequently, peer debriefers were used as ‘critical friends’ who would help discuss and strategize the grounded theory approach (Barber & Walczak, 2009; Patton, 2015b). Critiques and questions also offered a way to triangulate the analyst’s process and interpretations (Flick, 2018; Patton, 2015b).

Examples of the unitization (i.e., determining text segments for assigning codes) of quotations for coding and the rationale for sorting and organizing codes were discussed with the peer debriefers (Campbell et al., 2013). Table 5 depicts select examples of text excerpts and codes used for the different institutionalization categories: placement, exposure, typification, explanation, justification, taken-for-grantedness, and diffusion/translation.



Table 5.

Selected Grounded Theory Coding Examples for Institutionalization

Chapter Title (Fiscal Year)	“Program Evaluation” (2012)
Original text excerpt	“Evaluation is one component of the evidence infrastructure that plays a role in a wide range of decision making. The best government programs embrace a culture where performance measurement and evaluation are regularly used and complement one another.”
Thematic Codes	Evaluation is one component of a strong “evidence infrastructure” Performance measurement can only take you so far in answering questions about impact Other analytic methods are complementary to evaluation
Institutionalization Category	Placement
Category Definition	Aligning evaluation relationally to established field meanings.
Chapter Title (Fiscal Year)	“Program Evaluation and Data Analytics” (2013)
Original text excerpt	“In addition, an inter-agency working group is beginning to share best practices across the Federal Government and to discuss issues, such as how to do a better job disseminating evidence of what works, integrating cost-effectiveness analysis into evaluations, and making better use of administrative data for evaluation and other data analytics purposes.”
Thematic Codes	Inter-agency working group to promote stronger evaluations in the Federal government OMB to provide flexible guidance on implementing strong research designs
Institutionalization Category	Exposure
Category Definition	Making evaluation accessible to the target audience.
Chapter Title (Fiscal Year)	“Program Evaluation and Data Analytics” (2014)
Original text excerpt	“It is a culture that values rapid, operationally-focused experiments that can quickly boost program efficiency, effectiveness and customer service, while at the same time equally valuing longer-term evaluation focused on more fundamental questions about program strategy.”
Thematic Codes	“Doing so allows us to more effectively serve the American people.” Improve government efficiency and effectiveness

Institutionalization Category	Pursuit of stronger evidence and better results by emphasizing scientific rigor
Category Definition	Evaluation help to better serve Americans Appeal/Mobilization Making evaluation attractive and relevant.
Chapter Title (Fiscal Year)	“Program Evaluation and Data Analytics” (2015)
Original text excerpt	“OMB and Federal agencies are working together to develop common standards and guidelines for research and evaluation, i.e., ‘common evidence standards’.”
Thematic Codes	Develop robust “what works” repositories across a wide range of programs Developing evidence and evaluation standards
Institutionalization Category	Typification
Category Definition	Generalizing the meaning and applicability of evaluation.
Chapter Title (Fiscal Year)	“Program Evaluation” (2011)
Original text excerpt	“One of the challenges to doing evidence-based policy making is that sometimes it is hard to say whether a program is working well or not. Historically, evaluations have been an afterthought when programs are designed—and once programs have been in place for a while it can be hard to build a constituency for a rigorous evaluation.”
Thematic Codes	Evaluations should be part of the program design from the start Evaluations contribute new knowledge to inform policy decisions
Institutionalization Category	Explanation
Category Definition	Creating a link between evaluation and a relevant problem.
Chapter Title (Fiscal Year)	“Program Evaluation and Data Analytics” (2015)
Original text excerpt	“By instilling a culture of learning into Federal programs, the Administration will build knowledge so that spending decisions are based not only on good intentions, but also on strong evidence that yield the highest social returns on carefully targeted investments.”
Thematic Code	Evaluations justify and maximize government investments
Institutionalization Category	Justification
Category Definition	Creating legitimacy for evaluation
Chapter Title (Fiscal Year)	“Building and Using Evidence to Improve Government Effectiveness” (2018)

Original text excerpt	“Build or support independent evaluation offices to conduct rigorous, independent evaluations.”
Thematic Codes	Increasing investment in evaluation Rigorous and credible evaluations abide by coordinated principles and processes
Institutionalization Category	Taken-for-Grantedness
Category Definition	Establishing and maintaining evaluation as an essential function.
Chapter Title (Fiscal Year)	“Building and Using Evidence to Improve Government Effectiveness” (2020)
Original text excerpt	“These evaluations provide an opportunity to test different program strategies against one another or a control group, and enable an agency to go beyond questions of overall program effectiveness and consider the effects of variations in program approaches, including which services are offered and how.”
Thematic Codes	Evaluations address organizational learning by posing and answering strategic questions Agencies are allowed and encouraged to test innovative strategies
Institutionalization Category	Diffusion/Translation
Category Definition	Making evaluation communicable across all disciplines.

*Note.* This table shows an example of one text excerpt (i.e., quotation) for each of the categories of institutionalization (as described by Meyer et al., 2018) from the *Analytical Perspectives* evaluation chapters. Excerpts could have more than one open code assigned, and the codes were sorted into institutionalization categories as code groups.

## Results and Interpretations

**Document metadata.** The organization of the *Analytical Perspectives* yielded interesting findings about the arrangement of the evaluation chapters. These findings were noted by recording features of the document metadata (Table 6). Although the practice of evaluation has been present in the U.S. Federal Government since the mid-1960s (Lemire, Fierro, et al., 2018), FY 2011—during President Barack Obama’s first term—was the first time there was a chapter dedicated to “Program Evaluation”. Despite

its long-standing function in the government, it took nearly 50 years for evaluation to be prioritized as an intentional investment in the President’s Budget. Since appearing in the FY 2011 *Analytical Perspectives*, evaluation has held a place as an individual chapter in the President’s Budget, but the characteristics and focus of the chapter has varied slightly with each iteration of the document.

For both FY 2011 and 2012, the title of this chapter was “Program Evaluation”, but then this changed to “Program Evaluation and Data Analytics” for the subsequent three fiscal years (FY 2013-2015). In FY 2016, the first shift from evaluation to a focus on evidence was made and this was the only year that dedicated the chapter to discussing the role of administrative data on evidence-based policymaking. For President Obama’s last President’s Budget (i.e., during in his second term), the title in FY 2017<sup>2</sup> changed to, “Building the Capacity to Produce and Use Evidence” to be inclusive of evaluations, administrative data, statistics, observations, implementation studies, and performance measurement as facets of the government’s evidence infrastructure. Beginning in FY 2018, President Donald Trump’s Administration renamed this chapter to, “Building and Using Evidence to Improve Government Effectiveness” shifting the focus again from capacity to evidence use for the purpose of improving government.

In addition to chapter titles, the position of these chapters also changed over time. In FY 2011, the “Program Evaluation” chapter was first placed as the eighth chapter in the *Analytical Perspectives*. In FY 2015—not in alignment with a title change which had

---

<sup>2</sup> The development, authorization, appropriation, and publication of the President’s Budget occurs over approximately 12 months and can be delayed if Congress cannot arrive at an agreement about budget decisions. Thus, although President Donald Trump was elected in November 2016, he was inaugurated in January 2017 and the first President’s Budget published under his Administration was for FY 2018.

occurred two years prior—“Program Evaluation and Data Analytics” was elevated to be the seventh chapter overall. In FY 2018, the transition of Administration and chapter title aligned with moving the “Building and Using Evidence to Improve Government Effectiveness” chapter to be sixth in order. The ordering of chapters could be equated to the primacy and, therefore, preeminence of the topics that are featured earlier rather than later (Lund, 1925). Previous research has found that the ordering of information in communication has influenced decision-making processes because of an observed primacy effect in persuasion (Bansback et al., 2014; Lana, 1963; Lund, 1925).

Furthermore, each chapter provides rationale fundamental to decisions explained in later chapters, which suggests that the content in earlier chapters are foundational to succeeding chapters and allows the reader to believe it is of primary importance above other chapters (Lund, 1925). For example, studies found that jurors could be influenced to decide in favor of the party that argued first (Lawson, 1968; Stone, 1969); patients felt that treatments listed first were of primary importance (Bansback et al., 2014); and consumers tend to select businesses listed earlier in the Yellow Pages (Lohse, 1997). Lastly, as the chapters were elevated in position, the number of pages also increased with each version of the chapter within an Administration. Chapter length denotes the density of the content. The amount of real estate dedicated to cover a topic relays the extent of importance (and need for clarity) for authors to spend increasingly greater time and energy to provide the information. With the exception of a change in Administration, this chapter continues to grow each year, only adding to the discourse and prioritization of evaluation and evidence in the Federal government.

Table 6.

Metadata for Analytical Perspectives Evaluation and Evidence Chapters

<b>Fiscal Year</b>	<b>Administration</b>	<b>Chapter Title</b>	<b>Chapter No.</b>	<b>No. of Pages</b>
2011	Barack Obama	Program Evaluation	8	2
2012	Barack Obama	Program Evaluation	8	3
2013	Barack Obama	Program Evaluation and Data Analytics	8	4
2014	Barack Obama	Program Evaluation and Data Analytics	8	5
2015	Barack Obama	Program Evaluation and Data Analytics	7	6
2016	Barack Obama	Building Evidence with Administrative Data	7	9
2017	Barack Obama	Building the Capacity to Produce and Use Evidence	7	10
2018	Donald Trump	Building and Using Evidence to Improve Government Effectiveness	6	4
2019	Donald Trump	Building and Using Evidence to Improve Government Effectiveness	6	6
2020	Donald Trump	Building and Using Evidence to Improve Government Effectiveness	6	8

*Note.* Fiscal year (FY) 2011 was the first year that a distinct chapter on evaluation was featured in the *Analytical Perspectives*.

**Prevalent words.** The contents of Table 4 were displayed as a line graph to illustrate the prevalence of these select key terms in the *Analytical Perspectives* over time (Figure 6). Afterward, the spike and drops in the graph were noted. This observation will be expound upon in the third article of this series because of the addition of other data that provide context to this graph. The vertical dotted lines in this figure indicate the first President’s Budget for a new Presidential Administration (i.e., blue lines for Democratic and red lines for Republican). FY 1996 was the start of the Bill Clinton Administration, FY 2002 the start of the George W. Bush Administration, FY 2010 the start of the Barack Obama Administration, and FY 2018 the start of the Donald Trump Administration. With these demarcations, we can observe trends in the prevalence of these key terms during years which those Administrations authored and authorized the *Analytical Perspectives*.

Comparing data from Table 4 and Table 6 provides some additional depth to the analyses. For example, the word “evidence-based” did not appear in the *Analytical Perspectives* until FY 2008. In FY 2005 and 2006, there was an uptick of the word “evidence”, which may have inspired a shift toward using evidence to inform decision making processes. In FY 2011, there is a large jump in prevalence for nearly all of the key terms. This marked increase, also apparent in Figure 6, reflects founding the first chapter dedicated to evaluation in the *Analytical Perspectives*. In FY 2013 through 2015, there is sudden spike in the prevalence of “analy\*”, which aligns with the change in the chapter title to, “Program Evaluation and Data Analytics” during those years. In FY 2016, the occurrences of “statistic\*” dramatically increases while all other counts decrease. This fluctuation relates to the sudden change in focus from evaluation to

administrative data. When the chapter shifted focus to producing and using evidence in FY 2017, there is another surge in the prevalence of the other terms. In FY 2018, another drop is observed but this aligns with the transition in Administration and dramatic shortening of the document from 10 pages in FY 2017 to only four pages in FY 2018.





Figure 6. Line Graph of Word Prevalence for Searched Terms

Note. This figure displays the prevalence of key terms searched in the *Analytical Perspectives* documents. Vertical dotted lines represent the first document published under a new U.S. Presidential Administration (i.e., blue for Democrat and red for Republican).

In order to observe patterns in overall word prevalence (i.e., not just of select terms), the most highly prevalent words in each evaluation chapter were investigated to compare differences in the frequency (i.e., density of terms) over time (Patton, 2015b). Comparison across years may indicate a shift in the prioritization or emphasis of terms in the documents at different points in time. Using the word cloud function in ATLAS.ti, frequencies for all words in these chapters were produced (Figure 7).

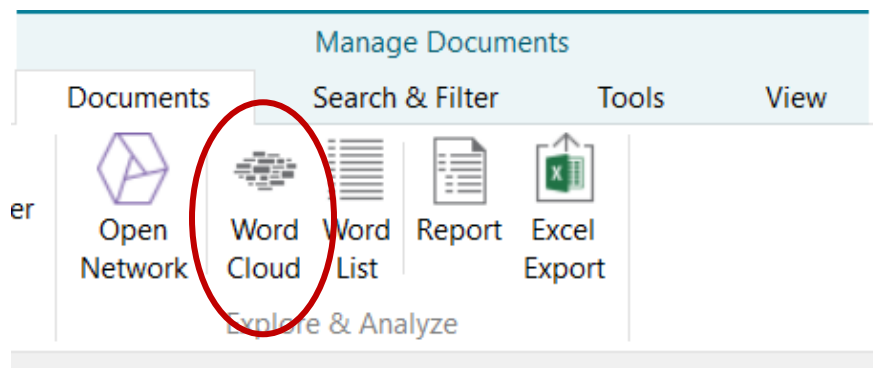


Figure 7. Snapshot of Word Cloud Function in ATLAS.ti

*Note.* The red circle shows the icon for the word cloud function in ATLAS.ti, which produces a word cloud based on word prevalence for the selected document from the document manager view.

A sliding bar in the upper left-hand section of the toolbar can be toggled to set the threshold for the minimum count for inclusion in the word cloud (shown in Figure 8). Words were included if they had 10 or more occurrences in the text. Conjunctions and other insignificant words (i.e., and, or, a, etc.) were manually excluded.

These counts were collected for each chapter and summed for the overall total across all chapters. The full list of the most highly prevalent words for all 10 evaluation chapters are displayed in Appendix A, but a truncated list is provided in Table 7 to highlight the counts for the 15 most prevalent words present at three pivotal timepoints—FY 2011, 2016, and 2017. Upon constructing the frequency tables, these data were reviewed and compared with document metadata (Table 6) to elucidate the reasoning for these patterns. The FY 2011 evaluation chapter was selected because it was the very first evaluation chapter featured in the *Analytical Perspectives*. Although the OMB published other documents prior to this chapter supporting program evaluation (i.e., the 2009 Memo M-10-01 “Increased Emphasis on Program Evaluation”), FY 2011 was the first President’s Budget to dedicate a chapter to reflect the prioritization of evaluation. The FY 2016 chapter was included because of the shift in focus toward administrative data in this year. Finally, the FY 2017 chapter was selected because it had the highest prevalence of “evaluation” across all evaluation chapters (as shown in Figure 6).

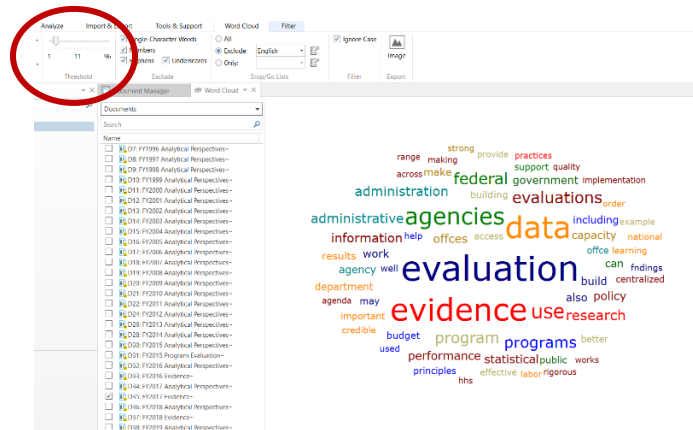


Figure 8. Snapshot of Word Cloud Produced by ATLAS.ti

Table 7.

Mostly Highly Prevalent Words in the Evaluation Chapters

Prevalence Rank	FY 2011		FY 2016		FY 2017	
	Word	Count	Word	Count	Word	Count
1	Program	38	Data	255	Evaluation	149
2	Evaluation	37	Administrative	89	Evidence	98
3	Department	27	Statistical	83	Agencies	89
4	Effectiveness	20	Program	78	Data	86
5	Agencies	17	Use/Using	74	Use/Using	80
6	Evidence	16	Agencies	63	Program	75
7	Building	11	Federal	57	Information	41
8	Capacity	11	Access	49	Research	39
9	Improve	9	Information	47	Building	37
10	Administration	9	Budget	35	Federal	36
11	Strong	9	Improve	33	Work(s)	36
12	Education	9	Evaluation	32	Policy	33
13	Federal	8	Evidence	30	Implementation	33
14	Government	8	Research	29	Administrative	32
15	Work(s)	8	Policy	29	Statistical	29

*Note.* This table displays the 15 most highly prevalent words in the evaluation chapters of the *Analytical Perspectives*. The data for all 10 evaluation chapters from FY 2011 to 2020 are displayed in Appendix A. This table features three years where there are notable shifts in word prevalence.

When the evaluation chapter was first introduced in FY 2011, words like “program” and “effect(ive)” were often paired with “evaluation”, which paints a picture of an emphasis on using evaluation to measure program effectiveness. The top five words in this list are nouns, which suggests descriptive language aimed to place and define new institutional logics (Loewenstein, 2014; Mark & Henry, 2004; Meyer et al., 2018; Nigam & Ocasio, 2010). Given the fact that this chapter was only two pages long in FY 2011, this explains why the counts were relatively low compared to other years. In FY 2016, the most highly prevalent words looked remarkably different from the preceding and succeeding years. Whereas every other year "evaluation" was the first or second most common word featured in the chapter, it was the 12th most common word in FY 2016.

That year, the shift to the generation and use of administrative data removed "Program Evaluation" from the title and “evaluation” was, therefore, scarcely mentioned (i.e., 32 times) within the chapter compared to the 84 times it was mentioned in FY 2015 and 149 time in FY 2017. As such, “data” was the clear leader in this chapter focused on administrative data with “administrative” and “statistical” as distant second- and third-most prevalent words, respectively. The words in this list explicate nouns describe in FY 2011 to provide additional details about the specific type of evidence (i.e., administrative data), analysis (i.e., statistical), evaluation (i.e., program), and effectiveness (i.e., utilization). These words are not only elaborations of descriptors, they also hint at action or behavior (Haveman, 1993; Li, 2017; Mark & Henry, 2004; Meyer et al., 2018).

In FY 2017, despite dropping the word evaluation from the title, “evaluation” becomes the most prevalent word in this chapter on “Building the Capacity to Produce

and Use of Evidence,” which legitimizes evaluation as an important evidence-generating activity. In fact, FY 2017 is only one of three evaluation chapters included in this study to have “evaluation” as its most prevalent word. In FY 2012 and FY 2020, “evaluation” was even more prevalent than “program” compared to preceding years, but neither surpassed FY 2017. These numbers also reflect the large spikes shown in Figure 6 compared to adjacent years. However, FY 2017 was also the longest evaluation chapter at 10 pages whereas FY 2012 was four pages long and FY 2020 was eight. In FY 2020, although the title of the chapter still did not explicitly include evaluation, this was the first chapter to include information about the Evidence Act, which resulted in a marked increase in discussing evaluation. The rise of other words in the FY 2017 chapter such as “research” and “policy” served to establish evaluation as a rigorous field of inquiry that informs policy decisions (Loewenstein et al., 2012).

Word clouds, weighted by word frequencies, were generated to display the most highly prevalent words in each chapter (using data from Appendix A). In Figure 9, a word cloud displays the aggregate of the 25 most highly prevalent words in the evaluation chapters across all 10 years (FY 2011-2020). Additionally, Table 8 displays the counts of those words. This list was increased from 15 to 25 to capture the fluctuations in the top 15 words across other years. In some years, the most highly prevalent words were not among the highest 15 in other years. By increasing the number to 25 words, the most prevalent in any of the evaluation chapters could be included.



Figure 9. Word Cloud of Aggregated Most Prevalent Words

*Note.* This word cloud displays the aggregated counts of the 25 most highly prevalent words across all 10 years of the evaluation chapters in the *Analytical Perspectives*. This graphic was generated at <https://www.wordclouds.com> using the counts displayed in Table 8. The most highly prevalent word is “evaluation” (n = 736) and the least prevalent word is “support” (n = 120).

Table 8.

Aggregated Counts of Most Highly Prevalent Words

Prevalence Rank	Word	Count
1	Evaluation	736
2	Program	731
3	Data	679
4	Evidence	564
5	Agencies	503
6	Use/Using	440
7	Federal	252
8	Policy	234
9	Building	215
10	Improve	213
11	Effective(ness)	203
12	Analysis	192
13	Government	189
14	Administrative	189
15	Research	183
16	Information	182
17	Performance	178
18	Work(s)	159
19	Administration	148
20	Department	146
21	Access	127
22	Budget	124
23	Learning	123
24	Example	123
25	Support	120

**Evidence of institutionalization.** The aforementioned framework presented by Meyer et al. (2018) was used to organize open codes to investigate patterns of institutionalization (example coding shown in Table 5). After these codes were sorted, the code-document function was used in ATLAS.ti to cross-tabulate the occurrences of the eight levels of institutionalization (i.e., placement, exposure, appeal/mobilization,



typification, explanation, justification, taken-for-grantedness, and diffusion/translation) across documents (i.e., the evaluation chapters of the *Analytical Perspectives* documents).

Figure 10 shows the icon in ATLAS.ti that is used to create this table.

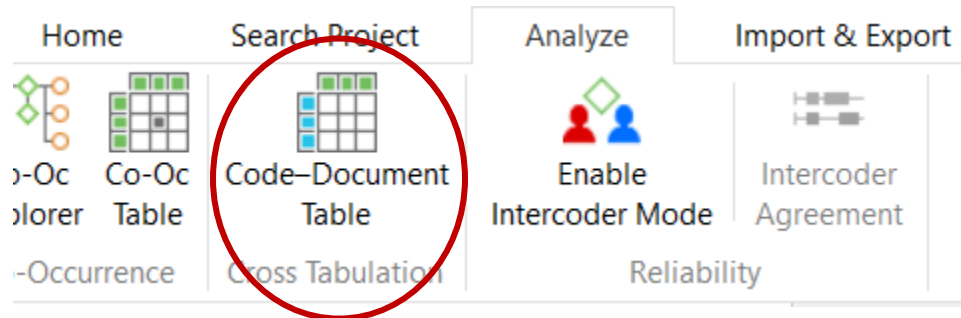


Figure 10. Snapshot of Code-Document Table Function in ATLAS.ti

*Note.* The red circle shows the icon for the code-document cross-tabulation function in ATLAS.ti, which produces a table with counts (i.e., density) of codes across documents. This icon can be found under the “Analyze” tab from the document manager view.

The resultant table from this cross-tabulation is shown in Table 9. This table illustrates the change in the tone of language over time across the levels of institutionalization (i.e., from pre-institutionalization to semi-institutionalization to full institutionalization) as seen in the text from chapters dedicated to evaluation and evidence in the *Analytical Perspectives*. Closer investigation of these counts suggests there is an observable progression from pre-institutionalization toward full institutionalization.

Table 9.

Mapping Data to Levels of Institutionalization

<b>Category</b>	<b>FY 2011</b>	<b>FY 2012</b>	<b>FY 2013</b>	<b>FY 2014</b>	<b>FY 2015</b>	<b>FY 2016</b>	<b>FY 2017</b>	<b>FY 2018</b>	<b>FY 2019</b>	<b>FY 2020</b>	<b>Total</b>
<i>Pre-Institutionalization</i>											
Placement	3	6	5	7	9	2	6	5	4	3	50
Exposure	3	6	6	9	10	5	6	1	2	2	50
Appeal/Mobilization	1	6	6	8	7	7	8	0	1	5	49
<i>Semi-Institutionalization</i>											
Typification	4	8	12	16	19	5	10	4	5	5	88
Explanation	3	10	11	12	11	5	10	8	8	8	86
Justification	4	4	12	12	20	16	23	8	9	11	119
<i>Full Institutionalization</i>											
Taken-for-Grantedness	3	3	7	8	7	7	16	9	12	14	86
Diffusion/Translation	0	0	1	1	2	4	6	2	7	9	32

*Note.* This table displays the counts of the occurrences of the institutionalization codes adopted from the Meyer et al. (2018) framework.

Language in the first five years (i.e., FY 2011-2015) initially focused on placing evaluation in relation to other similar fields, exposing it as an accessible and desirable practice, and portraying evaluation to appear attractive and logical (examples shown in Table 5). Although evaluation already existed in the government prior to the first evaluation chapter, the *Analytical Perspectives* were used to construct a shared understanding of evaluation (Colyvas & Jonsson, 2011). For example, early evaluation chapters focused on instituting the importance of evidence and the inadequacies of existing methods and demonstrated how evaluations are perfectly positioned to address those concerns. Evaluation was compared and contrasted with familiar, established, and reputable fields of inquiry within the Federal Government (Freidson, 1986; Haveman, 1993; Loewenstein et al., 2012). In FY 2011, this excerpt came after stating the importance of improving performance,

Performance measurement is a critical tool managers use to improve performance, but often cannot conclusively answer questions about how outcomes would differ in the absence of a program or if a program had been administered in a different way. That is where program evaluations pay a critical role. (p. 91)

However, early language was vague and mostly spoke about the practice of evaluation. This language evolved to generalize the concept of evaluation to other possible applications by explaining its role in contributing to a culture of continuous feedback and improvement to address public policy issues. FY 2014 and 2015 featured a list of commitments regarding a culture of learning that the government should aspire to

build to be more efficient and effective in its service to the American people. These seven vision statements affirm (p. 91 in FY 2014 and p. 65 in FY 2015):

“It is a culture that keeps asking, “How can we do things better?” and approaches public policy and management challenges with humility about what we know or don’t know about what works.

It is a culture that values rapid, operationally-focused experiments that can quickly boost program efficient, effectiveness and customer service, while at the same time equally valuing longer-term evaluation focused on more fundamental questions about program strategy.

It is a culture that sees program evaluation and performance measurement as valuable, complementary tools, since each has different strengths.

It is a culture that believes in using data to drive decision-making and is not satisfied with anecdotal evidence, since intuition about what works is often wrong.

It is a culture where people are open to changing their minds and practices based upon evidence.

It is a culture that is committed to publicly disseminating results from evaluations in an open and transparent manner, never suppressing evidence because it is politically inconvenient.

It is a culture that sees improved program performance not as a destination that can be reached with the right tool or strategy, but as a process of ongoing program refinement, since new challenges will always arise and new knowledge and innovations can always bring better outcomes and efficiencies.” (p. 91)

The language of these documents continued to increase in detail and generalizability to various needs across the Federal government. These chapters in FY 2014 and FY 2015 also featured the creation of initiatives to highlight model evaluations; funds to invest in more evaluations; repositories to build a knowledge base of standards in evaluation; and communities of practice (i.e., via workshops and working groups) to support sharing evaluation best practices.

The chapters used the foundation of a shared understanding and increased access to quality evaluations to progress toward the legitimization of evaluation as a viable function, which later chapters continued to substantiate as having an essential role for government effectiveness. In fact, in the FY 2015 evaluation chapter, it was stated that evaluation should not be an option,

“Research, statistics, and evaluation are part of any comprehensive effort to use data and evidence to serve the American people in more cost-effective ways. Funding for these areas should never be viewed as a luxury, but rather as an essential element of running effective government programs.” (p. 68)

In the most recent chapters related to evidence and evaluation (i.e., FY 2018, 2019, and 2020), they describe an undeniable need to hire and equip evaluation practitioners, officers, and advocates for the diffusion of evaluation across every policy arena. These documents effectively translated evaluation to every context in government by explicitly highlighting examples of evaluations in different agencies. By providing a diverse coverage of examples of evaluation use across different sectors (e.g., homeland security, justice, education, health and human services, housing and urban development, labor), the evaluation chapters justify and diffuse value in evaluation across the entire Federal government. As such, although it cannot be definitely stated that evaluation has been fully institutionalized in the U.S. Federal Government, these data suggest that the language in the *Analytical Perspectives* chapters approach diffusion and translation of evaluation widely across the government. The following section will further discuss these findings and the implications for evaluation practice.

## **Discussion and Implications**

In summary, the nature and content of the *Analytical Perspectives* greatly changed over the last 25 years. The inaugural chapter dedicated to evaluation in FY 2011 was a pivotal change demonstrated by a marked increase in the prevalence of the word “evaluation” (Figure 6). In addition to a greater prevalence in mentioning evaluation, the chapters grew in content length and the chapters gradually elevated in chapter order over time. Some shifts in organization and content could be attributed to the transition of Presidential Administration; implementation of new initiatives and policies; or shifting focus between evaluation, performance, evidence, and administrative data. The language in the evaluation chapters of the *Analytical Perspectives* suggest that the Federal government has moved through stages of diffusing value in evaluation toward fully institutionalizing evaluation. This section will revisit the conceptual model for diffusing value in evaluation presented at the beginning of this paper based on empirical data and discuss key areas that have promoted institutional change.

**Realizing the process.** In using the Meyer et al. (2018) framework as a guide to map the process of institutionalization in the text, a clearer delineation of the organizational change process in the U.S. Federal Government became exceedingly apparent through analysis of the *Analytical Perspectives* documents. What emerged from this analysis was an understanding that messaging in the documents were instrumental to building signs (i.e., a word or phrase that signifies a constructed meaning) to influence the progression toward institutionalization (Gee, 2014; Li, 2017). Whereas before the types of influences adapted from Mark and Henry (2004) were merely presented as

nominal categories (Figure 3), categories were modified to accommodate the data and more explicitly demonstrated that cognitive and motivational influences implied a readiness for institutional change (i.e., pre-institutionalization), behavioral influences indicated adoption (i.e., semi-institutionalization), and structural influences were evidence of full institutionalization (Armenakis & Harris, 2002; Meyer et al., 2018).

Therefore, the data inspired us to modify the Meyer et al. (2018) framework and assign the existing categories (i.e., placement, exposure, etc.) to align with types influences (Figure 11). Language displayed clear efforts to place, expose, and mobilize evaluation to influence readers' thoughts and feelings; typify and explain evaluation to motivate buy-in; legitimize evaluation as an essential function to influence behavior; and justify and diffuse value in evaluation to influence structural changes for implementation. The way in which the *Analytical Perspectives* built tasks around each of these types of influences was by conveying change messages (cognitive); emphasizing executive buy-in (motivational); creating principles and standards (behavioral); and establishing centralized evaluation units (structural).

## INFLUENCES

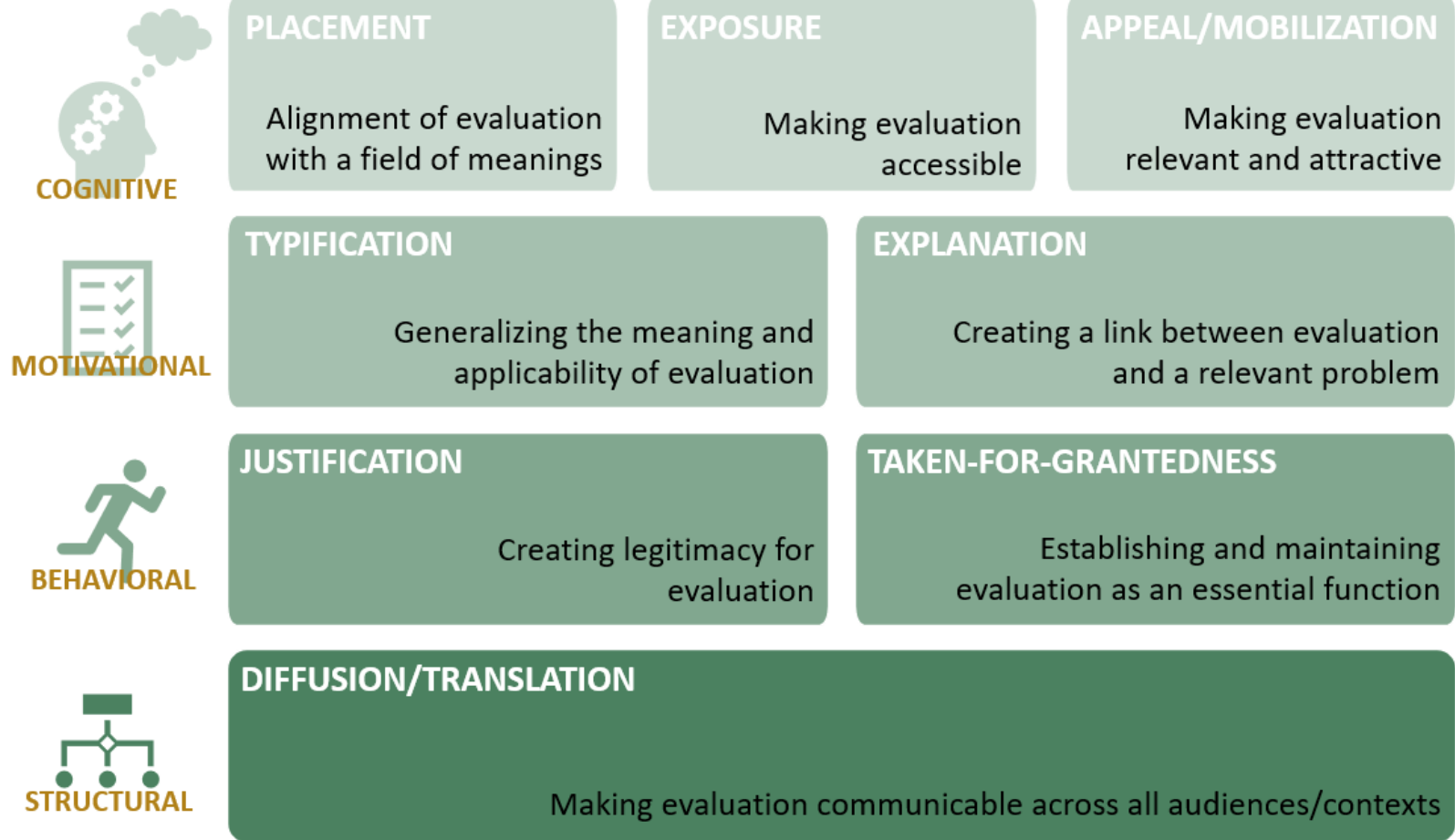


Figure 11. Modified Institutionalization Framework Aligned to Types of Influences



**Conveying persuasive change messages.** The first three years of evaluation chapters in the *Analytical Perspectives* (i.e., FY 2011, 2012, and 2013) were dedicated mostly to conveying consistent and convincing change messages to persuade an uptake of evaluation in the Federal government (Armenakis & Harris, 2002; Bolman & Deal, 2017). The intention was to socialize people to a cultural norm of evaluation by constructing a new value system and relating it to familiar concepts, for example, performance measurements and statistical analyses (Haveman, 1993; van Leeuwen, 2007; Warrick, 2017). This statement from the FY 2011 chapter perfectly touches on the placement, exposure, and appeal of evaluation,

“A central pillar of good government is a culture where answering such questions is a fundamental part of program design and where agencies have the capacity to use evidence to invest more in what works and less in what does not.” (p. 91)

The emphasis on the centrality of evaluation (i.e., the placement phase of institutionalization), giving power and access to the Federal agencies to perform evaluations (i.e., exposure), and the focus on making smart investments (i.e., appeal/mobilization) are instrumental in influencing cognitive affect (Armenakis & Harris, 2002; Colyvas & Jonsson, 2011; Mark & Henry, 2004). This chapter also makes evaluation more approachable to agencies by calling attention to a previous memo from former OMB Director Peter Orszag titled, “Increased Emphasis on Program Evaluation” (M-10-01); giving agencies access to other ongoing evaluations; forming an inter-agency working group to allow constructive exchange; and launching a \$100 million initiative to fund agencies to do new evaluations. Therefore, early language was instrumental in transferring an established value *of* evaluation to motivate a value *in* evaluation.

**Emphasizing executive buy-in.** In FY 2016, the “Building Evidence with Administrative Data” chapter on evaluation opened with this quote from the “Remarks on Promise Zones” speech that President Obama gave in 2014,

“We’ve got Democratic and Republican elected officials across the country who are ready to roll up their sleeves and get to work. And this should be a challenge that unites us all. I don’t care whether the ideas are Democrat or Republican. I do care that they work. I do care that they are subject to evaluation...” (p. 65)

Associating the President’s voice here at the start of this chapter is a strategy to build consensus through executive buy-in. This quote acknowledges the division between Democratic and Republican officials but urges bipartisan unity under a shared goal to do and use more evaluations (i.e., explanation).

The prioritization of evidence was codified in 2016 when President Obama signed the Commission for Evidence-Based Policymaking Act (P.L 114-140) into law appointing a commission to study the Federal government’s capacity for evidence-building and policymaking. The FY 2016 evaluation chapter highlighted this Act as a major step in “Setting the Stage for Future Progress.” The joint bipartisan sponsorship between then House Speaker Paul Ryan (Republican) and Senator Patty Murray (Democrat) emboldened the political will of other legislators to make this issue a top priority. The Commission investigated current uses of data in the Federal government to inform program and policy decision-making (i.e., typification) then submitted recommendations to Congress (Groves & Schoeffel, 2018). They reported redundant, inefficient, and harmful practices that hamstring knowledge brokering (Commission on Evidence-Based Policymaking, 2017; Godfrey et al., 2010; Liebman, 2018; Olejniczak et al., 2016). Their findings and recommendations offered evaluation as a plausible solution

for the issues they discovered (i.e., explanation), which were generalized across the government and not focused on any particular agency or policy area (i.e., typification).

**Creating principles and standards.** In FY 2014, the evaluation chapter in the *Analytical Perspectives* stated that the Federal government would begin sharing best practices for evaluation. In FY 2015, an entire section of the evaluation chapter discussed the importance of building capacity for evaluation and mechanisms that agencies could use to begin sharing best practices through working groups and forums. Unlike previous years where best practices were instructive for new adopters, the purpose of sharing best practices in FY 2015 and later centered on creating an appetite and knowledge base for standards of practice (i.e., justification). Citing examples published by the Administration for Children and Families (ACF) and the Department of Labor (DoL), the FY 2017 evaluation chapter legitimized evaluation (i.e., typification) as an essential routine (i.e., taken-for-grantedness) by stating the need for evaluation policies that establish core principles.

The evaluation chapter on became a pivotal document for federal evaluation—and is still featured by the OMB as a foundational text for current evaluation practice<sup>3</sup>—because it successfully introduced principles of evaluation. It stated that,

“Establishing these standards is an important building block in furthering agencies’ capacity to routinely build and use high-quality evidence to improve program performance, and help evaluation offices maintain standards for their programs across administrations and changes in personnel.” (p. 71)

A preliminary framework of the fundamental principles was proposed in the FY 2017

---

<sup>3</sup> Retrieved at [https://obamawhitehouse.archives.gov/sites/default/files/omb/budget/fy2017/assets/ap\\_7\\_evidence.pdf](https://obamawhitehouse.archives.gov/sites/default/files/omb/budget/fy2017/assets/ap_7_evidence.pdf)

*Analytical Perspectives* and included rigor, relevance, independence, transparency, and ethics. This four-year evolution demonstrates a progression from statements of intention, implementation, to diffusion, which shows a commitment to institutionalizing evaluation. In March 2020, OMB codified the “Program Evaluation Standards and Practices” in the M-20-12 memo, which includes relevance and utility, rigor, independence and objectivity, transparency, and ethics.

While the FY 2017 Budget was debated in the 114<sup>th</sup> Congress, the National Academies of Sciences, Engineering, and Medicine (NAS) convened a public workshop in October 2016, whereby thought leaders voiced their support of these five principles (National Academies of Sciences, 2017). In addition to highlighting the evaluation policies from ACF and DoL, this group also sought mimetic isomorphism by drawing inspiration from the Federal Statistical System, who published their own *Principles and Practices for a Federal Statistical Agency* in 1992 (DiMaggio & Powell, 1983; Haveman, 1993; National Research Council, 1992). They concluded by discussing the tricky but necessary process of institutionalizing the principles into practice (Colyvas & Jonsson, 2011). They acknowledged the change agents who could either drive or impede this process to bring about awareness of the potential opportunities and obstacles that may be encountered (Espedal, 2017; Oliver, 1992). Lastly, they noted that structural changes would be required to coordinate agency-wide operationalization and implementation of the proposed policies and principles of evaluation practice (Higgins et al., 2012).

**Establishing centralized evaluation units.** The FY 2017 chapter on “Building the Capacity to Produce and Use Evidence” was appropriately named because this was

the first year that centralized evaluation units were mentioned in the *Analytical Perspectives* to build greater capacity to *do* and *use* evaluation. Furthermore, the FY 2020 evaluation chapter was the first chapter published after the Evidence Act was enacted, which echoed the mandate to appoint CEOs. These chapters highlighted the importance of centralized evaluation offices as a key strategy to strengthen and broker evidence (i.e., translation) to policymakers, which make them an invaluable part of the organizational structure (Abbate & Coppolino, 2011; Campbell et al., 2011; Olejniczak et al., 2016).

These centralized units also coordinate sharing methods and data (i.e., diffusion/translation) with other programs or agencies (Cabatoff, 2000; Strang & Meyer, 1993). Higgins et al. (2012) also discussed the value of leveraging internal teams to facilitate institutional change. As the FY 2017 chapter states,

“Centralized or chief evaluation offices are often a key component of implementing evaluation policies reflective of the core principles discussed above.” (p. 76)

However, the FY 2020 chapter acknowledged that only some agencies already had established offices whereas others are still in the process of building these teams. Every agency has a unique baseline culture and capacity for evaluation and the strategies and resources for implementation must reflect the needs of that agency. Institutionalizing structural changes in an organization require general capacity support, but centralized evaluation units are uniquely equipped to provide specialized capacity for agency-specific needs (Hall & Hord, 2015; Nolton et al., 2020b; Ting, 2009).

**Limitations and future research.** This study is not without its limitations. The *Analytical Perspectives* were downloaded from the internet as PDF documents, which

made some of the text unclear. Although it is still possible to code as images with QDA software, this may have affected the auto-coding and word counting processes. The paragraphs in these documents were also formatted as fully justified with hyphenated words that carried to the next line. This might have presented an issue if one of the auto-coded word searches included words that were hyphenated and split across two lines. In these cases, these words would not have been captured in the count because the auto-coding search may not have recognized the word fragments as a match to the word. However, these occurrences are rare and should not greatly affect count accuracy.

Using the *Analytical Perspectives* (intentionally) limits observations about organizational change to how messages were conveyed through these documents; however, there are other modes of influence—both internal and external to the government—that also construct tasks, signs, and values systems (Gee, 2014; van Leeuwen, 2007). The *Analytical Perspectives* were chosen because they uniquely speak to cross-government decisions, but still acknowledge individuality of the agencies (i.e., microlevels). These documents are also tied to the annual Budget, which inherently make the language focused on explaining and justifying allocation and investment decisions. The *Analytical Perspectives* are also heavily influenced by political agenda, which influences the language that is used. Chelimsky (1987) stated that it is, in fact, impossible to separate government evaluation from politics. Although it would not be the goal to remove the influence of politics from these documents, we acknowledge that it is a sociocultural factor that frames all government and policy processes.

Due to the way in which the *Analytical Perspectives* are authored and edited between fiscal years, there are instances where the language is stagnate or verbatim from preceding years. As such, slowed progression toward diffusion may be a result of consistency in reinforcing the political agenda rather than evidence of obstacles stalling progress toward institutionalizing evaluation. In contrast, large shifts may be a result of a new Administration attempting to establish their own identity and principles. The desire for distinctiveness between Administrations may give the illusion of either progression or regression but may not reflect reality. Lastly, language does not always dictate behavior or practice, but it does suggest a desire for reality. As such, the documents may not accurately represent how institutionalizing evaluation has truly unfolded. This continues to rationalize future research to investigate other sources of influence. The third manuscript in this series will include interviews with participants to observe their perceptions, reactions, and responses to institutional change.

## **Conclusion**

Content and discourse analyses of the *Analytical Perspectives* suggest that evaluation is moving toward institutionalization in the U.S. Federal Government. There are clear trends in the increase in prevalence of key terms related to evaluation through these documents over time and discursive features demonstrate an elevation of evaluation as an essential function of government. Although diffusion is not evenly distributed throughout all areas of the government, this common, cross-cutting document attempts to align institutional logics with attractive, accessible, generalizable, and translatable language (Meyer et al., 2018; Nigam & Ocasio, 2010). Since FY 2017, the language has

shifted to influence behavioral and structural change by creating standards and policies for evaluation practice and establishing centralized evaluation units to facilitate implementation (Higgins et al., 2012; Olejniczak et al., 2016). Investing in capacity building for greater production and use of evaluation exhibits a commitment to the diffusing value in evaluation across the Federal government.

As each chapter on evaluation and evidence matured between FY 2011 to 2020, the text spent less time placing and mobilizing evaluation as a novel idea. The language evolved to emphasize the ubiquity of evaluation and the trajectory over the last 10 years suggests that it will continue to move toward more diverse applications across the government (Green, 2004). Although political climate and agendas may cause progress to be punctuated, the evolution of evaluation has been steadily advancing toward institutionalization. This study provides a method to capture the use of government budget narratives to influence the culture and capacity of evaluation. These findings can help government officials understand the origins and development of fundamental beliefs that inspire or constrain the process of institutionalizing evaluation in the Federal government and how to foster that progress (Caracelli, 2006). Understanding the perceptions of stakeholders who experienced the ebb and flow of this process will build on trends observed in the *Analytical Perspectives* to identify other levers that were effective or ineffective in driving institutional change but not reported in these documents. The next manuscript in the series will combine findings from this present study with lived experiences to map factors that have influenced the institutionalization of evaluation in the U.S. Federal Government.



## References

- Abbate, T., & Coppolino, R. (2011). Knowledge creation through knowledge brokers: Some anecdotal evidence. *Journal of Management Control*, 22, 359-371.  
<https://doi.org/10.1007/s00187-011-0144-5>
- Alkin, M. C., & King, J. A. (2016). The historical development of evaluation use. *American Journal of Evaluation*, 37(4), 568-579.  
<https://doi.org/10.1177/1098214016665164>
- Armenakis, A. A., & Harris, S. G. (2002). Crafting a change message to create transformational readiness. *Journal of Organizational Change Management*, 15(2), 169-183. <https://doi.org/10.1108/09534810210423080>
- Bansback, N., Li, L. C., Lynd, L., & Bryan, S. (2014). Exploiting order effects to improve the quality of decisions. *Patient Education and Counseling*, 96(2), 197-203. <https://doi.org/10.1016/j.pec.2014.05.021>
- Barber, J. P., & Walczak, K. K. (2009). *Conscience and critic: Peer debriefing strategies in grounded theory research*. American Educational Research Association, San Diego, California.
- Berger, P., & Luckman, T. (1967). *The social construction of reality—A treatise in the sociology of knowledge*. Anchor Doubleday.

- Bolman, L. G., & Deal, T. E. (2017). *Reframing organizations: Artistry, choice, and leadership* (6th ed.). John Wiley & Sons, Inc.
- Brown, A. D., Ainsworth, S., & Grant, D. (2012). The rhetoric of institutional change. *Organization Studies*, 33(3), 297-321. <https://doi.org/10.1177/0170840611435598>
- Cabatoff, K. (2000). Translating evaluation findings into "policy language". In R. K. Hopson (Ed.), *How and why language matters in evaluation. New Directions for Evaluation*, 86, 43-54. <https://doi.org/10.1002/ev.1171>
- Campbell, D., Donald, B., Moore, G., & Frew, D. (2011). Evidence check: Knowledge brokering to commission research reviews for policy. *Evidence & Policy*, 7, 97-107. <https://doi.org/10.1332/174426411X553034>
- Campbell, J. L., Quincy, C., Osserman, J., & Pedersen, O. K. (2013). Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement. *Sociological Methods & Research*, 42(3), 294-320. <https://doi.org/10.1177/0049124113500475>
- Caracelli, V. J. (2006). Enhancing the policy process through the use of ethnography and other study frameworks: A mixed-method strategy. *Research in the Schools*, 3(1), 84-92.
- Charmaz, K., Thornberg, R., & Keane, E. (2018). Evolving grounded theory and social justice inquiry. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (5th ed., pp. 411-443). SAGE Publications, Inc.

- Chelimsky, E. (1987). The politics of program evaluation. *New Directions for Program Evaluation*, 1987(34), 5-21. <https://doi.org/10.1002/ev.1450>
- Colyvas, J. A., & Jonsson, S. (2011). Ubiquity and legitimacy: Disentangling diffusion and institutionalization. *Sociological Theory*, 29(1), 27-53.
- Commission on Evidence-Based Policymaking. (2017). *The promise of evidence-based policymaking: Report of the commission on evidence-based policymaking*. Washington, DC Retrieved from <https://www.cep.gov/report/cep-final-report.pdf>.
- D'Eredita, M. A., & Barreto, C. (2006). How does tacit knowledge proliferate? An evidence-based perspective. *Organization Studies*, 27, 1821-1841. <https://doi.org/10.1177/0170840606067666>
- DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48, 147-160.
- Espedal, B. (2017). Understanding how balancing autonomy and power might occur in leading organizational change. *European Management Journal*, 35, 155-163. <https://doi.org/10.1016/j.emj.2016.08.005>
- Flick, U. (2018). Triangulation. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (5th ed., pp. 444-461). SAGE Publications, Inc.

- Freidson, E. (1986). *Professional powers: A study of the institutionalization of formal knowledge*. University of Chicago Press.
- Friedland, R., & Alford, R. (1991). Bringing society back in: Symbols, practices, and institutional contradictions. In W. W. Powell & P. J. DiMaggio (Eds.), *The new institutionalism in organizational analysis* (pp. 232-263). University of Chicago Press.
- Gee, J. P. (2011). Discourse analysis: What makes it critical? In R. Rodgers (Ed.), *An introduction to critical discourse analysis in education* (2nd ed., pp. 23-45). Routledge.
- Gee, J. P. (2014). *An introduction to discourse analysis: Theory and method*. Routledge.
- Godfrey, L., Funke, N., & Mbizvo, C. (2010). Bridging the science-policy interface: A new era for South African research and the role of knowledge brokering. *South African Journal of Science*, 106(5/6), 1-8.  
<https://doi.org/10.4102/sajs.v106i5/6.247>
- Green, S. E. (2004). A rhetorical theory of diffusion. *Academy of Management Review*, 29(4), 653-669.
- Groves, R. M., & Schoeffel, G. J. (2018). Use of administrative records in evidence-based policymaking. *Annals of the American Academy of Political and Social Sciences*, 378, 71-80. <https://doi.org/10.1177/0002716218766508>

- Hall, G. E., & Hord, S. M. (2015). *Implementing change: Patterns, principles, and potholes* (4th ed.). Pearson Education, Inc.
- Haskins, R. (2018). Evidence-based policy: The movement, the goals, the issues, the promise. *Annals of the American Academy of Political and Social Sciences*, 678, 8-37. <https://doi.org/10.1177/0002716218770642>
- Haveman, H. A. (1993). Follow the leader: Mimetic isomorphism and entry into new markets. *Administrative Science Quarterly*, 38(4), 593-627.
- Higgins, M. C., Weiner, J., & Young, L. (2012). Implementation teams: A new lever for organizational change. *Journal of Organizational Behavior*, 33, 366-388. <https://doi.org/10.1002/job.1773>
- Kinarsky, A. R. (2018). The evaluation landscape: U.S. foundation spending on evaluation. In S. B. Nielsen, S. Lemire, & C. A. Christie (Eds.), *The evaluation marketplace: Exploring the evaluation industry. New Directions for Evaluation*, 160, 81-96. <https://doi.org/10.1002/ev.20342>
- Krippendorff, K. (2019). *Content analysis: An introduction to its methodology* (4th ed.). SAGE Publications, Inc.
- Lana, R. E. (1963). Controversy of the topic and the order of presentation in persuasive communications. *Psychological Reports*, 12(1), 163-170. <https://doi.org/10.2466/pr0.1963.12.1.163>

- Lawson, R. G. (1968). Order of presentation as a factor in jury persuasion. *Kentucky Law Journal*, 56(3), 523-555.
- Lemire, S., Fierro, L. A., Kinarsky, A. R., Fujita-Conrads, E., & Christie, C. A. (2018). The U.S. federal evaluation market. In S. B. Nielsen, S. Lemire, & C. A. Christie (Eds.), *The evaluation marketplace: Exploring the evaluation industry. New Directions for Evaluation*, 160, 63-80. <https://doi.org/10.1002/ev.20343>
- Lemire, S., Nielsen, S. B., & Christie, C. A. (2018). Toward understanding the evaluation market and its industry—Advancing a research agenda. In S. B. Nielsen, S. Lemire, & C. A. Christie (Eds.), *The evaluation marketplace: Exploring the evaluation industry. New Directions for Evaluation*, 160, 145-163. <https://doi.org/10.1002/ev.20339>
- Li, Y. (2017). A semiotic theory of institutionalization. *Academy of Management Review*, 42(3), 520-547. <https://doi.org/10.5465/amr.2014.0274>
- Liebman, J. B. (2018). Using data to more rapidly address difficult U.S. social problems. *Annals of the American Academy of Political and Social Sciences*, 675, 166-181. <https://doi.org/10.1177/0002716217745812>
- Loewenstein, J. (2014). Take my word for it: How professional vocabularies foster organizing. *Journal of Professions and Organizations*, 1, 65-83. <https://doi.org/10.1093/jpo.jot004>

- Loewenstein, J. E., Ocasio, W., & Jones, C. (2012). Vocabularies and vocabulary structure: A new approach linking categories, practices, and institutions. *The Academy of Management Annals*, 6(1), 41-86.  
<https://doi.org/10.1080/19416520.2012.660763>
- Lohse, G. L. (1997). Consumer eye movement patterns on yellow pages advertising. *Journal of Advertising*, 26(1), 61-73.  
<https://doi.org/10.1080/00913367.1997.10673518>
- Lund, F. H. (1925). The psychology of belief: A study of its emotional and volitional determinants. *The Journal of Abnormal and Social Psychology*, 20(1), 63-81; 174-195. <https://doi.org/10.1037/h0076047>
- Maier, M. A. (2017). Content analysis: Advantages and disadvantages. In M. Allen (Ed.), *The SAGE encyclopedia of communication research methods* (pp. 240-242). SAGE Publications, Inc.
- Manson, H. (2016). Systematic reviews are not enough: policymakers need a greater variety of synthesized evidence. *Journal of Clinical Epidemiology*, 73, 11-14.  
<https://doi.org/10.1016/j.jclinepi.2015.08.032>
- Mark, M. M., & Henry, G. T. (2004). The mechanisms and outcomes of evaluation influence. *Evaluation*, 10(1), 35-57. <https://doi.org/10.1177/1356389004042326>
- Maxwell, J. A. (2010). Using numbers in qualitative research. *Qualitative Inquiry*, 16(6), 475-482. <https://doi.org/10.1177/1077800410364740>

- Maynard, R., Goldstein, N., & Smith Nightingale, D. (2016). Program and policy evaluations in practice: Highlights from the federal perspective. In L. R. Peck (Ed.), *Social experiments in practice: The what, why, when, where, and how of experimental design & analysis*. *New Directions for Evaluation*, 152, 109-135. <https://doi.org/10.1002/ev.20209>
- Meyer, R. E., Jancsary, D., Höllerer, M. A., & Boxenbaum, E. (2018). The role of verbal and visual text in the process of institutionalization. *Academy of Management Review*, 43(3), 329-418. <https://doi.org/10.5465/amr.2014.0301>
- Morse, J. (2018). Reframing rigor in qualitative inquiry. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (5th ed., pp. 796-817). SAGE Publications, Inc.
- National Academies of Sciences. (2017). *Principles and practices for federal program evaluation: Proceedings of a workshop*. <https://www.nap.edu/catalog/24716/principles-and-practices-for-federal-program-evaluation-proceedings-of-a>
- National Research Council. (1992). *Principles and Practices for a Federal Statistical Agency*. National Academy Press.
- Neuendorf, K. A. (2017). *The content analysis guidebook* (2nd ed.). SAGE Publications.
- Nigam, A., & Ocasio, W. (2010). Event attention, environmental sensemaking, and change in institutional logics: An inductive analysis of the effects of public



attention to Clinton's health care reform initiative. *Organization Science*, 21(4), 823-841. <https://doi.org/10.1287/orsc.1090.0490>

Nolton, E. C., Varier, D., Powell, M. G., Caracelli, V. J., & Hopson, R. K. (2020).

*Introducing a conceptual framework and method of mapping influences on the process of institutionalizing evaluation in the U.S. Federal Government.*

Olejniczak, K., Raimondo, E., & Kupiec, T. (2016). Evaluation units as knowledge

brokers: Testing and calibrating an innovative framework. *Evaluation*, 22(2), 168-189. <https://doi.org/10.1177/1356389016638752>

Oliver, C. (1992). The antecedents of deinstitutionalization. *Organization Studies*, 13(4),

563-588.

Pasachoff, E. (2017). Two cheers for evidence: Law, research, and values in education

policymaking and beyond. *Columbia Law Reviews*, 117, 1933-1972.

Patton, M. Q. (2000). Overview: Language matters. In R. K. Hopson (Ed.), *How and why*

*language matters in evaluation*. *New Directions for Evaluation*, 86, 5-16.

<https://doi.org/10.1002/ev.1168>

Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and*

*practice*. SAGE Publications, Inc.

Pawson, R. (2006). *Evidence-based policy: A realist perspective*. SAGE Publications,

Inc.

- Peräkylä, A., & Ruusuvuori, J. (2018). Analyzing talk and text. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (5th ed., pp. 669-691). SAGE Publications, Inc.
- Phillips, N., Lawrence, T. B., & Hardy, C. (2004). Discourse and institutions. *The academy of Management Review*, 29(4), 635-652.
- Rapley, T. (2007). *Doing conversation, discourse and document analysis*. SAGE Publications, Inc.
- Rizvi, F. (2004). Theorizing the global convergence of educational restructuring. In S. Lindblad & T. Popkewitz (Eds.), *Educational restructuring: International perspectives on traveling policies*. Information Age Publishing.
- Schlauffer, C., Stucki, I., & Sager, F. (2018). The political use of evidence and its contribution to democratic discourse. *Public Administration Review*, 78, 645-649.  
<https://doi.org/10.1111/puar.12923>
- Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, 22, 63-75.
- Shore, C., & Wright, S. (2011). Introduction. Conceptualising policy: Technologies of governance and the politics of visibility. In S. Shore, S. Wright, & D. Però (Eds.), *Policy worlds: Anthropology and the analysis of contemporary power* (pp. 1-25). Bergahn Books.

- Spall, S. (1998). Peer debriefing in qualitative research: Emerging operational models. *Qualitative Inquiry*, 4(2), 280-292.
- Stack, K. (2018). The Office of Management and Budget: The quarterback of evidence-based policy in the federal government. *Annals of the American Academy of Political and Social Sciences*, 678, 112-123.  
<https://doi.org/10.1177/0002716218768440>
- Stone, V. A. (1969). A primacy effect in decision-making by jurors. *Journal of Communication*, 19(3), 239-247. <https://doi.org/10.1111/j.1460-2466-1969.tb00846.x>
- Strang, D., & Meyer, J. W. (1993). Institutional conditions for diffusion. *Theory and Society*, 22, 487-511.
- Suddaby, R., & Greenwood, R. (1995). Rhetorical strategies of legitimacy. *Administrative Science Quarterly*, 50, 35-67.
- Ting, M. M. (2009). Organizational capacity. *The Journal of Law, Economics, & Organization*, 27, 245-270. <https://doi.org/10.1093/jleo/ewp021>
- U.S. General Accounting Office. (1990). *Prospective evaluation methods*. (GAO/PEMD-10.1.10). Washington, DC.
- van Leeuwen, T. (2007). Legitimation in discourse and communication. *Discourse & Communication*, 1(1), 91-112. <https://doi.org/10.1177/1750481307071986>

- Warrick, D. D. (2017). What leaders need to know about organizational culture. *Business Horizons*, 60, 395-404. <https://doi.org/10.1016/j.bushor.2017.01.011>
- Wedel, J. R. (2019). *The mapping method: A guide to charting corruption and influence processes* [Working paper forthcoming]. Global Integrity.
- Weiss, C. H. (1988). Evaluation for decisions: Is anybody there? Does anybody care? *Evaluation Practice*, 9(1), 5-20.
- Zilber, T. B. (2002). Institutionalization as an interplay between actions, meanings, and actors: The case of a rape crisis center in Israel. *Academy of Management Journal*, 45(1), 234-254.
- Zucker, L. G. (1977). The role of institutionalization in cultural persistence. *American Sociological Association*, 42(5), 726-743.

## Mapping Influences on Institutionalizing Evaluation in the U.S. Federal Government

Federal evaluation plays a large role in demonstrating the effectiveness of social programs and policies (Kinarsky, 2018; Maynard et al., 2016). Evaluations have the unique ability to provide diverse evidence sources necessary to support complex policymaking decisions, but the ways in which evaluations have been used have vastly evolved over the last 60 years (Haskins, 2018; Lemire, Nielsen, et al., 2018; Manson, 2016; Maynard et al., 2016; Pawson, 2006; Pawson et al., 2011). Although evaluative activities have existed to some degree in the U.S. Federal Government since the 1960s, there have been a number of catalysts and obstacles in the process of diffusing evaluation as a legitimate and essential function (Lemire, Fierro, et al., 2018; Nolton et al., 2020a). The growth in evaluation use in the Federal Government has reflected increasing value in evaluation over time (Nolton et al., 2020a, 2020b); however, the process and mechanisms by which this growth has occurred is not fully understood.

This is the third article in a series investigating the influences on institutionalizing evaluation in the U.S. Federal Government. The first article introduced a framework to conceptualize the process of institutionalizing evaluation in the Federal government and offered an ethnographic mapping method as a means to study the process. The second article in this series discovered—through content and discourse analyses of the *Analytical Perspectives*<sup>4</sup> of the annual U.S. President’s Budget—there was evidence that evaluation

---

<sup>4</sup> *Analytical Perspectives* are appended narratives to the proposed annual U.S. President’s Budget that explain the analyses, rationale, and intentions of financial allocations proposed by the U.S. Office of Management and Budget (OMB).

has been approaching institutionalization in the Federal government (Nolton et al., 2020a). The purpose of this manuscript is to build on previous theories and findings to map the landscape of influences (i.e., actors, events, policies, and cultures) that led to progress toward institutionalizing evaluation in the U.S. Federal Government over time.

**Current understanding of the process.** Although other scholars have written about the history of evaluation methods (Fitzpatrick et al., 2011), the history of evaluation use (Caracelli, 2000), and the factors that impede evaluation use (Chelimsky, 2015), these pieces were limited by scope, relevance, or time. Since those contributions, there have been many other factors that have led to institutional changes in the Federal government. The second paper in this series offered grounding in our understanding of how positions and perceptions may be shaped by central government documents, such as the *Analytical Perspectives* (Nolton et al., 2020a). The U.S. Office of Management and Budget (OMB) authored these documents to defend and rationalize budget decisions for the U.S. Federal Government. However, the language, contents, and arrangement of the documents are heavily influenced by the political arena. Large swings in the prevalence of key terms (i.e., evaluation, analysis, statistics, evidence, and evidence-based policymaking), number of pages or chapters, and focus of the content varied based on Presidential Administrations and current policies.

Although there have been observable patterns in the discourse related to evaluation over the last 25 years, it was not until a distinct chapter on evaluation was introduced to the *Analytical Perspectives* when institutional change became evident across the Federal government. Although it was not the first document written by the

OMB to promote evaluation, the *Analytical Perspectives* first featured a dedicated chapter on “Program Evaluation,” in 2011 and provided a shared understanding of the value *of* evaluation (i.e., statement a worth) in effort to create an appetite to build greater value *in* evaluation (i.e., belief and investment in that worth). In order to increase evaluation use, influence, and capacity at micro- and meso-levels of government, specific budget documents were intended to diffuse value in evaluation by influencing cognition, motivation, behavior, and structure for institutional change (see Figure 12). By conveying persuasive change messages; emphasizing executive buy-in; creating principles and standards of evaluation; and establishing centralized evaluation units, the language in the *Analytical Perspectives* gradually moved toward institutionalizing evaluation.

**What lies beyond, between, and beneath?** However, the language in these documents are limited to what the authors at OMB wanted to convey in those narratives (i.e., per the presiding Administration). Consequently, the *Analytical Perspectives* reflect the desired institutional change through political rhetoric, but it is not clear if their efforts to diffuse value in evaluation are effectively translated into reality and what other internal or external factors were effective or ineffective in influencing the institutionalization of evaluation in the Federal government. It is also not clear that their efforts are perceived in the ways the government hoped they would be when the documents were published. Despite the potential for a central government document to mold and align institutional logics, there are other factors at play. These additional factors can be identified through learning about the lived experiences of individuals who observed changes in evaluation from their positions in and around the Federal government.

## INFLUENCES



### PLACEMENT

Alignment of evaluation with a field of meanings

### EXPOSURE

Making evaluation accessible

### APPEAL/MOBILIZATION

Making evaluation relevant and attractive



### TYPIIFICATION

Generalizing the meaning and applicability of evaluation

### EXPLANATION

Creating a link between evaluation and a relevant problem



### JUSTIFICATION

Creating legitimacy for evaluation

### TAKEN-FOR-GRANTEDNESS

Establishing and maintaining evaluation as an essential function



### DIFFUSION/TRANSLATION

Making evaluation communicable across all audiences/contexts

Figure 12. Modified Institutionalization Framework Aligned to Types of Influences



## Methods

**Study design.** Based on social anthropological principles, mapping the process of diffusing value in evaluation across the Federal government is a useful method of inquiry to better understand the sociocultural factors that influence institutional change over time (Wedel, 2019). The mapping method used for this study accounts for the various public policy actors and processes at play. This method utilizes a comprehensive and multidirectional approach—including a variety of relevant data sources—to effectively study “through”, “out”, and “up” the meanderings of these complex processes (Burawoy, 2009; Peck & Theodore, 2015; Shore & Wright, 2011; Stone, 2008; Wedel, 2019). The potential data sources, therefore, are not typically predetermined in the design of a mapping study because relevant sources (i.e., people, documents, policies, events, etc.) emerge as the data develop.

The full mapping method is described in the first manuscript of this series, which includes a theoretical framework of diffusing value in evaluation; the comparison of this mapping technique to other similar methods; and the rationale for the appropriateness of the mapping method for this topic (Nolton et al., 2020b). The current mapping study was approved by the George Mason University Institutional Review Board (#1502800-1; see Appendix B for approved materials). In addition to analyzing the *Analytical Perspectives* documents, a total of 15 semi-structured interviews were conducted with key informants who could account for the changes they observed in federal evaluation over time to ground developing theories of this process (Charmaz et al., 2018; Patton, 2015b; Zilber, 2002). The following six questions guided the empirical methods:

1. Who have been the **parties** (i.e., players, social networks, and organizations) involved, both formally and informally, in federal evaluation?
2. What have been the parties' respective interests, **agendas**, incentives, goals, motivations, operating assumptions, and expectations with regard to federal evaluation?
3. Who have been the parties' **sponsors** (i.e., funders or providers of other indispensable resources) and what are their interests, agendas, incentives, goals, motivations, operating assumptions, and expectations with regards to federal evaluation activities?
4. What has been the larger context(s)/circumstances of **power** and **resources** in which the parties are embedded? How have features of the larger context(s)/circumstances constrain, enable, or otherwise affect the parties involved in federal evaluation efforts?
5. To what extent has the **structure** that facilitates federal evaluation activities or efforts institutionalized (or becoming so) even if it remains informal/unincorporated and the players change over time?
6. What has been the **ecosystem** in which the various parties operate and how does it affect federal evaluation efforts? What has been the "culture(s)" and mindsets of the milieus, organization(s), or venues in which the parties operate (i.e., norms and mindsets of a community of practice, norms and ethics of a profession) and how do these affect federal evaluation efforts?

**Participant selection.** Informants were selected based on their expertise in one or more of the following areas: federal evaluation, evaluation use/influence/capacity, or the field of evaluation. A list of known experts in these areas were initially compiled and contacted for participation (Wedel, 2019). Recruitment grew quickly, as a snowball sampling technique allowed each participant to volunteer names of other stakeholders who might also have interesting pieces of the overarching story to tell (Patton, 2015b). Their primary role (i.e., educator or practitioner), employment setting (i.e., university, government, consulting, or philanthropy), years of evaluation experience, and years in the

Federal government are shown in Table 10. Although some individuals do not have Federal government experience, they are thought leaders who have influenced the practice of evaluation at-large in the US and their perspectives on shifts in federal evaluation from an external vantage were of interest to this study.

Table 10.

Demographic Background of Interview Participants

Role	Setting	Evaluation Experience (Years)	Federal Experience (Years)
Educator	University	40+ §	0
Practitioner	Government	25	18
Educator	University	40+ §	0
Practitioner	Philanthropy	40+ §	0
Practitioner	Consulting	30+ §	0
Practitioner	Consulting	45	5*
Practitioner	Government	10	9
Practitioner	Government	50+ §	35*
Practitioner	Consulting	50+	40*
Practitioner	Government	20+	4
Educator	University	35+ §	0
Educator	University	40+	0
Practitioner	Philanthropy	25	10*
Practitioner	Government	22	29
Practitioner	Government	50+ §	40*

Note. \*denotes previously employment in the Federal government, but not presently. §denotes previous or current leadership in the American Evaluation Association<sup>5</sup>.

**Interview procedures.** Potential informants were solicited to participate in interviews via email correspondence. Based on availability and convenience, in-person

<sup>5</sup> “The American Evaluation Association is a professional association of evaluators devoted to the application and exploration of program evaluation, personnel evaluation, technology, and many other forms of evaluation.” American Evaluation Association. (n.d.). *About AEA*. <https://www.eval.org/p/cm/ld/fid=4>

meetings, video conference sessions, or telephone calls were scheduled with informants who agreed to participate. All interviews were conducted over a three-month period (November 2019 and January 2020). All participants were provided with the informed consent form and interview questions electronically ahead of the interview.

Semi-structured interview questions were tailored to be relevant to each of participants based on their backgrounds and experiences (Brinkmann, 2018; Patton, 2015b). The researcher is formally trained to conduct semi-structured qualitative interviews. She has expert knowledge of the topic and related areas, which allowed her to construct questions and prompts to elicit responses that would answer the research questions. In most cases, materials were provided to participants a week in advance with a reminder notification of the interview but, due to scheduling, some had almost one month to prepare for the interview. With the permission of the informants, all interviews were audio recorded and professionally transcribed using an online transcription service (<https://www.rev.com>). Upon receipt of the transcripts, data were cleaned for errors and areas that the transcriptionist noted were inaudible were resolved by the researcher by reviewing the transcripts alongside the audio recordings and field notes.

**Coding and analytic procedures.** Transcripts were uploaded to ATLAS.ti 8.4.24.0 for Windows for coding (ATLAS.ti Scientific Software Development GmbH; Berlin, Germany, 2020). ATLAS.ti is a qualitative data analysis (QDA) instrument, which allows data to be coded, linked, and visualized. Using a robust software for QDA is helpful to analyze multiple sources of data of varying formats (i.e., text, video, audio, image files) and/or large quantities of data. For the present study, portable document

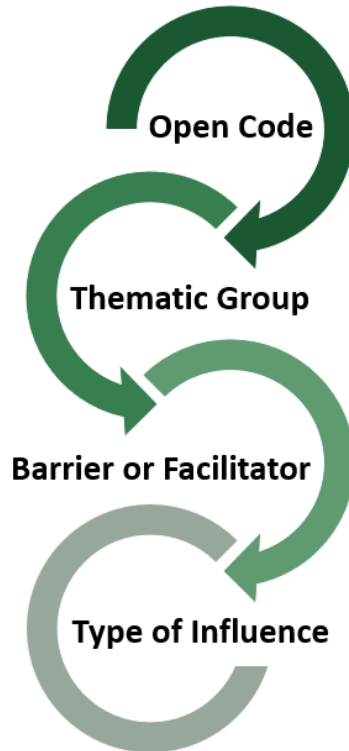
format (PDF) and Microsoft Word files were imported and analyzed in ATLAS.ti, which allowed codes within each of the documents to be shared and linked across files. This coding and analytic method served to ground and clarify existing theories based on empirical data (Charmaz et al., 2018; Patton, 2015b).

This portion of the study utilized coding techniques to identify themes and build a chronological account of influential factors. The evaluation chapters of the *Analytical Perspectives*—described in the second manuscript in this series—were also analyzed using these coding techniques (Nolton et al., 2020a). The researcher coded all of the files for continuity of labeling and interpretation. This allowed her to have first-hand knowledge of the interview and field notes that were integral to capturing the gestures and tone of quotations (Gee, 2014). Many informants offered references to resources that continued to deepen and widen a developing understanding. These resources (i.e., books, policies, documents, reports, articles, people, etc.) and others uncovered throughout the study were used to provide additional context, corroborate findings, and strengthen the credibility of interpretations (Denzin, 1978; Flick, 2018; Greene et al., 1989; Patton, 1999, 2015b). The coding methods and process of establishing validity used in this study are described in the following sections.

***Coding for themes.*** Coding text allows researchers to capture and recapture diverse codes for varying levels of depth with each pass (Carspecken, 1996). Researchers are, therefore, not limited because the design can be developmental and the researcher's data dictionaries can also emerge from the data (Neuendorf, 2017). Small units of

quotations within the interview transcripts and *Analytical Perspectives* chapters were openly coded (i.e., the initial phase of coding used in a grounded theory approach) to identify emerging themes (Campbell et al., 2013; Charmaz et al., 2018; Patton, 2015b; Peräkylä & Ruusuvuori, 2018). Open codes used to identify thematic patterns were initially developed from interview transcripts then the same and new codes were applied were applied when coding the *Analytical Perspectives* evaluation chapters.

Using this grouping feature in ATLAS.ti allowed codes to be assigned to more than one code group. In addition to coding for themes, two other code groupings were used to contextualize the codes (shown in Figure 13). For the first step, two code groups were used which sorted codes as either barriers or facilitators to institutionalizing evaluation. This was done because whole thematic code groups (e.g., “leadership”) could not be automatically considered either a barrier or facilitator. For example, there were some leaders who promoted institutionalization, but there were also some who opposed this process. Likewise, whole code groups (i.e. “legislations and policies”) did not all have the same type of influence on diffusing value in evaluation. Therefore, the second type of code grouping sorted relevant codes into types of influences (i.e., cognitive, motivational, behavioral, and structural), which were developed in the first manuscript of this series (Nolton et al., 2020b). This method of grouping allowed codes to further inform the developing theory of the institutionalization process, the ecosystem, and the influential factors involved, but not all thematic codes had an influence on the process in the Federal government and were not assigned to those groups if they did not apply.



*Figure 13.* Code Grouping and Sorting Procedure

*Note.* This figure demonstrates the code sorting process into code groups. After assignment to a thematic group, codes were also evaluated to see if they were a barrier or facilitator to the process of institutionalizing evaluation in the U.S. Federal Government. If they were, their level of influence on this process was sorted into separate code groups.

***Coding for chronology.*** The sequencing of factors over time became an important feature to document in order to recognize and make judgments about organizational contexts and circumstances. For example, it was not enough to note that the Reagan Administration had an influence on evaluation use. The chronological position of the Reagan Administration in the 1980s allowed us to make inferences about the origins and motivations for other decisions that were made in or soon after the 1980s. Therefore, mapping influential factors was done across the conceptual framework and along a

developing timeline. Where discrete influences could be timestamped with a year or range of years, it was noted directly at the beginning of the code label. When sorted alphabetically, the codes were listed first and in chronological order, which could be easily differentiated from the thematic codes. Similar to the thematic coding method, chronology codes could increase in density (i.e., coding occurrences to unique quotations) by assigning these codes each time they were mentioned in the text, which also allowed the researcher to identify factors which were mentioned by more than one source. The chronology codes were then also sorted into code groups in the same way described above for the thematic codes.

*Code cleaning and sorting.* Including both thematic and chronological codes, the initial open coding process resulted in nearly 800 unique codes. Some codes were similar in meaning, but the coder used slightly different terms to describe the essence of the quotation. After coding was completed, codes were merged and/or renamed to better reflect the intended labeling and reduce the number of codes. After refining names and merging similar codes, there were a total of 462 codes—223 of which were chronology codes. As mentioned in the second manuscript of this series, there were eight codes reserved for the institutionalization framework used to code the evaluation chapters of the *Analytical Perspectives* and five codes used in the auto-coding process to capture the prevalence of five key terms. This left a total of 226 thematic codes.

The code groups were the first pass at identifying common factors or themes. Instead of creating broad groups (i.e., “policies”) which would have included many relevant codes, the initial code groups were made to be as specific as possible (e.g.,



“evaluation policy”, “Evidence Act”, or “legislations and policies”). These code groups were then grouped into larger buckets while still maintaining subgroup differentiation to appreciate the different types of, for example, policies that were influential in facilitating or impeding the institutionalization of evaluation. Figure 14 displays an example of this process. It shows there were 21 open codes (applied to 149 quotations) assigned to the code group labeled “evaluation policy”; however, the larger category of “policies” as a factor eventually contained a total of 63 codes, including codes from two other code groups. The two other code groups were “Evidence Act”, which had 16 codes applied to 140 quotations, and another group encompassing all other “legislations and policies”, which included 26 codes applied to 130 quotations. For example, a total of 14 quotations from eight different informants (and two more quotations from the *Analytical Perspectives*) mentioned the significance of the Evidence Act and its role in elevating the field of evaluation. These quotations were coded as, “Evidence Act is the most significant event that has happened to give evaluation increased status,” which was assigned to the “Evidence Act” code group.

One caveat to the coding process for chronology is that the labels only included the year or range of years. However, this meant the coding process would fail to capture the sequencing of events within a single year which, in some cases, might be important to capture. Sorting alphabetically meant events that happened in the same year were then ordered alphabetically by the word(s) following the four-digit year. This was not necessarily an accurate depiction of the true ordering of events, but careful consideration was given whenever there was a grouping of codes in the same year. Where sequencing

within a year was known and important to note, alphabetical letters were added to provide the correct orientation (e.g., 2003a and 2003b). Clarification was only necessary when an event occurred in direct response to another and the sequencing was critical in understanding the motivation of the reaction.

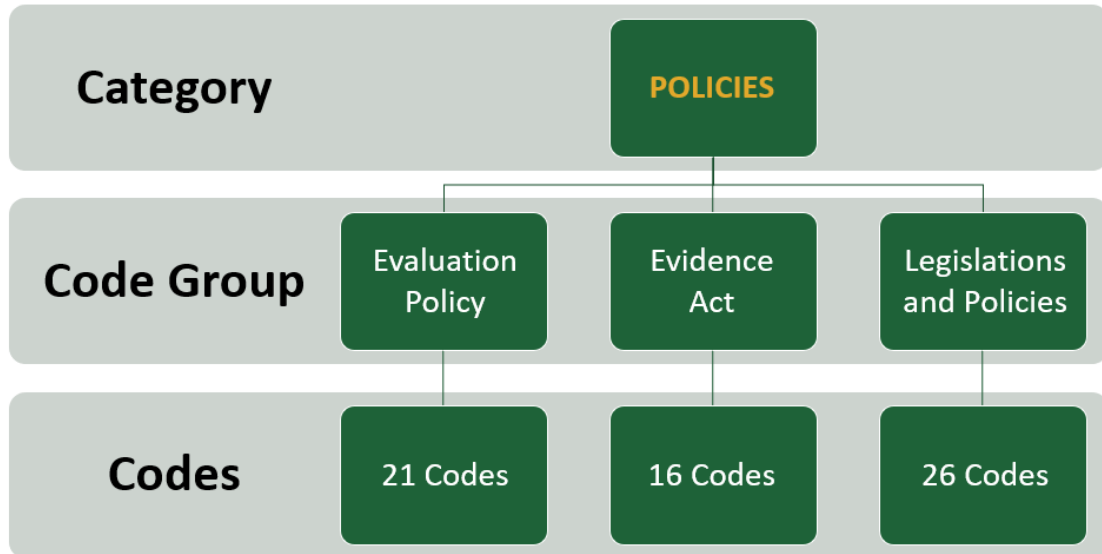


Figure 14. Example of Code Sorting and Bucketing Process

*Note.* This figure displays an example of how code groups were sorted and combined to create categories of factors that influenced the process of institutionalizing evaluation.

For example, the Institute of Education Science (IES) at the Department of Education (ED) issued a call for response to an impending statement that they were going to publish about prioritizing experimental designs—specifically, randomized controlled trials (RCTs). The American Evaluation Association (AEA) responded by encouraging IES to not make such a strong claim about the preeminence of RCTs and effectively dismiss other rigorous methods of systematic inquiry. Later, IES published their statement without substantial edits and no evidence that they had considered suggestions

made by the AEA. The ordering of these events was critical in understanding why informants reported so much tension and frustration during this period of time. The impact of this event will be described more fully in the results section of this manuscript.

Another challenge was—although they were given time to prepare for the interviews—some informants could not recall exact dates, especially participants who were recalling events that occurred several decades ago. In most cases, the exact year may not have been important to note, but statements like, “sometime in the late 90s,” translated to a code such as, “1990s later years...,” to make the code sortable within the existing chronological structure. Participants were also encouraged to follow-up with the interviewer after the completion of the interviews if they remembered additional information that would be pertinent to include. Several participants did contact the interviewer after the interview to provide additional details or resources to corroborate or elaborate their statements. The interviewer also researched and obtained additional information when the informants were unable to provide detail. *Establishing rigor and validity.* As mentioned in the second manuscript of this series, there were many different techniques employed to ensure rigorous and credible methods (Nolton et al., 2020a; Shenton, 2004). The first manuscript in this series introduced the conceptual framework and mapping method used for this study (Nolton et al., 2020b). The researcher arrived at the proposed framework and method of inquiry through a thorough review of existing models and methods to determine the appropriateness of the approaches employed for this study (Patton, 2015b; Shenton,

2004). A grounded theory approach was used to continually revisit and revise the developing theories throughout the study (Charmaz et al., 2018; Patton, 2015b).

Technologically, the use of audio recordings of the interviews and a robust QDA platform both allowed more accuracy and capability compared to coding or transcribing by hand, respectively (Shenton, 2004). These methods were supplemented by detailed field notes. The study was also designed by a researcher who: has expert knowledge about the topic and is without strong political leanings; built her qualitative research skills in conducting these analyses; carefully selected a relatively large number of diverse informants for interviews; triangulated methods and data by comparing multiple alternatives and sources; grounded her coding technique to an existing theoretical framework and revisited theories by grounding them in empirical data; thoroughly documented her methods and interpretations; and received approval from peers to scrutinize the process. The following sections further describe these efforts to establish validity.

Prior to conducting this study, the author had worked in a centralized evaluation unit of a federal agency and studied research methods, evaluation, and public policy. Academic and professional training in these areas allowed the researcher to design a robust qualitative study with subject matter expertise about the intricacies of the field (Patton, 2015b). Through these experiences, the researcher discovered that the *Analytical Perspectives* demonstrated government value in evaluation (i.e., investment in the worth of evaluation). The researcher began building connections with key players in the field who could help her learn about federal evaluation. The researcher used these

relationships to continue building a network within the evaluation community and sought out individuals with diverse backgrounds to participate in interviews.

Including 15 interviews is a relatively large number for a qualitative study. All informants willfully participated because they were able to decline or withdraw at any time, even after scheduling or completing the interview (Shenton, 2004). The researcher established rapport with informants by meeting several of them in-person prior to requesting their participation; being candid about the use of their interviews for this study; and guaranteeing anonymity of their responses. These approaches allowed informants to speak honestly and freely without fear of reprimand for sharing their perceptions and opinions about past and current events (Shenton, 2004).

Their diverse backgrounds and demonstrated expertise were also integral to establishing credibility of interview data (Shenton, 2004). The inclusion of informants with no experience in the Federal government was intentional in providing a different vantage that could triangulate the perspectives of informants reporting on the same or similar events from their perspective within the government (Flick, 2018; Shenton, 2004). Staggering informants by background (i.e., not interviewing all Federal employees first or vice versa) allowed their stories and perspectives to be compared and checked with each other throughout the study. The researcher also performed member checking frequently during the interviews as a way to immediately verify how informants' statements were heard or interpreted (Morse, 2018; Shenton, 2004). Informants were allowed to strike, add to, or amend what they said at any point during or after the interview.

As described in the second manuscript of this series, the researcher consulted two

evaluators with expert knowledge in QDA and ATLAS.ti for peer debriefing (Barber & Walczak, 2009; Shenton, 2004; Spall, 1998). Comments and memos were documented in ATLAS.ti and adjustments to the process were made and documented, as necessary. The coder and peer debriefers met regularly to discuss the coder's methods, findings, and interpretations as they were developing (Campbell et al., 2013). Peer debriefers acted as 'critical friends' who helped discuss and strategize the grounded theory approach (Barber & Walczak, 2009; Patton, 2015b). Critiques and questions also offered a way to triangulate the analyst's process and interpretations (Patton, 2015b; Shenton, 2004). The coding techniques described in this manuscript were used for both the interviews and the *Analytical Perspectives* documents making pattern and comparative analyses possible across multiple data sources (Flick, 2018). Other forms of external review were achieved by engaging IRB review, consulting peers for scrutiny, and other methodological and content experts who did not participate as informants (Patton, 2015b; Shenton, 2004).

## **Results and Interpretations**

**Actors, agendas, and assumptions.** As previously mentioned, there were a total of 223 chronology codes that spanned over a 67-year period (1953-2020) that emerged from the interviews and *Analytical Perspectives* evaluation chapters. The most significant events from this list were separately grouped to highlight pivotal moments in history toward the institutionalization of evaluation. These events were selected because of the emphasis placed on them by informants; how frequently these events were mentioned in interviews or documents; and the impact (i.e., positive or negative) they had on

evaluation. As mentioned earlier, the significance of these events were also triangulated with other sources (Patton, 1999).

Figure 15 displays these significant events along a timeline to further illustrate the sequencing of these events over time. Events shown in dark green were ones that were directly related to the Federal government and items in a lighter green specifically involve people within the Federal government. Events shown in gold are ones that occurred in an unspecified or international context. Events in blue involve the American Evaluation Association (AEA) and events in red are ones that had negative influences on or deinstitutionalized evaluation—as stated by informants—regardless of the setting in which it occurred. Although other influential events occurred outside of this list, whether mentioned in interviews and documents or not, the following provides an overview of the most significant events which emerged from the empirical data. Elaborations provide additional context according to how informants justified the impact of these events on diffusing or restricting value in evaluation.

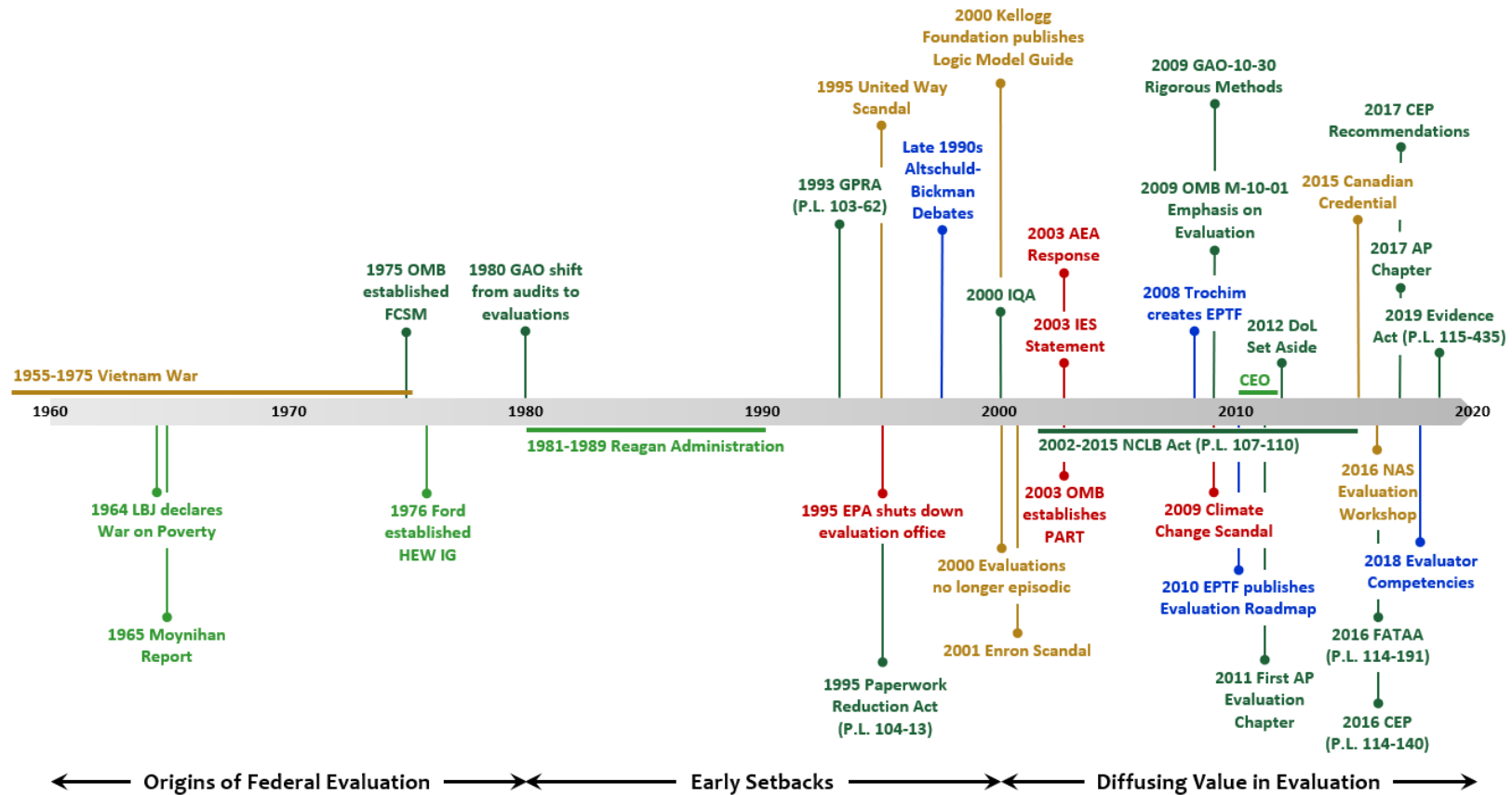


Figure 15. Mapping a Chronology of Significant Events



*Origins of federal evaluation.* Informants reported that there were two main events that influenced the introduction of evaluation in the Federal Government: the Vietnam War (1955-1975) and President Lyndon B. Johnson declaring a “War on Poverty” (1964). It was stated by informants that wars, in general, placed a great economic demand on countries and created a greater need to justify government investments. President Johnson’s famous State of the Union address launched a series of social programs and policies aimed at addressing the growing rates of poverty in America (Bailey & Duquette, 2014). Shortly after, informants noted a controversial report published by Senator Daniel P. Moynihan in 1965, which highlighted the underlying sociocultural issues that were perpetuating poverty (Moynihan, 1965). This called the effectiveness of existing social programs and policies into question, which inspired evaluations of those initiatives.

“Moynihan's book and some research that was showing that welfare was not helping people, but on the contrary, making them dependent, so that there were all sorts of questions being raised politically, and the idea that, wouldn't it be nice if we had some scientific studies that could look at these questions in there?”

In 1971, the British Central Policy Review Staff (CPRS) was founded as a part of a major reorganization of government to make it more efficient and effective. They were a diverse team of experts who produced high quality evaluations of policy decisions to inform strategy (Blackstone & Plowden, 1988). Twelve years later, however, Prime Minister Margaret Thatcher disbanded CPRS because it did not align with her political agenda, which effectively deprioritized evidence utilization in policymaking processes. Meanwhile, the U.S. Federal Government established the Federal Committee on Statistical Methodology (FCSM) in 1975 and President Gerald Ford created the Office of

the Inspector General for the Department of Health, Education, and Welfare (HEW) in 1976. The then U.S. General Accounting Office (GAO; established in 1921 and changed its name in 2004 to the Government Accountability Office) also began shifting their focus from financial audits to performance audits. In 1980, GAO established the Institute for Program Evaluation, which became the Program Evaluation and Methodology Division from 1983-1996.

*Early setbacks and missteps for evaluation.* On the heels of these changes came the inauguration of President Ronald Reagan in 1981, which completely shifted the use of evaluation in the Federal government. His particularly conservative brand of economic policies was later dubbed “Reaganomics” (Blanchard, 1987). Several informants attested instead of using evaluations for program improvement and organizational learning—classic Democratic uses for evaluation—President Reagan restructured government by,

“looking at 27 programs and what their budgets were for the last five years, take an average, reduce it by 25%, and divide it among the 50 states...we’re going to give the money away to the states and let them do what they think is best.”

In fact, informants reported that Reagan’s use of evaluations caused them to be seen as a mechanism to cutting funding or whole programs, which led to fear of and apprehension to evaluation. Of the informants who were working in evaluation—in any setting—in the 1980s, they reported that the impact of Reaganomics on evaluation were profound. Even informants who were not working in the Federal government felt the impact of Reaganomics:

“There had been a lot of funding prior to Reagan both for evaluation—it was a growing field—and also for work on utilization...[with] the Reagan Administration, boom, the bottom fell out, funding dried up.”

The organizational culture for evaluation across the Federal government during this time was scarce, which caused many evaluators (especially at HEW) to leave the government.

In 1993, the Government Performance and Results Act (GPRA; P.L. 103-62) was enacted, which brought greater emphasis on developing performance measures as evidence for program performance. This Act and the concurrent development of robust methods of inquiry contributed to the worldwide “Evidence Movement” (Hansen & Rieper, 2009). Although GPRA had the potential of promoting evaluation as an essential evidence-generating function, upon closer examination, it primarily focused on performance measurement and did not require evaluations.

“GPRA did not prioritize evaluation. It barely even mentioned it. It was performance measurement that was required.”

“[GPRA] encourages, but does not require program evaluation, which was a signal not to do it...the Chief Financial Officer had no intention of investing staff and time into conducting or supporting program evaluations.”

Informants described this as a misstep in evaluation history because the field of evaluation failed to take ownership of the opportunity GPRA could have provided to prioritize evaluations. Instead, the focus (and funds) shifted to support performance measurement and led to the disbandment of evaluation offices at agencies like the U.S. Environmental Protection Agency (EPA) in 1995. As such, the abandonment of intentionality in differentiating terminologies led to the conflation of evaluation with other disciplines, which resulted in evaluation to be considered an activity with some application across several disciplines rather than a distinct discipline that uses varying methodological and analytical tools (Madison, 2000; Scriven, 2008). Informants reported being frustrated with the lack of clarity when communicating with colleagues,

“Every time I did an evaluation, some seemed to act as if I was speaking a foreign language. It was very annoying.”

*Demand for credible evidence.* Around this time, the Paperwork Reduction Act (PRA) of 1995 (P.L. 104-13) was enacted mandating Federal government information activities to be more efficient, effective, and economical by reducing the burden of paperwork, records management, threats to privacy, and number of information collection requests. Five years later, pursuant to §515—commonly referred to as the Information Quality Act—of the Treasury and General Government Appropriations Act (P.L. 106-554), this legislation amended PRA to state information (i.e., including statistical information) generated by the government must also abide by standards of quality, objectivity, utility, and integrity (Copeland & Simpson, 2004). Later, the FY 2017 evaluation chapter of the *Analytical Perspectives* highlighted the Information Quality Act as a significant mandate to promote credible evidence.

Additionally, three scandals sparked international attention and reinforced the urgency for greater evidence accountability—the United Way Scandal in 1995, Enron Scandal in 2001, and Climate Change Scandal in 2009.

“But in reality, I think the times we were most effective was when some wakeup call made people have to think about planning or evaluation. And at that point, you've got this teachable moment and going in and helping facilitate the discussion ends up being a better way for the teaching points to really take.”

Although the first two events served to call evaluators to action because company funds were not handled appropriately at these corporations, the Climate Change Scandal involved a breach of emails from within the Climatic Research Unit at the University of East Anglia (Norwich, England), which suggested climate change was a hoax. Although

researchers later denied the authenticity of those claims, the impact of this event led to widespread public distrust in the credibility of evidence provided by the academy and government, especially in the midst of already contentious debates about the validity of climate change research (Freudenburg & Muselli, 2013).

*Proliferation of robust evaluation methods.* Informants who were consultants around this time saw more government funding for evaluation and stated the rapid proliferation of robust evaluation theories and methods continued through the early 2000s. In 2000, the Kellogg Foundation commissioned the development and dissemination of the “Logic Model Development Guide,” which became a widely used tool in evaluation practice. Informants felt that increasing accessibility helped build capacity of evaluators and provided uniformity to practice,

“the Logic Model Development Guide...was the first publication that really codified and popularized the technique and made it available.”

It was also around this time that informants felt evaluations were no longer episodic (i.e., infrequent), but that they were becoming ubiquitous. They reported,

“I think 20 years ago [circa 2000] it popped up to me that what I was seeing is not just an episodic evaluation...evaluation has become nearly ubiquitous across lots of social Federal, state, and local programs.”

“[the Federal government] went through a period of backing off from evaluation about 20 years ago [circa 2000] and now...it’s continually becoming more intense and the demand has been higher.”

However, informants also reported that the development of robust methods led to ranking the robust-ness and scientific-ness of these methods. This created groupings and division between quantitative and qualitative methodologies, which are now known as the ‘Paradigm Wars’ (Lincoln et al., 2006; Patton, 2015b).

*Prioritization of experimentation.* Informants recalled an example of this when OMB Director Peter Orszag published a memo (M-10-01) in October 2009 titled, “Increased Emphasis on Program Evaluations.” However, this official memo promoting the production and use of evaluation actually included a call for proposals for a new evaluation initiative that defined “rigorous” evaluations as those that employed RCTs. The instructions for proposals required that agencies provide details about the rigor of their evaluation with a focus on experimental methods.

“OMB issued a memo about the importance of evaluation, but it was very pro-RCT...do RCTs, [because] RCTs are rigorous.”

One month later, GAO released a report (GAO-10-30) titled, “Program Evaluation: A Variety of Rigorous Methods Can Help Identify Effective Interventions,” which acknowledged several other methods that could be useful in studying impact, not just RCTs. In fact, this report also discussed instances where random assignment was not appropriate for evaluations. However, there were several other areas of government that were on board with endorsing experimental methods as the gold standard.

As briefly mentioned earlier, the IES at ED released a statement in 2003 prioritizing RCTs as the most rigorous approach to generating evidence. Prior to the publication, the AEA submitted a rebuttal to an open call for response stating that RCTs are not the only studies that may lead to inferences about causality (American Evaluation Association, 2003). AEA’s position statement also advocated for robust qualitative methods that were not mentioned in the IES statement. Despite AEA’s remarks, IES published the statement with no amendments or incorporation of recommendations made by AEA (Institute of Education Science, 2003). Informants admitted this led AEA leaders

and members to question the Association's prominence as the premier organization representing the field of evaluation, if unable to influence a major government statement on evaluation.

“How is it we think we're the ‘American Evaluation Association’? If some agency was thinking about a major change of regulations around medicine, they're not going to put that out for review in the Federal Register before they've talked with people at the American Medical Association. We started using the phrase ‘not at the table’.”

Despite establishing Guiding Principles for the field of evaluation in 1994 (American Evaluation Association, 2004), these efforts did not seem to be enough to elevate AEA as a leading voice to inform field-defining decisions. Informants recalled that, in being inclusive of varying evaluation methods, pluralism in AEA was misunderstood to condone anti-rigor and the field needed to develop professional standards.

“In some ways the field has evolved and grown tremendously, and in lots of great ways. It's been very inclusive, and that's one thing that I've always loved about AEA, nobody was ever excluded.”

“We also got a reputation, at the association, as being anti-rigor and anti-quantitative evaluation.”

***Addressing skepticism about evaluation.*** These concerns resurfaced previous discussions from the late 1990s, when a series of articles in the *American Journal of Evaluation* addressed the topic of legitimizing evaluation through professionalization (Altschuld, 1999; Bickman, 1999; Worthen, 1999). Some of the most prominent thought leaders in evaluation were unable to agree on the direction the field of evaluation should take toward professionalization, but the discussions raised many existential questions about defining the practice and identity of evaluators. AEA struggled with pursuing credentialing because experts like then (1999) AEA President Michael Scriven worried

about the litigiousness of America and felt pursuing this level of professionalization would have consequences to consider.

“Scriven and the Board said, ‘uh uh, not doing it...the US is so litigious that anything we do, somebody is going to sue us, and we don’t want to get in that business’.”

Informants stated that there was good reason to worry about this issue because there were events around this time that caused people to be dismissive or weary of evaluation.

In 2002, the No Child Left Behind Act (P.L. 107-110) was enacted by President George W. Bush, which shifted all educational priorities and funding for evaluations of educational programs and policies to development, monitoring, and support standard assessments.

“No Child Left Behind gets past... now they had to test. They had to add budget for tests, [which] cut all the evaluation money out.”

This changed the culture of evaluation in education and changed the way government generated evidence about the effectiveness of educational programs. One year later, OMB established the Performance Assessment Rating Tool (PART), which was used to ensure program improvement was informed by evidence (U.S. Office of Management and Budget, n.d.). Similar to GPRA, although PART placed an emphasis on evidence-building activities, evaluation was not a prominent part of the instrument with only a three-point scale for assessing evaluations. As such, PART offered many opportunities, but also created several challenges given the vague criteria provided as guidance for program evaluations (U.S. General Accounting Office, 2004). Several informants commented on the negative influence PART had on evaluation and there appears to be lingering skepticism about the legitimacy of evaluation from that period because it was



undervalued in both GPRA and PART. This concern continues to make evaluation experts weary of new policies.

“The [PART] tool had different color-coded dots for different programs, but the lingering effect of that is very negative in the agencies. They viewed it as a way to just cut budgets or eliminate programs, which in some cases, it did happen that way...so there’s still lingering skepticism about evaluation and performance management because of PART. It’s still there.”

*Diffusing value in evaluation.* Since then, evaluation in the Federal government appears to be on the rise. In effort to give evaluation a seat at the (proverbial) federal evaluation table and prevent being excluded from future field-defining language, AEA President William Trochim established the Evaluation Policy Task Force (EPTF) in 2008. This group put AEA at the forefront of evaluation discussions in the government to stay ahead of impending statements or legislations regarding evaluation practice and policy. The EPTF published “An Evaluation Roadmap for a More Effective Government” in 2010, which was later revised in 2013, 2016, and 2019 (American Evaluation Association, 2019). Informants state this document allowed AEA to voice consensus to influence the institutionalization of evaluation in the Federal government.

“In writing the roadmap, we really tried to represent a broad constituency...and used the roadmap as guidance for creating evaluation policies.”

In 2018, AEA also published the “Evaluator Competencies,” which described hard and soft skills required of a competent evaluator, which were modeled from the criteria for the Canadian Evaluation Society Credentialed Evaluator designation (2015), which pushed the international conversation of professionalizing evaluation.

In 2010, Dr. Jean Grossman became the first Chief Evaluation Officer (CEO) for the U.S. Department of Labor (DoL). As informants could recall, Dr. Grossman was the

first CEO in the Federal government. The creation of this position was a structural influence after years of establishing a culture of value and trust in evaluation at the DoL. In 2011, the *Analytical Perspectives* featured its first chapter on “Program Evaluation” under the Obama Administration. In 2012, DoL allocated a set aside fund for evaluation. This practice was initiated by the Department of Health & Human Services (HHS) in 1988, but no other agency had made a similar investment in evaluation until DoL (Riggin et al., 1995). In 2016, the Foreign Aid Transparency and Accountability Act (P.L. 114-191) was enacted, which required foreign aid agencies to develop and implement rigorous policies and plans to conduct evaluations.

“Once you have an evaluation policy, it takes a life of its own and people are able to develop more of a culture and demand.”

That year, President Obama also enacted the bipartisan Commission for Evidence-Based Policymaking (CEP) Act (P.L. 114-140), which allowed the government to study the factors for inefficient processes. Notable evaluation leaders—such as Dr. Demetra Nightingale, who served as the CEO of DoL from 2011-2016—provided input to the CEP. As mentioned in the second manuscript of this series, the 2017 *Analytical Perspectives* was the first evaluation chapter to include guidance on establishing evaluation principles, evaluation policies, and centralized evaluation offices (Nolton et al., 2020a). The proposed principles were discussed and agreed upon at a public convening of evaluation experts (National Academies of Sciences, 2017). In 2017, the CEP delivered recommendations to Congress, which became the impetus for the bipartisan Foundations for Evidence-Based Policymaking Act (P.L. 115-435)—

commonly referred to as the Evidence Act—enacted in January 2019 by President Donald Trump.

“The Foreign Aid Transparency and Accountability Act required international aid agencies—all foreign aid—to have rigorous evaluation policies and evaluation strategies...and now the Evidence Act. All of them are moving toward institutionalizing the role of evaluation in the Federal government.”

According to this timeline, it appears federal evaluation enters a new era about every 20 years. It will be interesting to see how the next 20 years following the Evidence Act will unfold. The following section further describes the context, circumstances, and barriers/facilitators of institutionalization.

**Context, circumstances, and facilitators of institutionalization.** A total of 47 codes were assigned as barriers to institutionalizing evaluation and 46 that were facilitators. Both groups were analyzed to identify common themes (Figure 16). Eight themes emerged for each group and many themes in one group had an opposing counterpart in the other category. For example, the most common barrier that emerged was defunding or divesting in evaluation. In contrast, the most common facilitator for institutionalizing evaluation was increasing funding or investments in evaluation.

Other barriers not previously mentioned as negative influences were slow processes for institutional change, turnover or transitions in leadership, political agendas, lacking a clear definition and identity for evaluators, and risk aversion. Informants shared that the momentum of institutional change must be constant and continuous. When the process was too slow, high staff turnover rates or a change in leadership (i.e., reaching a term limit or otherwise) in the government led to a lack of continuity and disrupted the drive necessary to continue forward progress.

“new people come, then building and maintaining a culture of learning where there’s trust between the evaluators and the program or policy people is ongoing work.”

In situations where new leadership came in with a new political agenda, their priority was to create a distinct and unique platform. Therefore, reorganization may not be a genuine opposition to evaluation as much as an attempt to diverge from “the other guy’s” mission.

Many informants also reported that the general sense across the government is that evaluators lack a clear identity or a practice that can be clearly defined. Informants affirm statements made in the first manuscript highlighting the ongoing struggle with professionalizing evaluation. The issue with not professionalizing quickly enough is that trust and value in evaluation might be slipping away.

“I would say even among the people that I hired most of them did not brand themselves as evaluators. They thought of themselves as analysts who did program evaluation as opposed to program evaluators... I think there are people who just felt that it just didn't matter to the agency, even if it mattered at Federal government level. Part of that was driven by the lack of resources given to it.”

Additionally, evaluations can inform program and policy decisions by measuring impact of innovative approaches to tackling social issues. Agencies that are averse to taking risks (i.e., investing in programs or policies that have no guarantee of having a positive impact)—or fear evaluations will reveal negative impacts—shy away from doing and using evaluations and performance measurement. OMB tried to dissuade this fear by encouraging agencies to take reasonable risks in program investments. They did this by introducing a tiered grant model in the FY 2015 *Analytical Perspectives* evaluation chapter whereby agencies could only be eligible for the highest tier of funding if they implemented and sought to test the impact of new approaches.

- Defunding/divesting in evaluation
- Deprioritizing evidence generation and use
- Damaged reputation
- Slow progress in changing organizational culture and/or structure
- Turnover or transitioning leadership
- Political agendas
- Lacking a clear definition of evaluation and identity for evaluators
- Risk aversion impeding innovation and growth

## BARRIERS

## FACILITATORS

- Increased funding/investing in evaluation
- Demanding accountability and greater ROI
- Giving evaluators an elevated status in an organization
- Advocating for evaluation
- Desire to increase effectiveness and efficiency
- Focusing on evidence and evaluation utilization
- Identifying evaluation as an esoteric field of practice requiring a versatile skillset
- Establishing value and demand in evaluation

*Figure 16.* Common Barriers and Facilitators of Institutionalizing Evaluation

Other facilitators of institutionalizing evaluation included demanding accountability, a greater desire for more effective and efficient government, and placing a focus on evidence utilization. As mentioned earlier, informants noted well-publicized scandals as positive influences on diffusing value in evaluation. These events were credited to a lack of accountability and processes to properly manage complex organizations. After these events occurred, people began demanding more evaluation for monitoring and organizational learning. They became cautionary tales that inspired more organizations to institutionalize evaluation as a preventative measure.

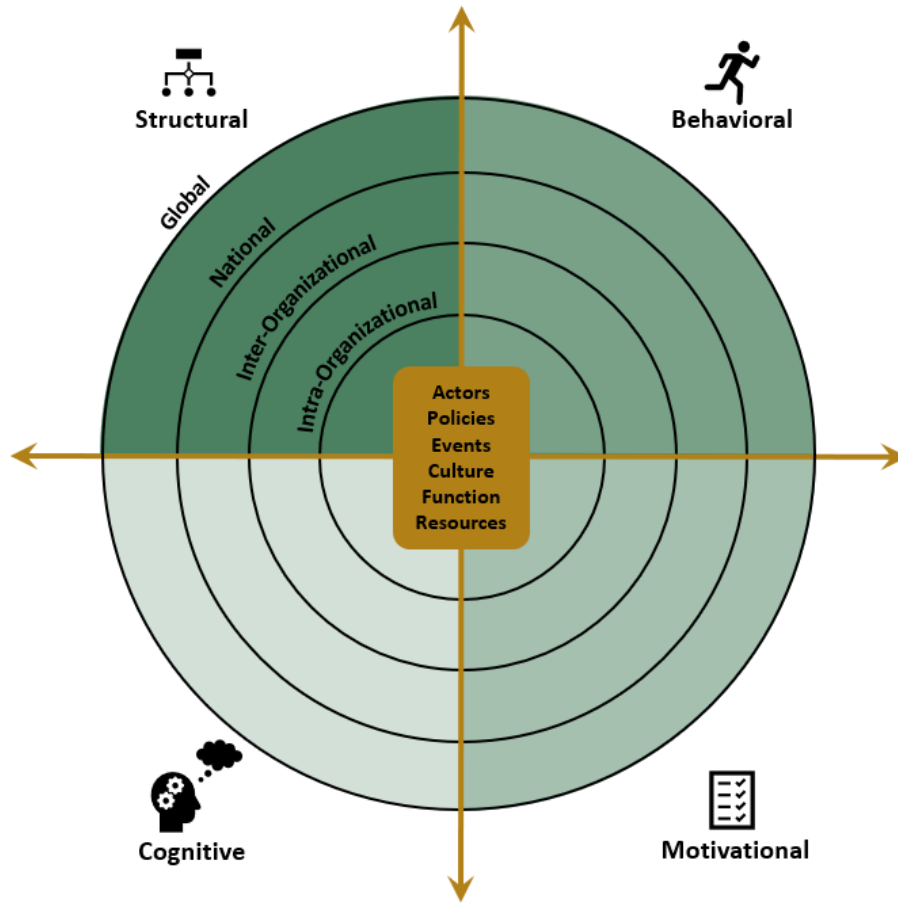
“The scandal at the national level led to them saying, ‘you now must measure outcomes’.”

The desire to make government more effective and efficient was a major motivation for increasing a demand in evaluation. Agencies were interested in testing their policies and programs to discover ways in which they could more justly and responsibly serve Americans. This emphasis led to a prioritization of producing and using evidence to inform program and policy decisions. Evaluations, therefore, were seen as a useful tool of inquiry to provide that evidence to make responsible and effective budget decisions.

**Institutional ecosystem and process.** After investigating the codes for types of influence, the themes reflect what was discovered in the content and discourse analyses in the previous study suggesting that the pattern of influences is cumulative (Nolton et al., 2020a). Interview data corroborate the pattern observed in the *Analytical Perspectives* that the process of diffusing new ideas occurs as a progression from cognitive, motivational, behavioral, to structural influences. Therefore, Figure 17 displays a modification of the original conceptual framework presented in the first manuscript of

this series to reflect the observed relationship between types of influences, types of factors, and layers of the organizational ecosystem (Nolton et al., 2020b).

Beginning from the bottom left-hand portion of this figure, the new model in Figure 17 illustrates that the factors must first influence cognitive affect before (moving to the right toward) motivating, acting on (behavioral), and building an infrastructure for institutional change. This was observed by following the logical patterns of events and factors that informants discussed (e.g., when “X” happened, then “Y” happened; or the only reason why “Y” happened was in response to “X”) and analyzing the sequencing and linkages between the codes. Instead of the types of influences diffusing evaluation across the organizational levels, the factors (i.e., actors, policies, events, resources, etc.) are the vehicles of diffusion that permeate in all directions. The types of influence those factors have on institutional change are the resultant outputs, which manifest as different outcomes. As such, Figure 18 adds a modification to a figure that was shown in the second manuscript in this series to represent the corresponding manifestations of increasing value in evaluation (Nolton et al., 2020a). This progression in valuing evaluation is mapped to the stages of institutionalization adapted from Meyer et al. (2018), which substantiates the relationship between these manifestations toward institutionalizing evaluation.



*Figure 17. Mapping the Process of Influence on Diffusing Value in Evaluation*

*Note.* This figure displays the observed process of diffusing value in evaluation in the U.S. Federal Government based on empirical data. The process is cumulative and build from increasing levels of influence—cognitive, motivational, behavioral, and structural.

As stated in the second article (Nolton et al., 2020a), factors or language that influence cognition place evaluation in relation to other known field meanings, expose it as an accessible concept, and make it appealing to adopt. The resultant manifestations of building tasks that signify cognitive influence is an institutional belief that the field of



evaluation is esoteric in that it requires specialized knowledge and skills (Abbott, 1988; Freidson, 1986; Gee, 2014; Li, 2017).

"It also had to do with the kind of culture that was in that agency. How ready they were to move on evaluation or not. So that's the answer."

However, beliefs alone do not diffuse value in evaluation in complex organizations.

Colyvas and Jonsson (2011) previously distinguished ubiquity from legitimacy by providing a 2x2 matrix which showed it was possible for a new idea to be diffuse, but not accepted in the organizational culture. For example, compliance to a federal mandate to do and use more evaluations but not having an organizational culture that embraces evaluations as an essential function. In these cases, informants stated that if the mandate were to be lifted, evaluations would almost immediately cease.

"It's very hard to sustain a culture of evaluation in my experience. It really does take years and lots of, 'two steps forward, one step back'."

This is what Colyvas and Johnson described as a consequence of lacking "stickiness," (p. 30) that cause novel ideas to eventually lose traction.

## INFLUENCES



### PLACEMENT

Alignment of evaluation with a field of meanings

### EXPOSURE

Making evaluation accessible

### APPEAL/MOBILIZATION

Making evaluation relevant and attractive



### TYPIFICATION

Generalizing the meaning and applicability of evaluation

### EXPLANATION

Creating a link between evaluation and a relevant problem



### JUSTIFICATION

Creating legitimacy for evaluation

### TAKEN-FOR-GRANTEDNESS

Establishing and maintaining evaluation as an essential function



### DIFFUSION/TRANSLATION

Making evaluation communicable across all audiences/contexts

## MANIFESTATIONS



Believe evaluation is special



Believe evaluation is esoteric



Evaluation is valued



Advocacy for evaluation



Investment in evaluation



Utilization of evaluation



Building Capacity for evaluation

Figure 18. Manifestations of Influences Toward Institutionalization

Value in evaluation (i.e., belief and investment in worth) must then influence motivation, which results in greater investments in performing evaluations. Motivation was described by informants as an increased political will to support evaluation as a viable solution to policy issues. As a precursor to influences on behavioral, motivational influences were those that inspired individuals at any level of government to organize efforts to mobilize institutional change. However, actions are necessary to codify beliefs and strategies to diffuse value in evaluation. Therefore, behavior must be influenced to inspire greater advocacy for, investment in, and utilization of evaluation. These outcomes are effectively driven by various factors, for example, actors who voice their advocacy for evaluation, resources that are allocated to funding more evaluation, and policies that mandate the use of evaluation for evidence-based policymaking.

“To me, culture is not separate from capacity or resources, but culture drives whether there’s capacity and whether there is a willingness to take resources from one area and move it to another. Culture drives those decisions.”

Lastly, structural influences lead to changes to the institutional structure to build greater capacity and use of evaluation. As such, structural changes translate evaluation as an essential function that is diffuse and embedded in the organizational culture.

The codes sorted to groups that classify the type of influence code it had on institutionalizing evaluation were also analyzed for common characteristics for levers of change. Factors had both positive and negative effects on influencing value in evaluation. Factors that positively influenced thoughts and feelings of evaluation (i.e., cognitive) were the establishment of standards and principles that set evaluation apart as a specialized field. Informants reported that observing leadership buy-in, seeing the benefit

of evaluation to their work, and receiving set aside funding for evaluation motivated value in evaluation.

“It's quite difficult if you're talking about sustainability because you have so many systems, relationships that you have to deal with that you didn't before...I think the important thing is to get managers to understand that there has to be some funding to that.”

When motivated, actions (i.e., mostly enacting policies and publishing statements) codified value in evaluation and increased utilization of evaluation. These behavioral influences then inspired mechanisms to implement and sustain institutional change through structural changes like creating new occupational roles and operational units that support more production and utilization of evaluation.

In contrast, negative influences on cognitive affect were the result of misuse or misevaluation, which gave evaluation a bad reputation. Misconceptions about evaluation and lacking a delineation between evaluation and other fields of inquiry also caused people to not believe evaluation was special. For example, informants mentioned that some of their colleagues in the Federal government did not recognize the added value that evaluation could bring to policymaking beyond methods like economics, performance measurement, or statistics. They viewed evaluation as a burdensome practice which demotivated individuals from using evaluations. Informants stated that this caused some agencies to divest in evaluation and some agency leaders even spoke out against evaluation because it was viewed as a superfluous investment.

“Statements were made along the lines that, ‘Everyone knows what evaluation is, people do evaluation daily at their desk. We don’t really need a unit to support program evaluation because everyone knows what it is’.”

If not seen as a priority, no structural changes would be necessary to create a greater capacity to do evaluation. In fact, in some cases, structural changes were made to prioritize other functions over evaluation, such as establishing performance management and statistical units then subsuming evaluation under these entities.

It was also reported that the pace of progression was varied and punctuated within micro-level systems of the Federal government, which led to some stages of the process to occur simultaneously while others had long lulls between notable points of change. Additionally, the process described by informants was not a linear model. Structural changes to build greater capacity for evaluation made positive influences on how organizational actors perceived (i.e., think and feel) institutional value in evaluation. In fact, there is a continuous feedback mechanism where forward progress through these levels of influence further reinforce previous levels of influence. For example, increasing organizational investments (i.e., when OMB allocates more resources to conducting evaluations) in evaluation is an action that fortifies motivation and positive thoughts or feelings toward value in evaluation.

Alternatively, although there could be forward progress up to some level of influence, informants revealed that factors with a negative influence at any stage of the process could undo prior progress and cause the organization to steer away from further diffusing value in evaluation.

“I think what we do get is people at least feel like they have to pay homage to evaluation as something important. When you look at how many people are investing a lot in it and doing it competently, it's less.”

However, it was unlikely to successfully course correct with a positive influence at a later stage if negative influences at early levels of the process were instituted. For example, revisiting a previous example, policies that were implemented to require evaluations without first establishing value in evaluation led to evaluations being viewed as an activity done merely for the sake of compliance. This change was not sustainable and only resulted in an annoyance and frustration from program staff with needing to check a box for evaluation. The following section will discuss the implications of these findings for institutionalizing evaluation in the U.S. Federal Government.

### **Discussion and Implications**

Although the progress toward institutionalizing evaluation has seemed to ebb and flow in the Federal government over the last 60 years, findings suggest there are trends that can be used to guide institutional change. A variety of factors can have an effect on any level of influence or layer of the ecosystem. These factors can be strategically engaged to influence cognition, motivation, behavior, or structure depending on the needs of an organization. The implications for the field of evaluation and Federal government—separately or in unison—are outlined in the following sections to help readers operationalize the framework for diffusing value in evaluation to stimulate local influences on institutionalization.

**Cultivating a culture of evaluation.** Building a strong organizational culture of evaluation requires influencing the way people think and feel about evaluation (Green et al., 2009; Nolton et al., 2020a; U.S. General Accounting Office, 2003b). As found in this study, successful influences on institutionalizing evaluation began by cultivating a value

in evaluation to motivate behavioral and structural changes. Therefore, strengthening the organizational culture of evaluation is the most essential and fundamental aspect of institutionalization (Bolman & Deal, 2017; Warrick, 2017). Institutionalization requires a careful balance between authority and values, which can be intrinsically or extrinsically motivated and may manifest in organizations in different ways (Shah et al., 2017; Strang & Meyer, 1993; Zilber, 2002). It is a political process whereby individuals within an organization evaluate the morality and value of a new idea then rationalize and authorize the authenticity of that idea to be essential to the organization (van Leeuwen, 2007; Zilber, 2002).

Building a culture of evaluation means constructing value chains that inform and inspire institutional change decision making (Sroufe, 2017). These decisions include how to begin diffusing value in evaluation; how to sustain that momentum; and how to navigate opposition or barriers (Zucker, 1977). Informants discussed many methods they found to be effective in building a culture of evaluation. They emphasized that a culture of evaluation is predicated on a culture of trust (Thames & Webster, 2009). Of the informants who have done the heavy lifting of building a healthy culture of evaluation at their respective agencies, they stated that it took intentionality and persistence. They sought opportunities to first understand the baseline culture and capacity of evaluation at their agency and—whether positive or negative—investigated the origins or influences of those beliefs by building meaningful relationships with program and executive staff.

Understanding the organizational landscape allowed them to know where there was support, apprehension, or opposition. They worked to empower allies to advocate for

evaluation and facilitated discussions with those who were apprehensive or opposed to evaluation. The facilitated discussions often revealed that stakeholders had a misunderstanding of or lingering distrust in evaluation because of previous experiences—personally or historically. These conversations were instrumental to helping change agents navigate and ease tension brewing within the agency (Espedal, 2017; Fierro, 2016). Although most of the current literature on facilitation in evaluation is related to facilitating engagement with stakeholders when conducting evaluations, many of the same principles can be translated to facilitate the diffusion of value in evaluation.

Evaluation should not be instructive or declarative, but interactive and constructive. It is also important to establish trust with program staff and understand their needs for organizational learning (Stevahn & King, 2016). During this process, informants found that program staff were seeking to better understand how to ask evaluation questions or utilize evidence that is generated by evaluations. This created an opportunity for evaluators to bridge a gap between the evidence and evidence users by more thoughtfully presenting and justifying evidence for use (Pankaj & Emery, 2016). Informants were intentional about democratizing evidence to ensure evaluations methods could be generalizable to any context within the organization. Ultimately, informants were often in positions of authority and had positive reputations within their agencies. Thus, trust in the individual led to trust in their advocacy. The next section elaborates on the role of leaders and advocates in diffusing value in evaluation.

**Fostering leaders and advocates.** Change agents are the lynchpins for institutional change. They are often mid-level staff who not only emphasize the



importance of a culture of evaluation in organizations, but also influence behavior toward institutionalization (Espedal, 2017; Warrick, 2017). Effective agents of change can hold any type of role in an agency. The common characteristics making them successful in influencing change are their reputation in the agency and persistence in diffusing value in evaluation (Zucker, 1977). This process is methodical and strategic. Informants shared they were not only respected as skilled evaluators but were equally attentive to organizational theory. They studied the processes and outcomes of similar efforts to create new markets for statistics, economics, performance measurement, data science, auditing, and other analytic functions in the Federal government and, specifically, the trajectory of those fields in their agencies (DiMaggio & Powell, 1983; Haveman, 1993).

The mapping method used in this study provided a systematic approach to obtain the same (and more) information across the entire Federal government. However, the way in which informants learned about the culture and capacity at their agencies was rarely through formal ethnography. They primarily learned through informal conversations to develop institutional knowledge and the degree to which this learning phase was necessary depended on the length of their tenure and prior relationship with the agency (Higgins et al., 2012). Those who had established reputations as external consultants, who were contracted by agencies to provide rigorous evaluations, had smoother transitions into agencies and were, therefore, better positioned to influence the culture and capacity for evaluation. However, they observed that what was created could just as easily be undone if institutional change is not consistent and sustained.

Most informants reported that high turnover and transitioning leadership caused

the biggest disruption in the process of institutionalization (Green et al., 2009; Oliver, 1992). The consequences of turnover were not just a change in momentum but were often shifts in political agenda, which caused rhetoric to take a completely different direction. Even if the original cause is respectable—it was not that new leaders had little knowledge of or admiration for it—new leaders often insisted on instituting new social logics that were distinct from their predecessors (Enfield, 2000; Higgins et al., 2012; Nigam & Ocasio, 2010; Oliver, 1992).

Despite informants all being evaluation experts, they reported that some of the most influential actors in the process of institutionalizing evaluation at their agency were Secretaries, Deputy/Assistant Secretaries, and Directors who were not evaluators but had the power and autonomy to be catalysts for institutional change (Espedal, 2017). In fact, it was often more beneficial for the loudest advocates to be leaders who are not evaluators. Evaluators preaching about the value of evaluation were seen as professionals who are passionate about their craft, but non-evaluation leaders who spread the gospel of evaluation represented executive buy-in and value that have been externally assessed (Zilber, 2002). However, reputable and passionate evaluators are still encouraged to self-advocate and foster relationships with leaders who have the authority to mobilize widespread change. The next section discusses a need to train and hire qualified evaluators that can be instrumental in this process.

**Equipping evaluators and evaluation units.** Establishing evaluation as a distinct part of an agency requires hiring individuals with specialized skills in evaluation to perform unique studies. Hiring a professional from another discipline and simply tasking

them with the responsibility to do evaluations diminishes the specialization of evaluation because it creates the perception that anyone can do evaluations without formal training. For example, trade skills lose value when people begin to teach themselves to perform tasks on their own instead of hiring professionals. The ease of learning a trade without formal training threatens the establishment of the professional market. Despite the growing market, there is no job series for evaluation in the Federal government and most who perform evaluative activities hold positions that do not require formal education in evaluation (Maynard et al., 2016). Many must build evaluation skills *post hoc* according to job demands but may not develop evaluative thinking (Milstein et al, 2002; Parsons et al, 2016). Thinking evaluatively goes beyond the mere act of conducting evaluation studies. Evaluative thinking is a form of critical thinking that distinguishes evaluation as a specialized discipline requiring distinct knowledge about evaluation theories and methods (Abbott, 1988; Freidson, 2001; Schwandt, 2018b; Vo & Archibald, 2018; Vo et al., 2018; Worthen, 1994). Therefore, training and hiring evaluators should be intentionally and thoughtfully done to preserve evaluation as a distinctive practice.

The control, growth, and sustainability of evaluation as a profession requires the training of new practitioners in expert knowledge which validate skills that are above and beyond the capabilities of lay people (Abbott, 1988; Bogenschneider & Corbett, 2010; Freidson, 1989; Mertens, 1994; Muzio et al., 2013; Stufflebeam, 2001). Researchers articulate a need to strengthen the theoretical and methodological training of evaluators (Christie & Masyn, 2010; Jacob & Boisvert, 2010; Mark et al., 2009; Monroe et al., 2005). Currently, there are only a small number of formal evaluation degree programs in

the US and some related disciplines may only offer a single (elective) course on program evaluation. A survey of university faculty teaching evaluation courses revealed there was high variability in content and mode of delivery in preparing evaluation professionals (Davies & MacKay, 2014; LaVelle & Donaldson, 2015). This body of literature and current study continue to call for professionalizing evaluation.

Since the late 1980s, GAO has offered training to support the work of those conducting Federal auditing and/or evaluation (Kingsbury & Hedrick, 1994). However, designing and delivering this type of training has been a challenge because of the variability in attendee background and needs (Kingsbury & Hedrick, 1994). It is difficult to develop a sense of purpose, professional development, and job assessments for ill-defined occupations. Approximately 25 years ago, GAO leadership reported concerns with a lack of shared terminology, methodologies, and guidance in federal evaluation (Kingsbury & Hedrick, 1994). What is still lacking in distinguishing evaluation as a distinct profession is a shared language to define clear qualifications of skilled evaluators. In fact, the field only continues to generate more evaluation jargon making it increasingly difficult to institutionalize a practice that is so dynamic and unwieldy (Patton, 2000). Walk and Handy (2018) stated that employees will resort to job crafting when they feel a misfit to their assigned duties but can regain a sense of control if they understand their purpose and value to the organization.

The U.S. Office of Personnel Management (OPM) develops standards for recognized professional job families in the public sector by justifying unique distinctions and classifying tiers of those positions. These standards provided by OPM then factor

into decisions on situating a series of jobs within that occupation family along the General Schedule of compensation. Lacking the appropriate organizational infrastructure to establish evaluation as a profession devalues and mystifies the work, which results in evaluation becoming a subsidiary to related jobs (i.e., mathematics and statistics, education research, analysis, etc.) that are recognized job families in the public sector. Subsuming evaluation as merely a task under other roles curtails the prevalence and rigor of an activity that holds the power to make significant decisions about social programs and policies (Alkin, 2013; Lemire, Fierro, et al., 2018; Maynard et al., 2016). Although there needs to be more recognition of evaluation as a distinct discipline, it is not encouraged for any advocate to create the illusion that evaluation is the sole proprietor of evidence generation. Instead, it is important for stakeholders to acknowledge the unique value rigorous evaluations have in contributing to a robust evidence infrastructure.

**Establishing an evidence infrastructure.** In 2018, the President’s Management Agenda included the Federal Data Strategy (OMB M-19-01) with the tagline, “Leveraging Data as a Strategic Asset” prominently branded on the web page (Federal Data Strategy, n.d.). The goal of the Federal Data Strategy is to provide guidance on how the government should generate, manage, access, and use Federal data for decision-making, and evaluations are acknowledged among other evidence-generating activities to inform evidence-based policymaking—the process by which a critical synthesis of available evidence is done to ethically inform organizational policy decisions (Bond, 2018; Greenhalgh & Russell, 2009; Haskins, 2018; McKay et al., 2015; Pawson et al.,

2011). However, this infrastructure is comprised of multiple disciplines, which requires mutual coordination and collaboration.

The *Analytical Perspectives* has offered considerable guidance on the process of building a robust evidence infrastructure and the value of centralized evaluation offices in this process. Early adopters—before the Evidence Act was enacted in 2019—elevated evaluation by appointing Chief Evaluation Officers (CEOs) as senior executives who were responsible for coordinating all evaluation activities in the agency. Although all agencies with Chief Financial Officers (CFOs) are now mandated to appoint CEOs, the motivation for this change is different from the agencies who appointed CEOs on their own volition. Informants who have witnessed both processes noted simply appointing CEOs for compliance would make institutionalization a futile (and painful) process.

The *Analytical Perspectives* suggest agencies should cross-collaborate with each other about shared needs and methods. This study confirms that the evidence infrastructure and practices within the Federal government are affected by events and other entities outside of the government—even internationally. The evidence infrastructure, therefore, is not limited to internal personnel in the government. External vendors who provide evaluation services for Federal agencies can form a unique bridge between government and non-government layers of the evaluation ecosystem, but they may lack the authority to operationalize institutional change (Gopalakrishnan & Zhang, 2017). By also having direct relationships with other clients in various settings, consultants have a unique ability to offer innovative techniques; challenge professional advancement by sharing practices and methods from other contexts; and broker

knowledge to refine the overall evidence infrastructure across the Federal government (Glennerster, 2012; Olejniczak, 2017; Pawson et al., 2011).

**Limitations and future research.** This study has a few limitations and opportunities to note. The timeline of significant events was generated based only on the information provided in the interviews and from key events mentioned in the *Analytical Perspectives* evaluation chapters. However, this was delimited by providing interview participants with questions a minimum of one week ahead of the interviews which allowed them to prepare comprehensive responses. Additionally, nearly all participants arrived prepared with documents and notes, which decreases the likelihood that informants would have failed to mention details that were important to note. Several participants also provided additional resources that validated their perspectives or verified dates, which were used to corroborate data. Furthermore, the purpose of this study was not to create an exhaustive list of any and all factors that have influenced the diffusing value in evaluation in the U.S. Federal Government, as this would be nearly impossible to fully uncover.

It is also important to reiterate that the findings of this study are not (meant to be) generalizable to every agency in the government, nor does it describe the process at any one agency. This study attempted to include a diverse sample of interview participants and included all available evaluation chapters of the *Analytical Perspectives* to observe common factors of change across the Federal government. The pace and pattern of institutional change occurred at different magnitudes, rates, and frequencies at different layers of the ecosystem and with varying degrees of success. Despite evidence that

evaluation, at-large, is approaching institutionalization in the U.S. Federal Government, there are instances where evaluation was deinstitutionalized, and the impact of those decisions should not be discounted. The disbandment of evaluation units like the CPRS in Great Britain and the evaluation office at EPA are only two examples mentioned by informants of many instances where evaluation was met with strong opposition. Informants briefly discussed the profound impact that these and other decisions had on the culture of evaluation and evidence utilization in government—some even decades after disbandment of these offices, but conspicuously related. These events should be further investigated and traced to related corollaries that have already occurred and monitored for repercussions that may have unrealized or developing consequences.

The inclusion of only evaluation experts may also be viewed as a limitation. Although the intent was to interview individuals whose work was affected by changes in the field of evaluation, future research should be intentional about including the voices of non-evaluators—including naysaying decision makers who (un)knowingly opposed evaluation. Furthermore, research is needed on decision making, in general, in organizations for or against institutionalizing new markets. Although the findings of this study simplified the process to four types of influences with corresponding manifestations, there are an array of decisions being made at each level. The system may be more fragile and agile than what was conveyed through the *Analytical Perspectives* and interviews. Future research should investigate the decision-making processes of the actors involved in the complex process of institutionalization—particularly, the chain of



influence between factors (e.g., an event that creates the political will to develop new legislation which inspires an advocate to initiate organizational change).

There was general consensus of urgency to professionalize the field of evaluation—not just within the Federal government, but also broadly as a professional practice. Informants expressed there is an untapped potential for evaluators that feel their identity in the Federal government is not clearly defined. This remains a major barrier to institutionalizing evaluation. Although there have been concerted efforts to work toward professionalizing evaluation, informants reported there are varying degrees of awareness or recognition of those efforts inside the Federal government. This study offers some potential effective (and ineffective) levers in institutionalizing evaluation in similar contexts. Future research should repeat the mapping method with individual agencies to better understand the unique milieus that lead to micro-level shifts and their relationship to macro-level changes across the system.

## **Conclusion**

The process for institutionalization begins at the micro-levels and influences structural changes at the macro-levels (Li, 2017). One of the tools The U.S. Federal Government has used is the *Analytical Perspectives* to codify a system-wide values chain that prioritizes evaluation as an essential function for a more efficient and effective government (van Leeuwen, 2007). Interviews with experts in evaluation practice informed the mapping of influential factors on institutionalizing evaluation. Many positive and negative factors were identified that have had cognitive, motivational, behavioral, or structural influences on institutionalizing evaluation, but the pace and

pattern of institutional change occurs at different magnitudes, rates, and frequencies within the U.S. Federal Government.

Influences on institutional change manifest in different ways, ranging from a belief system of value in evaluation to building a robust infrastructure to do and use more evaluation. Building an organizational culture of evaluation is at the core of diffusing value in evaluation and the levels of influence are progressive, which suggest earlier stages must be established to sustain higher levels of influence. In a complex organization like the Federal government, this means incremental changes within individual agencies can catalyze the process toward full institutionalization of evaluation across the entire government. Therefore, cultivating a healthy culture of evaluation, fostering leaders and advocates, equipping evaluators and evaluation units, and establishing a robust evidence infrastructure are essential to institutionalizing evaluation in the U.S. Federal Government (Armenakis et al., 1993; Shah et al., 2017).

## References

- Abbott, A. (1988). *The system of professions: An essay on the division of expert labor*.  
The University of Chicago Press.
- Alkin, M. C. (Ed.). (2013). *Evaluation roots: A wider perspective of theorists' views and influences* (2nd ed.). SAGE Publishers, Inc.
- Altschuld, J. W. (1999). The certification of evaluators: Highlights from a report submitted to the Board of Directors of the American Evaluation Association. *American Journal of Evaluation*, 20(3), 481-493.
- American Evaluation Association. (2003). *Scientifically Based Evaluation Methods*.  
<https://www.eval.org/p/cm/ld/fid=95>
- American Evaluation Association. (2004). *AEA Guiding Principles for Evaluators*.  
Washington, DC Retrieved from <http://www.https://www.eval.org/p/cm/ld/fid=51>.
- American Evaluation Association. (2019). *An evaluation roadmap for a more effective government*.
- American Evaluation Association. (n.d.). *About AEA*. <https://www.eval.org/p/cm/ld/fid=4>
- Armenakis, A. A., Harris, S. G., & Mossholder, K. W. (1993). Creating readiness for organizational change. *Human Relations*, 46, 681-703.

- Bailey, M. J., & Duquette, N. J. (2014). How Johnson fought the War on Poverty: The economics and politics of funding at the Office of Economic Opportunity. *Journal of Economic History*, 74(2), 351-388.
- Barber, J. P., & Walczak, K. K. (2009). *Conscience and critic: Peer debriefing strategies in grounded theory research*. American Educational Research Association, San Diego, California.
- Bickman, L. (1999). AEA, bold or timid? *American Journal of Evaluation*, 20(3), 519-520.
- Blackstone, T., & Plowden, W. (1988). *Inside the think tank: Advising the Cabinet 1971-1983*. William Heinemann Ltd.
- Blanchard, O. J. (1987). Reaganomics. *Economic Policy*, 2(5), 15-56.
- Bogensneider, K., & Corbett, T. J. (2010). *Evidence-based policymaking: Insights from policy-minded researchers and research-minded policymakers*. Taylor and Francis Group, LLC.
- Bolman, L. G., & Deal, T. E. (2017). *Reframing organizations: Artistry, choice, and leadership* (6th ed.). John Wiley & Sons, Inc.
- Bond, G. R. (2018). Evidence-based policy strategies: A typology. *Clinical Psychology Science and Practice*, 25, 1-14. <https://doi.org/10.1111/cpsp.12267>

- Brinkmann, S. (2018). The interview. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (5th ed., pp. 576-599). SAGE Publications, Inc.
- Burawoy, M. (2009). The extended case method: Race and class in postcolonial Africa. In *The extended case method: Four countries, four decades, four great transformations, and one theoretical tradition* (pp. 19-72). University of California Press.
- Campbell, J. L., Quincy, C., Osserman, J., & Pedersen, O. K. (2013). Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement. *Sociological Methods & Research*, 42(3), 294-320.  
<https://doi.org/10.1177/0049124113500475>
- Caracelli, V. J. (2000). Evaluation use at the threshold of the twenty-first century. In V. J. Caracelli & H. Preskill (Eds.), *The expanding scope of evaluation use. New Directions for Evaluation*, 88, 99-111. <https://doi.org/10.1002/ev.1194>
- Carspecken, P. F. (1996). *Critical Ethnography in Educational Research*. Routledge.
- Charmaz, K., Thornberg, R., & Keane, E. (2018). Evolving grounded theory and social justice inquiry. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (5th ed., pp. 411-443). SAGE Publications, Inc.
- Chelimsky, E. (2015). A strategy for improving the use of evaluation findings in policy. In C. A. Christie & A. T. Vo (Eds.), *Evaluation use and decision making in society: A tribute to Marvin C. Alkin* (pp. 73-90). Information Age Publishing Inc.

- Christie, C. A., & Masyn, K. E. (2010). Latent profiles of evaluators' self-reported practices. *The Canadian Journal of Program Evaluation*, 23, 225-254.
- Colyvas, J. A., & Jonsson, S. (2011). Ubiquity and legitimacy: Disentangling diffusion and institutionalization. *Sociological Theory*, 29(1), 27-53.
- Copeland, C. W., & Simpson, M. (2004). *The Information Quality Act: OMB's guidance and initial implementation*. The Library of the Congress.  
<https://fas.org/sgp/crs/RL32532.pdf>
- Davies, R., & MacKay, K. (2014). Evaluator training: Content and topic valuation in university evaluation courses. *American Journal of Evaluation*, 35, 419-429.  
<https://doi.org/10.1177/1098214013520066>
- Denzin, N. K. (1978). *Sociological methods: A sourcebook*. McGraw-Hill.
- DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48, 147-160.
- Enfield, N. J. (2000). The theory of cultural logic: How individuals combine social intelligence with semiotics to create and maintain cultural meaning. *Cultural Dynamics*, 12(1), 35-64. <https://doi.org/10.1177/092137400001200102>

Espedal, B. (2017). Understanding how balancing autonomy and power might occur in leading organizational change. *European Management Journal*, 35, 155-163.

<https://doi.org/10.1016/j.emj.2016.08.005>

Federal Data Strategy. (n.d.). GSA Technology Transformation Services.

<https://strategy.data.gov>

Fierro, R. S. (2016). Enhancing facilitation skills: Dancing with dynamic tensions. In R.

S. Fierro, A. Schwartz, & D. H. Smart (Eds.), *Evaluation and Facilitation. New*

*Directions for Evaluation*, 149, 31-42. <https://doi.org/10.1002/ev.20177>

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2011). Origins and current trends in modern program evaluation. In *Program evaluation: Alternative approaches and practical guidelines* (4th ed., pp. 38-63). Pearson Education.

Flick, U. (2018). Triangulation. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (5th ed., pp. 444-461). SAGE Publications, Inc.

Freidson, E. (1986). *Professional powers: A study of the institutionalization of formal knowledge*. University of Chicago Press.

Freidson, E. (1989). Theory and the professions. *Indiana Law Journal*, 64, 423-432.

Freidson, E. (2001). *Professionalism: The third logic*. University of Chicago Press.

Freudenburg, W. R., & Muselli, V. (2013). Reexamining climate change debates:

Scientific disagreement of scientific certainty argumentation methods (SCAMs)?

*American Behavioral Scientist*, 57(6), 777-795.

<https://doi.org/10.1177/0002764212458274>

Gee, J. P. (2014). *An introduction to discourse analysis: Theory and method*. Routledge.

Glennerster, R. (2012). The power of evidence: Improving the effectiveness of government by investing in more rigorous evaluation. *National Institute Economic Review*, 219, R4-14.

Gopalakrishnan, S., & Zhang, H. (2017). Client dependence and vendor innovation: The moderating role of organizational culture. *Industrial Marketing Management*, 66, 80-89. <https://doi.org/10.1016/j.indmarman.2017.07.012>

Green, S. E., Li, Y., & Nohria, N. (2009). Suspended in self-spun webs of significance: A rhetorical model of institutionalization and institutionally embedded agency. *The Academy of Management Journal*, 52(1).

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 11(3), 255-274. <https://doi.org/10.3102/01623737011003255>

Greenhalgh, T., & Russell, J. (2009). Evidence-based policymaking: A critique. *Perspectives in Biology and Medicine*, 52(2), 304-318.



- Hansen, H. F., & Rieper, O. (2009). The Evidence Movement: The development and consequences of methodologies in review practices. *Evaluation, 15*(2), 141-163.  
<https://doi.org/10.1177/1356389008101968>
- Haskins, R. (2018). Evidence-based policy: The movement, the goals, the issues, the promise. *Annals of the American Academy of Political and Social Sciences, 678*, 8-37. <https://doi.org/10.1177/0002716218770642>
- Haveman, H. A. (1993). Follow the leader: Mimetic isomorphism and entry into new markets. *Administrative Science Quarterly, 38*(4), 593-627.
- Higgins, M. C., Weiner, J., & Young, L. (2012). Implementation teams: A new lever for organizational change. *Journal of Organizational Behavior, 33*, 366-388.  
<https://doi.org/10.1002/job.1773>
- Institute of Education Science. (2003). *Identifying and implementing educational practices supported by rigorous evidence: A user friendly guide*.  
[https://ies.ed.gov/ncee/pubs/evidence\\_based/randomized.asp](https://ies.ed.gov/ncee/pubs/evidence_based/randomized.asp)
- Jacob, S., & Boisvert, Y. (2010). To be or not to be a profession: Pros, cons and challenges for evaluation. *Evaluation, 16*, 349-369.  
<https://doi.org/10.1177/1356389010380001>
- Kinarsky, A. R. (2018). The evaluation landscape: U.S. foundation spending on evaluation. In S. B. Nielsen, S. Lemire, & C. A. Christie (Eds.), *The evaluation*

*marketplace: Exploring the evaluation industry. New Directions for Evaluation, 160*, 81-96. <https://doi.org/10.1002/ev.20342>

Kingsbury, N., & Hedrick, T. E. (1994). Evaluator trainign in a government setting. In J. W. Altschuld & M. Engle (Eds.), *The preparation of professional evaluators: Issues, perspectives, and programs. New Directions for Evaluation, 62*, 61-70.

LaVelle, J. M., & Donaldson, S. I. (2015). The state of preparing evaluators. In J. W. Altschuld & M. Engle (Eds.), *Accreditation, certification, and credentialing: Relevant concerns for U.S. evaluators. New Directions for Evaluators, 145*, 39-52. <https://doi.org/10.1002/ev.20110>

Lemire, S., Fierro, L. A., Kinarsky, A. R., Fujita-Conrads, E., & Christie, C. A. (2018). The U.S. federal evaluation market. In S. B. Nielsen, S. Lemire, & C. A. Christie (Eds.), *The evaluation marketplace: Exploring the evaluation industry. New Directions for Evaluation, 160*, 63-80. <https://doi.org/10.1002/ev.20343>

Lemire, S., Nielsen, S. B., & Christie, C. A. (2018). Toward understanding the evaluation market and its industry—Advancing a research agenda. In S. B. Nielsen, S. Lemire, & C. A. Christie (Eds.), *The evaluation marketplace: Exploring the evaluation industry. New Directions for Evaluation, 160*, 145-163. <https://doi.org/10.1002/ev.20339>

Li, Y. (2017). A semiotic theory of institutionalization. *Academy of Management Review, 42*(3), 520-547. <https://doi.org/10.5465/amr.2014.0274>

Lincoln, Y. S., Lynham, S. A., & Guba, E. G. (2006). Paradigmatic controversies, contradictions, and emerging confluences, revisited. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE Handbook of Qualitative Research* (4th ed., pp. 97-128). SAGE Publications, Inc.

Madison, A. M. (2000). Language in defining social problems and in evaluating social programs. In R. K. Hopson (Ed.), *How and why language matters in evaluation. New Directions for Evaluation*, 86, 17-28. <https://doi.org/10.1002/ev.1169>

Manson, H. (2016). Systematic reviews are not enough: policymakers need a greater variety of synthesized evidence. *Journal of Clinical Epidemiology*, 73, 11-14. <https://doi.org/10.1016/j.jclinepi.2015.08.032>

Mark, M. M., Cooksy, L. J., & Trochim, W. M. K. (2009). Evaluation policy: An introduction and overview. In W. M. K. Trochim, M. M. Mark, & L. J. Cooksy (Eds.), *Evaluation policy and evaluation practice. New Directions for Evaluation*, 123, 3-11. <https://doi.org/10.1002/ev.302>

Maynard, R., Goldstein, N., & Smith Nightingale, D. (2016). Program and policy evaluations in practice: Highlights from the federal perspective. In L. R. Peck (Ed.), *Social experiments in practice: The what, why, when, where, and how of experimental design & analysis. New Directions for Evaluation*, 152, 109-135. <https://doi.org/10.1002/ev.20209>

- McKay, S., Murray, M., MacIntyre, S., & Kashyap, A. (2015). Evidence-based policymaking and the public interest: Lessons in legitimacy. *Town Planning Review*, 86(2), 133-154.
- Mertens, D. M. (1994). Training evaluators: Unique skills and knowledge. In J. W. Altschuld & M. Engle (Eds.), *The preparation of professional evaluators: Issues, perspectives, and programs. New Directions for Evaluation*, 62, 17-27.
- Meyer, R. E., Jancsary, D., Höllerer, M. A., & Boxenbaum, E. (2018). The role of verbal and visual text in the process of institutionalization. *Academy of Management Review*, 43(3), 329-418. <https://doi.org/10.5465/amr.2014.0301>
- Monroe, M. C., Fleming, L., Bowman, R. A., Zimmer, J. F., Marcinkowski, T., Washburn, J., & Mitchell, N. J. (2005). Evaluators as educators: Articulating program theory and building evaluation capacity. In E. Norland & C. Somers (Eds.), *Evaluating nonformal education programs and settings. New Directions for Evaluation*, 108, 57-71. <https://doi.org/10.1002/ev.171>
- Morse, J. (2018). Reframing rigor in qualitative inquiry. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (5th ed., pp. 796-817). SAGE Publications, Inc.
- Moynihan, D. P. (1965). *The negro family: The case for national action*.

- Muzio, D., Brock, D. M., & Suddaby, R. (2013). Professions and institutional change: Towards an institutionalist sociology of the professions. *Journal of Management Studies*, 50, 699-721. <https://doi.org/10.1111/joms.12030>
- National Academies of Sciences. (2017). *Principles and practices for federal program evaluation: Proceedings of a workshop*. <https://www.nap.edu/catalog/24716/principles-and-practices-for-federal-program-evaluation-proceedings-of-a>
- Neuendorf, K. A. (2017). *The content analysis guidebook* (2nd ed.). SAGE Publications.
- Nigam, A., & Ocasio, W. (2010). Event attention, environmental sensemaking, and change in institutional logics: An inductive analysis of the effects of public attention to Clinton's health care reform initiative. *Organization Science*, 21(4), 823-841. <https://doi.org/10.1287/orsc.1090.0490>
- Nolton, E. C., Varier, D., Powell, M. G., Caracelli, V. J., & Hopson, R. K. (2020a). *The diffusion of value in evaluation by the U.S. Federal Government through the Analytical Perspectives of the U.S. President's Budget*
- Nolton, E. C., Varier, D., Powell, M. G., Caracelli, V. J., & Hopson, R. K. (2020b). *Introducing a conceptual framework and method of mapping influences on the process of institutionalizing evaluation in the U.S. Federal Government*.

Olejniczak, K. (2017). The game of knowledge brokering: A new method for increasing evaluation use. *American Journal of Evaluation*, 38(4), 554-576.

<https://doi.org/10.1177/1098214017716326>

Oliver, C. (1992). The antecedents of deinstitutionalization. *Organization Studies*, 13(4), 563-588.

Pankaj, V., & Emery, A. K. (2016). Data placemats: A facilitative technique designed to enhance stakeholder understanding of data. In R. S. Fierro, A. Schwartz, & D. H. Smart (Eds.), *Evaluation and Facilitation. New Directions for Evaluation*, 149, 81-93. <https://doi.org/10.1002/ev.20181>

Patton, M. Q. (1999). Enhancing the quality and credibility of qualitative analysis. *Health Services Research*, 34(5), 1189-1208.

Patton, M. Q. (2000). Overview: Language matters. In R. K. Hopson (Ed.), *How and why language matters in evaluation. New Directions for Evaluation*, 86, 5-16.

<https://doi.org/10.1002/ev.1168>

Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice*. SAGE Publications, Inc.

Pawson, R. (2006). *Evidence-based policy: A realist perspective*. SAGE Publications, Inc.

- Pawson, R., Wong, G., & Owen, L. (2011). Known knowns, known unknowns, unknown unknowns: The predicament of evidence-based policy. *American Journal of Evaluation*, 32(4), 518-546. <https://doi.org/10.1177/1098214011403831>
- Peck, J., & Theodore, N. (2015). *Fast policy: Experimental statecraft at the thresholds of neoliberalism*. University of Minnesota Press.
- Peräkylä, A., & Ruusuvuori, J. (2018). Analyzing talk and text. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (5th ed., pp. 669-691). SAGE Publications, Inc.
- Riggin, L. J. C., Shipman, S. L., & York, R. L. (1995). Using a set-aside to encourage the evaluation of public health service programs. *Evaluation & The Health Professions*, 18(1), 3-12.
- Schwandt, T. A. (2018). Evaluative thinking as a collaborative social practice: The case of boundary judgment making. In A. T. Vo & T. Archibald (Eds.), *Evaluative thinking. New Directions for Evaluation*, 158, 125-137. <https://doi.org/10.1008/ev.20318>
- Scriven, M. (2008). The concept of a transdiscipline: And of evaluation as a transdiscipline. *Journal of MultiDisciplinary Evaluation*, 5(10), 65-66.
- Shah, N., Irani, Z., & Sharif, A. M. (2017). Big data in an HR context: Exploring organizational change readiness, employee attitudes and behaviors. *Journal of Business Research*, 70, 366-378. <https://doi.org/10.1016/j.busres.2016.08.010>

- Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information, 22*, 63-75.
- Shore, C., & Wright, S. (2011). Introduction. Conceptualising policy: Technologies of governance and the politics of visibility. In S. Shore, S. Wright, & D. Però (Eds.), *Policy worlds: Anthropology and the analysis of contemporary power* (pp. 1-25). Bergahn Books.
- Spall, S. (1998). Peer debriefing in qualitative research: Emerging operational models. *Qualitative Inquiry, 4*(2), 280-292.
- Sroufe, R. (2017). Integration and organizational change towards sustainability. *Journal of Cleaner Production, 162*, 315-329.  
<https://doi.org/10.1016/j.clepro.2017.05.180>
- Stevahn, L., & King, J. A. (2016). Facilitating interactive evaluation practice: Engaging stakeholders constructively. In R. S. Fierro, A. Schwartz, & D. H. Smart (Eds.), *Evaluation and Facilitation. New Directions for Evaluation, 149*, 67-80.  
<https://doi.org/10.1002/ev.20180>
- Stone, D. (2008). Global public policy, transnational policy communities, and their networks. *Policy Studies Journal, 36*(1), 19-38. <https://doi.org/10.1111/j.1541-0072-2007.00251.x>
- Strang, D., & Meyer, J. W. (1993). Institutional conditions for diffusion. *Theory and Society, 22*, 487-511.



- Stufflebeam, D. L. (2001). Interdisciplinary PhD programming in evaluation. *American Journal of Evaluation*, 22, 445-455.  
<https://doi.org/10.1177/109821400102200323>
- Thames, R. C., & Webster, D. W. (2009). *Chasing change: Building organizational capacity in a turbulent environment*. John Wiley & Sons, Inc.
- U.S. General Accounting Office. (2003). *An evaluation culture and collaborative partnerships help build agency capacity*. (GAO-03-454). Washington, DC: U.S. General Accounting Office.
- U.S. General Accounting Office. (2004). *OMB's performance rating tool presents opportunities and challenges for evaluating program performance*. (GAO-04-550T). Washington, DC.
- U.S. Office of Management and Budget. (n.d.). *Assessing Program Performance*.  
<https://georgewbush-whitehouse.archives.gov/omb/performance/index.html>
- van Leeuwen, T. (2007). Legitimation in discourse and communication. *Discourse & Communication*, 1(1), 91-112. <https://doi.org/10.1177/1750481307071986>
- Vo, A. T., & Archibald, T. (2018). New directions for evaluative thinking. In A. T. Vo & T. Archibald (Eds.), *Evaluative thinking. New Directions for Evaluation*, 158, 139-147. <https://doi.org/10.1002/ev.20317>

- Vo, A. T., Schreiber, J. S., & Martin, A. (2018). Toward a conceptual understanding of evaluative thinking. In A. T. Vo & T. Archibald (Eds.), *Evaluative thinking. New Directions for Evaluation*, 158, 29-47. <https://doi.org/10.1002/ev.20324>
- Walk, M., & Handy, F. (2018). Job crafting as reaction to organizational change. *The Journal of Applied Behavioral Science*, 54, 349-370.  
<https://doi.org/10.1177/0021886318777227>
- Warrick, D. D. (2017). What leaders need to know about organizational culture. *Business Horizons*, 60, 395-404. <https://doi.org/10.1016/j.bushor.2017.01.011>
- Wedel, J. R. (2019). *The mapping method: A guide to charting corruption and influence processes* [Working paper forthcoming]. Global Integrity.
- Worthen, B. R. (1994). Is evaluation a mature profession that warrants the preparation of evaluation professionals? In J. W. Altschuld & M. Engle (Eds.), *The preparation of professional evaluators: Issues, perspectives, and programs. New Directions for Evaluation*, 62, 3-15.
- Worthen, B. R. (1999). Critical challenges confronting certification of evaluators. *American Journal of Evaluation*, 20(3), 533-555.
- Zilber, T. B. (2002). Institutionalization as an interplay between actions, meanings, and actors: The case of a rape crisis center in Isreal. *Academy of Management Journal*, 45(1), 234-254.

Zucker, L. G. (1977). The role of institutionalization in cultural persistence. *American Sociological Association*, 42(5), 726-743.

## Appendix A

Tables of Highly Prevalent Words by Year

<b>Word</b>	<b>2011</b>	<b>2012</b>	<b>2013</b>	<b>2014</b>	<b>2015</b>	<b>2016</b>	<b>2017</b>	<b>2018</b>	<b>2019</b>	<b>2020</b>	<b>Total</b>
Evaluation	37	65	48	72	84	32	149	39	82	128	736
Program	38	47	63	84	105	78	75	57	94	90	731
Data	5	5	35	46	56	255	86	36	92	63	679
Evidence	16	20	32	58	84	30	98	59	85	82	564
Agencies	17	26	17	28	40	63	89	33	87	103	503
Use/Using	6	8	21	52	51	74	80	42	72	34	440
Federal	8	14	12	21	20	57	36	11	43	30	252
Policy	4	7	17	23	31	29	33	16	40	34	234
Building	11	9	7	9	11	16	37	28	46	41	215
Improve	9	12	11	24	18	33	19	21	40	26	213
Effective(ness)	5	13	26	24	32	12	21	14	33	23	203
Analysis	2	4	24	29	31	27	27	14	17	17	192
Government	8	13	15	20	20	17	24	17	32	23	189
Administrative	2	3	7	9	16	89	32	15	12	4	189
Research	5	5	10	20	19	29	39	17	20	19	183
Statistical	0	0	2	2	17	83	29	9	26	15	183
Information	4	3	8	13	11	47	41	14	17	24	182
Performance	7	9	9	24	28	17	25	16	20	23	178
Work(s)	8	6	10	27	23	17	36	11	10	11	159
Administration	9	12	16	14	20	16	27	6	17	11	148
Department	27	17	3	16	23	15	23	3	9	10	146
Access	0	1	3	3	6	49	15	1	19	30	127
Budget	1	4	2	10	11	35	17	6	19	19	124
Learning	1	4	4	12	7	3	18	11	20	43	123
Example	2	3	5	16	18	18	17	6	14	24	123
Support	5	6	12	13	17	11	18	10	14	14	120
Rigor(ous)	4	15	10	13	19	5	18	11	11	9	115
Practices	7	12	13	16	17	5	15	6	13	8	112
Education	9	12	8	13	13	22	12	12	6	3	110
Result	4	6	5	14	16	9	20	7	9	18	108
Better	0	3	11	21	20	15	16	11	5	6	108
Measurement	4	5	7	13	14	8	13	9	9	19	101
Outcomes	5	9	11	15	14	9	11	9	9	9	101
Help	2	5	8	11	21	11	15	11	5	11	100
Strategies	2	4	4	11	12	1	11	5	23	25	98
Strong	9	11	5	11	12	14	13	7	10	6	98
Questions	5	4	5	8	11	2	15	12	9	23	94
Implementation	1	2	10	5	10	3	33	8	7	15	94
Effort	6	5	6	11	14	15	13	2	4	10	86
Capacity	11	6	2	5	6	5	23	6	14	7	85
Innovation	4	4	6	15	16	13	10	3	2	4	77
Impact	3	7	5	7	14	9	8	6	4	13	76
Funding	6	8	9	11	11	9	4	2	4	12	76
Security	1	2	2	2	4	26	11	6	8	11	73
Approaches	4	5	7	9	13	3	10	5	3	13	72

Important	3	2	2	5	9	14	15	5	6	10	71
Infrastructure	1	2	4	5	4	12	8	13	7	5	61
Plans	3	3	0	2	2	0	12	4	11	23	60
Agenda	1	0	0	0	0	8	16	6	7	22	60
Quality	1	2	0	2	4	15	16	4	5	9	58
Tools	1	0	1	12	13	3	6	7	10	5	58
Effect	20	5	6	4	5	8	4	3	0	3	58
Available	4	3	4	3	7	15	8	1	7	5	57
Activities	1	3	1	5	3	6	6	1	10	20	56
Culture	2	3	5	15	13	1	4	3	2	7	55
Success	0	1	9	8	13	8	9	0	4	2	54
Privacy	0	0	1	2	3	19	7	7	6	6	51
Central(ized)	2	3	4	4	4	1	17	6	4	2	47
Act	0	2	1	1	2	8	4	1	6	21	46
Transparency	1	1	0	1	1	8	16	3	4	11	46
Bureau	0	0	1	1	1	26	9	4	1	3	46
Standards	0	0	2	8	10	9	9	0	6	0	44
Evidence-Building	0	0	0	0	1	0	1	3	14	21	40
Credible	2	2	2	2	2	0	22	5	1	2	40
Principles	0	0	0	0	0	3	17	2	12	3	37
Findings	1	1	1	1	2	1	14	2	4	6	33
Confidentiality	0	0	0	0	0	18	6	3	4	2	33
Integrity	0	1	1	1	1	2	9	0	2	9	26

## Appendix B

### IRB Approval Letter



Office of Research Integrity and Assurance

Research Hall, 4400 University Drive, MS 6D5, Fairfax, Virginia 22030  
Phone: 703-993-5445; Fax: 703-993-9590

DATE: October 15, 2019

TO: Marvin Powell, PhD  
FROM: George Mason University IRB

Project Title: [1502800-1] Evaluation and Evidence in the US Federal Government

SUBMISSION TYPE: New Project

ACTION: APPROVED  
APPROVAL DATE: October 15, 2019  
REVIEW TYPE: Expedited Review

REVIEW TYPE: Expedited review category #7

Thank you for your submission of New Project materials for this project. The George Mason University IRB has APPROVED your submission. This submission has received Expedited Review based on applicable federal regulations.

Please remember that all research must be conducted as described in the submitted materials.

Please remember that informed consent is a process beginning with a description of the project and insurance of participant understanding followed by a signed consent form unless the IRB has waived the requirement for a signature on the consent form or has waived the requirement for a consent process. Informed consent must continue throughout the project via a dialogue between the researcher and research participant. Federal regulations require that each participant receives a copy of the consent document.

Please note that any revision to previously approved materials must be approved by the IRB prior to initiation. Please use the appropriate revision forms for this procedure.

All UNANTICIPATED PROBLEMS involving risks to subjects or others and SERIOUS and UNEXPECTED adverse events must be reported promptly to the IRB office. Please use the appropriate reporting forms for this procedure. All FDA and sponsor reporting requirements should also be followed (if applicable).

All NON-COMPLIANCE issues or COMPLAINTS regarding this project must be reported promptly to the IRB.

This study does not have an expiration date but you will receive an annual reminder regarding future requirements.

Please note that all research records must be retained for a minimum of five years, or as described in your submission, after the completion of the project.

Please note that department or other approvals may be required to conduct your research in addition to IRB approval.

If you have any questions, please contact Kim Paul at (703) 993-4208 or kpaul4@gmu.edu. Please include your project title and reference number in all correspondence with this committee.

GMU IRB Standard Operating Procedures can be found here: <https://rdia.gmu.edu/topics-of-interest/human-or-animal-subjects/human-subjects/human-subjects-sops/>

This letter has been electronically signed in accordance with all applicable regulations, and a copy is retained within George Mason University IRB's records.



*Evaluation and Evidence in the US Federal Government*

**INFORMED CONSENT FORM**

**RESEARCH PROCEDURES**

This research is being conducted understand how evaluation has been institutionalized and evidence prioritized over time in the U.S. federal government. If you agree to participate, you will be asked to participate in a 45- to 60-minute interview regarding your perspective on this topic.

**RISKS**

There are no foreseeable risks for participating in this research.

**BENEFITS**

There are no benefits to you as a participant other than to further research in understanding the drivers of organizational change regarding evaluation and evidence in the U.S. federal government.

**CONFIDENTIALITY**

The data in this study will be confidential. Participants will be asked to provide their name and consent verbally on the audio recording at the beginning of the interview only to indicate consent. Upon transcription of the interview, the interviewee will be assigned a pseudonym and positions will be changed to be vague descriptors, such as "the Chief Evaluation Officer of a large federal agency" or "a subject matter expert in federal evaluation capacity building."

Data from interviews will be aggregated with other data sources (i.e., artifacts, policies, government reports, etc.) and other interviews. None of the data collected will be associated to your identity in any way. If direct quotes are used in any publications or presentations, they will not be associated to your identity in any way. This study is being conducted as part of a larger study and your perspectives will inform other methods but will not be featured individually in any dissemination of the study. The de-identified data could be used for future research without additional consent from participants.

Audio recordings will be stored on secured computers in secured offices within locked suites and building on the Fairfax campus of George Mason University. Records will be retained for five years, per George Mason University policy, and will be deleted at five years following the completion of the study. Only researcher will have access to these audio recordings.

**PARTICIPATION**

Your participation is voluntary, and you may withdraw from the study at any time and for any reason. If you decide not to participate or if you withdraw from the study, there is no penalty or loss of benefits to which you are otherwise entitled. There are no costs to you or any other party.



IRB: For Official Use Only

Project Number: 1502800-1

Page 1 of 2



In order to be included as an interview participant, you must be selected by the researchers as having unique perspectives and expertise in one or more of the following areas: federal evaluation, the profession of evaluation, evaluation capacity building, the Evidence Act, or evidence-based policymaking.

#### CONTACT

This research is being conducted by Esther Nolton at George Mason University. She may be reached at [echou@gmu.edu](mailto:echou@gmu.edu) for questions or to report a research-related problem. Esther is a Doctoral Candidate and her Dissertation Committee Co-Chairs, Dr. Marvin Powell ([mpowel11@gmu.edu](mailto:mpowel11@gmu.edu); 703-993-1723) or Dr. Rodney Hopson ([hopson@illinois.edu](mailto:hopson@illinois.edu)), may also be contacted regarding any issues or concerns with the study.

This research has been reviewed according to George Mason University procedures governing your participation in this research. You may contact the George Mason University Institutional Review Board office at 703-993-4121 if you have questions or comments regarding your rights as a participant in the research.

#### CONSENT

Participants will be asked to provide verbal consent for participation and to be audio recorded on the audio recording. Participants should state their name, professional title, and that they agree to being audio recorded to indicate their consent to participate.

Dear XXXX,

My name is Esther Nolton and I am a Doctoral Candidate at George Mason University. For my dissertation, I am interested in studying the factors that have influenced the institutionization of evaluation and prioritization of evidence in the U.S. federal government.

**Information about the Study**

This study will include content analyses of key documents and interviews with key stakeholders who are selected based on their expertise. Given your experiences and/or expertise, I would like to request approximately 45-60 minutes of your time to learn your perspectives on this topic.

All interview data will be confidential, which means your identity will not be associated with your interview transcription. Interviews will be used to inform other methods in a larger study. If any direct quotes become important to highlight in dissemination, only pseudonyms will be used.

**Scheduling the Interview**

If you are interested in participating, we have many options to make arrangements. I am located in the DC area and will also be attending Evaluation 2019 in Minneapolis. If you are located in the DC area or have upcoming travel to DC, I would be more than happy to coordinate our schedules and arrange time to meet in-person.

If you are not in the DC area and are attending Evaluation 2019, please let me know if you are available to meet in Minneapolis for the interview and I will make the arrangements.

**Contact Information**

Please do not hesitate to contact me at [echou@gmu.edu](mailto:echou@gmu.edu) if you have any additional questions regarding the study (IRB Number: 1502800-1). Additionally, you may also contact my Dissertation Committee Co-Chairs, Dr. Marvin Powell ([mpowell11@gmu.edu](mailto:mpowell11@gmu.edu)) or Dr. Rodney Hopson ([hopson@illinois.edu](mailto:hopson@illinois.edu)), for more information. Thank you for your time. I look forward to hearing from you!

Thank you,  
Esther Nolton

--

Esther C. Nolton, MEd, LAT, ATC, CSCS  
Doctoral Candidate  
College of Education and Human Development  
George Mason University  
[echou@gmu.edu](mailto:echou@gmu.edu)

## References

- Abbate, T., & Coppolino, R. (2011). Knowledge creation through knowledge brokers: Some anecdotal evidence. *Journal of Management Control*, 22, 359-371.  
<https://doi.org/10.1007/s00187-011-0144-5>
- Abbott, A. (1988). *The system of professions: An essay on the division of expert labor*. The University of Chicago Press.
- Alkin, M. C. (Ed.). (2013). *Evaluation roots: A wider perspective of theorists' views and influences* (2nd ed.). SAGE Publishers, Inc.
- Alkin, M. C., & King, J. A. (2016). The historical development of evaluation use. *American Journal of Evaluation*, 37(4), 568-579.  
<https://doi.org/10.1177/1098214016665164>
- 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. <https://doi.org/10.1177/1098214017717015>
- Altschuld, J. W. (1999). The certification of evaluators: Highlights from a report submitted to the Board of Directors of the American Evaluation Association. *American Journal of Evaluation*, 20(3), 481-493.

American Evaluation Association. (2003). *Scientifically Based Evaluation Methods*.

<https://www.eval.org/p/cm/ld/fid=95>

American Evaluation Association. (2004). *AEA Guiding Principles for Evaluators*.

Washington, DC Retrieved from <http://www>.

<https://www.eval.org/p/cm/ld/fid=51>.

American Evaluation Association. (2018). *AEA Evaluator Competencies*. Washington,

DC Retrieved from <https://www.eval.org/page/competencies>.

American Evaluation Association. (2019). *An evaluation roadmap for a more effective government*.

American Evaluation Association. (n.d.). *About AEA*. <https://www.eval.org/p/cm/ld/fid=4>

Amo, C., & Cousins, J. B. (2007). Going through the process: An examination of the operationalization of process use in empirical research on evaluation. In J. B. Cousins (Ed.), *Process use in theory, research, and practice. New Directions for Evaluation, 116*, 5-26. <https://doi.org/10.1002/ev.240>

Armenakis, A. A., & Harris, S. G. (2002). Crafting a change message to create transformational readiness. *Journal of Organizational Change Management, 15*(2), 169-183. <https://doi.org/10.1108/09534810210423080>

Armenakis, A. A., Harris, S. G., & Mossholder, K. W. (1993). Creating readiness for organizational change. *Human Relations, 46*, 681-703.

- Bailey, M. J., & Duquette, N. J. (2014). How Johnson fought the War on Poverty: The economics and politics of funding at the Office of Economic Opportunity. *Journal of Economic History*, 74(2), 351-388.
- Bansback, N., Li, L. C., Lynd, L., & Bryan, S. (2014). Exploiting order effects to improve the quality of decisions. *Patient Education and Counseling*, 96(2), 197-203. <https://doi.org/10.1016/j.pec.2014.05.021>
- Barber, J. P., & Walczak, K. K. (2009). *Conscience and critic: Peer debriefing strategies in grounded theory research*. American Educational Research Association, San Diego, California.
- Behrens, T. E. J., Muller, T. H., Whittington, J. C. R., Mark, S., Baram, A. B., Stachenfeld, K. L., & Kurth-Nelson, Z. (2018). What is a cognitive map? Organizing knowledge for flexible behavior. *Neuron*, 100, 490-509. <https://doi.org/10.1016/j.neuron.2018.10.002>
- Berger, P., & Luckman, T. (1967). *The social construction of reality—A treatise in the sociology of knowledge*. Anchor Doubleday.
- Bickman, L. (1999). AEA, bold or timid? *American Journal of Evaluation*, 20(3), 519-520.
- Blackstone, T., & Plowden, W. (1988). *Inside the think tank: Advising the Cabinet 1971-1983*. William Heinemann Ltd.

- Blake, S. C., & Ottoson, J. M. (2009). Knowledge utilization: Implications for evaluation. In J. M. Ottoson & P. Hawe (Eds.), *Knowledge utilization, diffusion, implementation, transfer, and translation: Implications for evaluation. New Directions for Evaluation*, 124, 21-34. <https://doi.org/10.1002/ev.311>
- Blanchard, O. J. (1987). Reaganomics. *Economic Policy*, 2(5), 15-56.
- Bogenschneider, K., & Corbett, T. J. (2010). *Evidence-based policymaking: Insights from policy-minded researchers and research-minded policymakers*. Taylor and Francis Group, LLC.
- Bolman, L. G., & Deal, T. E. (2017). *Reframing organizations: Artistry, choice, and leadership* (6th ed.). John Wiley & Sons, Inc.
- Bond, G. R. (2018). Evidence-based policy strategies: A typology. *Clinical Psychology Science and Practice*, 25, 1-14. <https://doi.org/10.1111/cpsp.12267>
- Brennan, L. K., Brownson, R. C., Kelly, C., Ivey, M. K., & Leviton, L. C. (2012). Concept mapping: Priority community strategies to create changes to support active living. *American Journal of Preventative Medicine*, 43(5S4), S337-S350. <https://doi.org/10.1016/j.amepre.2012.07.015>
- Brinkmann, S. (2018). The interview. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (5th ed., pp. 576-599). SAGE Publications, Inc.

- Brown, A. D., Ainsworth, S., & Grant, D. (2012). The rhetoric of institutional change. *Organization Studies*, 33(3), 297-321. <https://doi.org/10.1177/0170840611435598>
- Burawoy, M. (2009). The extended case method: Race and class in postcolonial Africa. In *The extended case method: Four countries, four decades, four great transformations, and one theoretical tradition* (pp. 19-72). University of California Press.
- Cabatoff, K. (2000). Translating evaluation findings into "policy language". In R. K. Hopson (Ed.), *How and why language matters in evaluation. New Directions for Evaluation*, 86, 43-54. <https://doi.org/10.1002/ev.1171>
- Campbell, D., Donald, B., Moore, G., & Frew, D. (2011). Evidence check: Knowledge brokering to commission research reviews for policy. *Evidence & Policy*, 7, 97-107. <https://doi.org/10.1332/174426411X553034>
- Campbell, J. L., Quincy, C., Osserman, J., & Pedersen, O. K. (2013). Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement. *Sociological Methods & Research*, 42(3), 294-320. <https://doi.org/10.1177/0049124113500475>
- Caracelli, V. J. (2000). Evaluation use at the threshold of the twenty-first century. In V. J. Caracelli & H. Preskill (Eds.), *The expanding scope of evaluation use. New Directions for Evaluation*, 88, 99-111. <https://doi.org/10.1002/ev.1194>

- Caracelli, V. J. (2006). Enhancing the policy process through the use of ethnography and other study frameworks: A mixed-method strategy. *Research in the Schools*, 3(1), 84-92.
- Carspecken, P. F. (1996). *Critical Ethnography in Educational Research*. Routledge.
- Centers for Disease Control and Prevention. (n.d.). *Brief 1: Overview of policy evaluation*. <https://www.cdc.gov/injury/pdfs/policy/Brief%201-a.pdf>
- Charmaz, K., Thornberg, R., & Keane, E. (2018). Evolving grounded theory and social justice inquiry. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (5th ed., pp. 411-443). SAGE Publications, Inc.
- Chelimsky, E. (1977). Improving the use of evaluation: An agency perspective. *The Prison Journal*, 57(1), 13-18. <https://doi.org/10.1177/003288557705700103>
- Chelimsky, E. (1987). The politics of program evaluation. *New Directions for Program Evaluation*, 1987(34), 5-21. <https://doi.org/10.1002/ev.1450>
- Chelimsky, E. (2015). A strategy for improving the use of evaluation findings in policy. In C. A. Christie & A. T. Vo (Eds.), *Evaluation use and decision making in society: A tribute to Marvin C. Alkin* (pp. 73-90). Information Age Publishing Inc.
- Christie, C. A. (2015). Setting the state for understanding evaluation use and decision making. In C. A. Christie & A. T. Vo (Eds.), *Evlauation use and decision making*



*in society: A tribute to Marvin C. Alkin* (pp. 1-10). Information Age Publishing, Inc.

Christie, C. A., & Lemire, S. T. (2019). Why evaluation theory should be used to inform evaluation policy. *American Journal of Evaluation*, 40(4), 490-508.

<https://doi.org/10.1177/1098214018824045>

Christie, C. A., & Masyn, K. E. (2010). Latent profiles of evaluators' self-reported practices. *The Canadian Journal of Program Evaluation*, 23, 225-254.

Clarke, J. (2005). Performing for the public: Doubt, desire, and the evaluation of public services. In P. du Gay (Ed.), *Values of Bureaucracy* (pp. 211-232). Oxford University Press.

Colyvas, J. A., & Jonsson, S. (2011). Ubiquity and legitimacy: Disentangling diffusion and institutionalization. *Sociological Theory*, 29(1), 27-53.

Commission on Evidence-Based Policymaking. (2017). *The promise of evidence-based policymaking: Report of the commission on evidence-based policymaking*.

Washington, DC Retrieved from <https://www.cep.gov/report/cep-final-report.pdf>.

Copeland, C. W., & Simpson, M. (2004). *The Information Quality Act: OMB's guidance and initial implementation*. The Library of the Congress.

<https://fas.org/sgp/crs/RL32532.pdf>

Cousins, J. B., Goh, S. C., Elliott, C. J., & Bourgeois, I. (2014). Framing the capacity to do and use evaluation. In J. B. Cousins & I. Bourgeois (Eds.), *Organizational capacity to do and use evaluation. New Directions for Evaluation*, 141, 7-23.

<https://doi.org/10.1002/ev.20076>

D'Eredita, M. A., & Barreto, C. (2006). How does tacit knowledge proliferate? An evidence-based perspective. *Organization Studies*, 27, 1821-1841.

<https://doi.org/10.1177/0170840606067666>

Datta, L. (2003). The evaluation profession and the government. In T. Kellaghan & D. L. Stufflebeam (Eds.), *International Handbook of Educational Evaluation* (pp. 345-360). Dordrecht: Kluwer Academic Publishers.

Davies, R., & MacKay, K. (2014). Evaluator training: Content and topic valuation in university evaluation courses. *American Journal of Evaluation*, 35, 419-429.

<https://doi.org/10.1177/1098214013520066>

Denzin, N. K. (1978). *Sociological methods: A sourcebook*. McGraw-Hill.

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48, 147-160.

Duignan, P. (2003). Mainstreaming evaluation or building evaluation capability? Three key elements. In J. J. Barnette & J. R. Sanders (Eds.), *The mainstreaming of evaluation. New Directions for Evaluation*, 99, 7-21. <https://doi.org/10.1002/ev.90>

- Enfield, N. J. (2000). The theory of cultural logic: How individuals combine social intelligence with semiotics to create and maintain cultural meaning. *Cultural Dynamics*, 12(1), 35-64. <https://doi.org/10.1177/092137400001200102>
- Espedal, B. (2017). Understanding how balancing autonomy and power might occur in leading organizational change. *European Management Journal*, 35, 155-163. <https://doi.org/10.1016/j.emj.2016.08.005>
- Federal Data Strategy. (n.d.). GSA Technology Transformation Services. <https://strategy.data.gov>
- Fierro, L. A., & Christie, C. A. (2016). Evaluator and program manager perceptions of evaluation capacity and evaluation practice. *American Journal of Evaluation*, 38, 376-392. <https://doi.org/10.1177/1098214016667581>
- Fierro, R. S. (2016). Enhancing facilitation skills: Dancing with dynamic tensions. In R. S. Fierro, A. Schwartz, & D. H. Smart (Eds.), *Evaluation and Facilitation. New Directions for Evaluation*, 149, 31-42. <https://doi.org/10.1002/ev.20177>
- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2011). Origins and current trends in modern program evaluation. In *Program evaluation: Alternative approaches and practical guidelines* (4th ed., pp. 38-63). Pearson Education.
- Flick, U. (2018). Triangulation. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (5th ed., pp. 444-461). SAGE Publications, Inc.

Freidson, E. (1986). *Professional powers: A study of the institutionalization of formal knowledge*. University of Chicago Press.

Freidson, E. (1989). Theory and the professions. *Indiana Law Journal*, 64, 423-432.

Freidson, E. (2001). *Professionalism: The third logic*. University of Chicago Press.

Freudenburg, W. R., & Muselli, V. (2013). Reexamining climate change debates: Scientific disagreement of scientific certainty argumentation methods (SCAMs)? *American Behavioral Scientist*, 57(6), 777-795.  
<https://doi.org/10.1177/0002764212458274>

Friedland, R., & Alford, R. (1991). Bringing society back in: Symbols, practices, and institutional contradictions. In W. W. Powell & P. J. DiMaggio (Eds.), *The new institutionalism in organizational analysis* (pp. 232-263). University of Chicago Press.

fvap.gov. (n.d.). *About Elections*. Federal Voting Assistance Program.  
<https://www.fvap.gov/info/about-absentee-voting/elections>

Gamoran, A. (2018). Evidence-based policy in the real world: A cautionary view. *Annals of the American Academy of Political and Social Sciences*, 678, 180-191.  
<https://doi.org/10.1177/0002716218770138>

- Gee, J. P. (2011). Discourse analysis: What makes it critical? In R. Rodgers (Ed.), *An introduction to critical discourse analysis in education* (2nd ed., pp. 23-45). Routledge.
- Gee, J. P. (2014). *An introduction to discourse analysis: Theory and method*. Routledge.
- Glennerster, R. (2012). The power of evidence: Improving the effectiveness of government by investing in more rigorous evaluation. *National Institute Economic Review*, 219, R4-14.
- Godfrey, L., Funke, N., & Mbizvo, C. (2010). Bridging the science-policy interface: A new era for South African research and the role of knowledge brokering. *South African Journal of Science*, 106(5/6), 1-8.  
<https://doi.org/10.4102/sajs.v106i5/6.247>
- Gopalakrishnan, S., & Zhang, H. (2017). Client dependence and vendor innovation: The moderating role of organizational culture. *Industrial Marketing Management*, 66, 80-89. <https://doi.org/10.1016/j.indmarman.2017.07.012>
- Gray, B., Purdy, J. M., & Ansari, S. (2015). From interactions to institutions: microprocesses of framing and mechanisms for the structuring of institutional fields. *Academy of Management Review*, 40(2), 115-143.  
<https://doi.org/10.5465/amr.2013.0299>
- Green, S. E. (2004). A rhetorical theory of diffusion. *Academy of Management Review*, 29(4), 653-669.

- Green, S. E., Li, Y., & Nohria, N. (2009). Suspended in self-spun webs of significance: A rhetorical model of institutionalization and institutionally embedded agency. *The Academy of Management Journal*, 52(1).
- Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 11(3), 255-274. <https://doi.org/10.3102/01623737011003255>
- Greenhalgh, T., & Russell, J. (2009). Evidence-based policymaking: A critique. *Perspectives in Biology and Medicine*, 52(2), 304-318.
- Greve, H. R., & Teh, D. (2018). Goal selection internally and externally: A behavioral theory of institutionalization. *International Journal of Management Review*, 20, S19-S38. <https://doi.org/10.1111/ijmr.12138>
- Grob, G. F. (2007). The evaluator's role in policy development. In R. Mohan & K. Sullivan (Eds.), *Promoting the use of government evaluations in policymaking*. *New Directions for Evaluation*, 112, 99-108. <https://doi.org/10.1002/ev.211>
- Groves, R. M., & Schoeffel, G. J. (2018). Use of administrative records in evidence-based policymaking. *Annals of the American Academy of Political and Social Sciences*, 378, 71-80. <https://doi.org/10.1177/0002716218766508>
- Governmentwide Direct-Hire Authority for Scientific, Technical, Engineering and Mathematics (STEM) Positions, (2018). <https://www.opm.gov/policy-data->

[oversight/hiring-information/direct-hire-authority/#url=Governmentwide-Authority](#)

- Hall, G. E., & Hord, S. M. (2015). *Implementing change: Patterns, principles, and potholes* (4th ed.). Pearson Education, Inc.
- Hanberger, A. (2001). Policy and program evaluation, civil, social, and democracy. *American Journal of Evaluation*, 22, 211-228.
- Hanneman, R. A., & Riddle, M. (2005). *Introduction to social network methods*. University of California, Riverside.
- Hansen, H. F., & Rieper, O. (2009). The Evidence Movement: The development and consequences of methodologies in review practices. *Evaluation*, 15(2), 141-163. <https://doi.org/10.1177/1356389008101968>
- Hardy, C., & Maguire, S. (2010). Discourse, field-configuring events, and change in organizations and institutional fields: Narratives of DDT and the Stockholm Convention. *Academy of Management Journal*, 53(6), 1365-1392.
- Haskins, R. (2018). Evidence-based policy: The movement, the goals, the issues, the promise. *Annals of the American Academy of Political and Social Sciences*, 678, 8-37. <https://doi.org/10.1177/0002716218770642>
- Haveman, H. A. (1993). Follow the leader: Mimetic isomorphism and entry into new markets. *Administrative Science Quarterly*, 38(4), 593-627.

Higgins, M. C., Weiner, J., & Young, L. (2012). Implementation teams: A new lever for organizational change. *Journal of Organizational Behavior*, 33, 366-388.

<https://doi.org/10.1002/job.1773>

Hogan, S. J., & Coote, L. V. (2014). Organizational culture, innovation, and performance: A test of Schein's model. *Journal of Business Research*, 67, 1609-

1621. <https://doi.org/10.1016/j.busres.2013.09.007>

Institute of Education Science. (2003). *Identifying and implementing educational practices supported by rigorous evidence: A user friendly guide*.

[https://ies.ed.gov/ncee/pubs/evidence\\_based/randomized.asp](https://ies.ed.gov/ncee/pubs/evidence_based/randomized.asp)

Jacob, S., & Boisvert, Y. (2010). To be or not to be a profession: Pros, cons and challenges for evaluation. *Evaluation*, 16, 349-369.

<https://doi.org/10.1177/1356389010380001>

Jennings, J., & Nagel, J. C. (2019). *Federal workforce statistics sources: OPM and OMB* (R43590). Congressional Research Service.

<https://fas.org/sgp/crs/misc/R43590.pdf>

Johnson, J. B. (1998). Toward a theoretical model of evaluation utilization. *Evaluation and Program Planning*, 21, 93-110.

Kadushin, C. (2012). *Understanding social networks: Theories, concepts, and findings*. Oxford University Press.



- Kane, M., & Trochim, W. M. K. (2007a). An introduction to concept mapping. In *Concept mapping for planning and evaluation* (pp. 1-26). SAGE Publications, Inc. <https://doi.org/10.4135/9781412983730>
- Kane, M., & Trochim, W. M. K. (2007b). Using concept mapping in evaluation. In *Concept mapping for planning and evaluation* (pp. 157-174). SAGE Publications, Inc. <https://doi.org/10.4135/9781412983730>
- Kaplan, A. (1964). *The conduct of inquiry: Methodology for behavioral science*. Harvard University Press.
- Kinarsky, A. R. (2018). The evaluation landscape: U.S. foundation spending on evaluation. In S. B. Nielsen, S. Lemire, & C. A. Christie (Eds.), *The evaluation marketplace: Exploring the evaluation industry. New Directions for Evaluation, 160*, 81-96. <https://doi.org/10.1002/ev.20342>
- King, J. A. (1988). Research on evaluation use and its implications for evaluation research and practice. *Studies in Educational Evaluation, 14*, 285-299.
- King, J. A., & Alkin, M. C. (2019). The centrality of use: Theories of evaluation use and influence and thoughts on the first 50 years of use research. *American Journal of Evaluation, 40*(3), 431-458. <https://doi.org/10.1177/1098214018796328>
- King, J. A., & Stevahn, L. (2013). *Interactive evaluation practice: Mastering the interpersonal dynamics of program evaluation*. SAGE Publications, Inc.

King, J. A., & Stevahn, L. (2015). Competencies for program evaluators in the light of adaptive action: What? So what? Now what? In J. W. Altschuld & M. Engle (Eds.), *Accreditation, certification, and credentialing: Relevant concerns for U.S. evaluators*. *New Directions for Evaluation*, 145, 21-37.

<https://doi.org/10.1002/ev.20109>

Kingsbury, N., & Hedrick, T. E. (1994). Evaluator trainign in a government setting. In J. W. Altschuld & M. Engle (Eds.), *The preparation of professional evaluators: Issues, perspectives, and programs*. *New Directions for Evaluation*, 62, 61-70.

Kirkhart, K. (2000). Reconceptualizing evaluation use: An integrated theory of influence. In V. J. Caracelli & H. Preskill (Eds.), *The expanding scopy of evaluation use*. *New Directions for Evaluation*, 88, 5-24. <https://doi.org/10.1002/ev.1188>

Krippendorff, K. (2019). *Content analysis: An introduction to its methodology* (4th ed.). SAGE Publications, Inc.

Labin, S. N., Duffy, J. L., Meyers, D. C., Wandersman, A., & Lesesne, C. A. (2012). A research synthesis of the evaluation capacity building literature. *American Journal of Evaluation*, 33, 307-338. <https://doi.org/10.1177/1098214011434608>

LaDuca, A. (1994). Validation of professional licensure examinations. *Evaluation & The Health Professions*, 17, 178-197.

- Lana, R. E. (1963). Controversy of the topic and the order of presentation in persuasive communications. *Psychological Reports*, 12(1), 163-170.  
<https://doi.org/10.2466/pr0.1963.12.1.163>
- LaVelle, J. M., & Donaldson, S. I. (2015). The state of preparing evaluators. In J. W. Altschuld & M. Engle (Eds.), *Accreditation, certification, and credentialing: Relevant concerns for U.S. evaluators. New Directions for Evaluators*, 145, 39-52. <https://doi.org/10.1002/ev.20110>
- Lawrenz, F., Kunz Kollmann, E., King, J. A., Bequette, M., Pattison, S., Grack Nelson, A., Cohn, S., Cardiel, C. L. B., Iacovelli, S., Ostgaard Eliou, G., J., G., Causey, L., Sinkey, A., Beyer, M., & Francisco, M. (2018). Promoting evaluation capacity building in a complex adaptive system. *Evaluation and Program Planning*, 69, 53-60. <https://doi.org/10.1016/j.evalprogplan.2018.04.005>
- Lawson, R. G. (1968). Order of presentation as a factor in jury persuasion. *Kentucky Law Journal*, 56(3), 523-555.
- Lemire, S., Fierro, L. A., Kinarsky, A. R., Fujita-Conrads, E., & Christie, C. A. (2018). The U.S. federal evaluation market. In S. B. Nielsen, S. Lemire, & C. A. Christie (Eds.), *The evaluation marketplace: Exploring the evaluation industry. New Directions for Evaluation*, 160, 63-80. <https://doi.org/10.1002/ev.20343>
- Lemire, S., Nielsen, S. B., & Christie, C. A. (2018). Toward understanding the evaluation market and its industry—Advancing a research agenda. In S. B. Nielsen, S.

Lemire, & C. A. Christie (Eds.), *The evaluation marketplace: Exploring the evaluation industry*. *New Directions for Evaluation*, 160, 145-163.

<https://doi.org/10.1002/ev.20339>

Leviton, L. C. (2014). Some underexamined aspects of evaluation capacity building.

*American Journal of Evaluation*, 35(1), 90-94.

<https://doi.org/10.1177/1098214013502844>

Li, Y. (2017). A semiotic theory of institutionalization. *Academy of Management Review*,

42(3), 520-547. <https://doi.org/10.5465/amr.2014.0274>

Liebman, J. B. (2018). Using data to more rapidly address difficult U.S. social problems.

*Annals of the American Academy of Political and Social Sciences*, 675, 166-181.

<https://doi.org/10.1177/0002716217745812>

Lincoln, Y. S., & Guba, E. (2004). The roots of fourth generation evaluation: Theoretical

and methodological origins. In M. C. Alkin (Ed.), *Evaluation roots: Tracing theorists' views and influences* (pp. 225-242). SAGE Publications, Inc.

Lincoln, Y. S., Lynham, S. A., & Guba, E. G. (2006). Paradigmatic controversies,

contradictions, and emerging confluences, revisited. In N. K. Denzin & Y. S.

Lincoln (Eds.), *The SAGE Handbook of Qualitative Research* (4th ed., pp. 97-

128). SAGE Publications, Inc.

- Loewenstein, J. (2014). Take my word for it: How professional vocabularies foster organizing. *Journal of Professions and Organizations*, 1, 65-83.  
<https://doi.org/10.1093/jpo.jot004>
- Loewenstein, J. E., Ocasio, W., & Jones, C. (2012). Vocabularies and vocabulary structure: A new approach linking categories, practices, and institutions. *The Academy of Management Annals*, 6(1), 41-86.  
<https://doi.org/10.1080/19416520.2012.660763>
- Lohse, G. L. (1997). Consumer eye movement patterns on yellow pages advertising. *Journal of Advertising*, 26(1), 61-73.  
<https://doi.org/10.1080/00913367.1997.10673518>
- Love, A. J. (1994). Should evaluators be certified? In J. W. Altschuld & M. Engle (Eds.), *The preparation of professional evaluators: Issues, perspectives, and programs*. *New Directions for Evaluation*, 62, 29-40.
- Lund, F. H. (1925). The psychology of belief: A study of its emotional and volitional determinants. *The Journal of Abnormal and Social Psychology*, 20(1), 63-81; 174-195. <https://doi.org/10.1037/h0076047>
- Madison, A. M. (2000). Language in defining social problems and in evaluating social programs. In R. K. Hopson (Ed.), *How and why language matters in evaluation*. *New Directions for Evaluation*, 86, 17-28. <https://doi.org/10.1002/ev.1169>

- Maier, M. A. (2017). Content analysis: Advantages and disadvantages. In M. Allen (Ed.), *The SAGE encyclopedia of communication research methods* (pp. 240-242). SAGE Publications, Inc.
- Manson, H. (2016). Systematic reviews are not enough: policymakers need a greater variety of synthesized evidence. *Journal of Clinical Epidemiology*, *73*, 11-14. <https://doi.org/10.1016/j.jclinepi.2015.08.032>
- Mark, M. M. (2001). Evaluation's future: Furor, futile, or fertile? *American Journal of Evaluation*, *22*(3), 457-479.
- Mark, M. M., Cooksy, L. J., & Trochim, W. M. K. (2009). Evaluation policy: An introduction and overview. In W. M. K. Trochim, M. M. Mark, & L. J. Cooksy (Eds.), *Evaluation policy and evaluation practice. New Directions for Evaluation*, *123*, 3-11. <https://doi.org/10.1002/ev.302>
- Mark, M. M., & Henry, G. T. (2004). The mechanisms and outcomes of evaluation influence. *Evaluation*, *10*(1), 35-57. <https://doi.org/10.1177/1356389004042326>
- Martin, S. A. (2015). *Social capital at the Capitol: A social network analysis of interest group influence in the 111th Congress* [University of Kentucky]. Lexington, KY.
- Maxwell, J. A. (2010). Using numbers in qualitative research. *Qualitative Inquiry*, *16*(6), 475-482. <https://doi.org/10.1177/1077800410364740>

- Maynard, R. (2018). The role of federal agencies in creating and administering evidence-based policies. *Annals of the American Academy of Political and Social Sciences*, 678, 134-144. <https://doi.org/10.1177/000271628768742>
- Maynard, R., Goldstein, N., & Smith Nightingale, D. (2016). Program and policy evaluations in practice: Highlights from the federal perspective. In L. R. Peck (Ed.), *Social experiments in practice: The what, why, when, where, and how of experimental design & analysis*. *New Directions for Evaluation*, 152, 109-135. <https://doi.org/10.1002/ev.20209>
- McDavid, J. C., & Huse, I. (2015). How does accreditation fit into the picture? In J. W. Altschuld & M. Engle (Eds.), *Accreditation, certification, and credentialing: Relevant concerns for U.S. evaluators*. *New Directions for Evaluation*, 145, 53-69. <https://doi.org/10.1002/ev.20111>
- McKay, S., Murray, M., MacIntyre, S., & Kashyap, A. (2015). Evidence-based policymaking and the public interest: Lessons in legitimacy. *Town Planning Review*, 86(2), 133-154.
- Mertens, D. M. (1994). Training evaluators: Unique skills and knowledge. In J. W. Altschuld & M. Engle (Eds.), *The preparation of professional evaluators: Issues, perspectives, and programs*. *New Directions for Evaluation*, 62, 17-27.
- Merwin, J. C., & Wiener, P. H. (1985). Evaluation: A profession? *Educational Evaluation and Policy Analysis*, 7, 253-259.

- Meyer, R. E., Jancsary, D., Höllerer, M. A., & Boxenbaum, E. (2018). The role of verbal and visual text in the process of institutionalization. *Academy of Management Review*, 43(3), 329-418. <https://doi.org/10.5465/amr.2014.0301>
- Monroe, M. C., Fleming, L., Bowman, R. A., Zimmer, J. F., Marcinkowski, T., Washburn, J., & Mitchell, N. J. (2005). Evaluators as educators: Articulating program theory and building evaluation capacity. In E. Norland & C. Somers (Eds.), *Evaluating nonformal education programs and settings. New Directions for Evaluation*, 108, 57-71. <https://doi.org/10.1002/ev.171>
- Morrill, C. (2008). Culture and organization theory. *Annals of the American Academy of Political and Social Sciences*, 619, 15-40. <https://doi.org/10.1177/0002716208320241>
- Morse, J. (2018). Reframing rigor in qualitative inquiry. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (5th ed., pp. 796-817). SAGE Publications, Inc.
- Moynihan, D. P. (1965). *The negro family: The case for national action*.
- Muzio, D., Brock, D. M., & Suddaby, R. (2013). Professions and institutional change: Towards an institutionalist sociology of the professions. *Journal of Management Studies*, 50, 699-721. <https://doi.org/10.1111/joms.12030>
- National Academies of Sciences. (2017). *Principles and practices for federal program evaluation: Proceedings of a workshop*.



<https://www.nap.edu/catalog/24716/principles-and-practices-for-federal-program-evaluation-proceedings-of-a>

National Research Council. (1992). *Principles and Practices for a Federal Statistical Agency*. National Academy Press.

Neuendorf, K. A. (2017). *The content analysis guidebook* (2nd ed.). SAGE Publications.

Nigam, A., & Ocasio, W. (2010). Event attention, environmental sensemaking, and change in institutional logics: An inductive analysis of the effects of public attention to Clinton's health care reform initiative. *Organization Science*, 21(4), 823-841. <https://doi.org/10.1287/orsc.1090.0490>

Nolton, E. C., Varier, D., Powell, M. G., Caracelli, V. J., & Hopson, R. K. (2020a). *The diffusion of value in evaluation by the U.S. Federal Government through the Analytical Perspectives of the U.S. President's Budget*

Nolton, E. C., Varier, D., Powell, M. G., Caracelli, V. J., & Hopson, R. K. (2020b). *Introducing a conceptual framework and method of mapping influences on the process of institutionalizing evaluation in the U.S. Federal Government*.

O'Grady, M. J. (2010). The role of the federal statistical system in evidence-based policymaking or how to make the statistical system essential. *Annals of the American Academy of Political and Social Sciences*, 631, 180-188. <https://doi.org/10.1177/0002716210374154>

- O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford University Press.
- Olejniczak, K. (2017). The game of knowledge brokering: A new method for increasing evaluation use. *American Journal of Evaluation*, 38(4), 554-576.  
<https://doi.org/10.1177/1098214017716326>
- Olejniczak, K., Raimondo, E., & Kupiec, T. (2016). Evaluation units as knowledge brokers: Testing and calibrating an innovative framework. *Evaluation*, 22(2), 168-189. <https://doi.org/10.1177/1356389016638752>
- Oliver, C. (1992). The antecedents of deinstitutionalization. *Organization Studies*, 13(4), 563-588.
- Ottoson, J. M., & Martinez, D. (2010). *An ecological understanding of evaluation use*. <https://www.rwjf.org/content/dam/web-assets/2010/10/an-ecological-understanding-of-evaluation-use>
- Pankaj, V., & Emery, A. K. (2016). Data placemats: A facilitative technique designed to enhance stakeholder understanding of data. In R. S. Fierro, A. Schwartz, & D. H. Smart (Eds.), *Evaluation and Facilitation. New Directions for Evaluation*, 149, 81-93. <https://doi.org/10.1002/ev.20181>
- Pasachoff, E. (2017). Two cheers for evidence: Law, research, and values in education policymaking and beyond. *Columbia Law Reviews*, 117, 1933-1972.

- Patton, M. Q. (1998). Discovering process use. *Evaluation*, 4, 225-233.
- Patton, M. Q. (1999). Enhancing the quality and credibility of qualitative analysis. *Health Services Research*, 34(5), 1189-1208.
- Patton, M. Q. (2000). Overview: Language matters. In R. K. Hopson (Ed.), *How and why language matters in evaluation. New Directions for Evaluation*, 86, 5-16.  
<https://doi.org/10.1002/ev.1168>
- Patton, M. Q. (2015a). Misuse: The shadow side of use. In C. A. Christie & A. T. Vo (Eds.), *Evaluation use and decision making in society: A tribute to Marvin C. Alkin* (pp. 131-147). Information Age.
- Patton, M. Q. (2015b). *Qualitative research & evaluation methods: Integrating theory and practice*. SAGE Publications, Inc.
- Pawson, R. (2006). *Evidence-based policy: A realist perspective*. SAGE Publications, Inc.
- Pawson, R., Wong, G., & Owen, L. (2011). Known knowns, known unknowns, unknown unknowns: The predicament of evidence-based policy. *American Journal of Evaluation*, 32(4), 518-546. <https://doi.org/10.1177/1098214011403831>
- Peck, J., & Theodore, N. (2015). *Fast policy: Experimental statecraft at the thresholds of neoliberalism*. University of Minnesota Press.

- Peräkylä, A., & Ruusuvuori, J. (2018). Analyzing talk and text. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (5th ed., pp. 669-691). SAGE Publications, Inc.
- Però, D. (2011). Migrants' practices of citizenship and policy change. In C. Shore, D. W. Wright, & D. Però (Eds.), *Policy worlds: Anthropology and the analysis of contemporary power* (pp. 244-263). Bergahn Books.
- Phillips, N., Lawrence, T. B., & Hardy, C. (2004). Discourse and institutions. *The academy of Management Review*, 29(4), 635-652.
- Preskill, H., & Boyle, S. (2008). A multidisciplinary model of evaluation capacity building. *American Journal of Evaluation*, 29(4), 443-459.  
<https://doi.org/10.1177/1098214008324182>
- Rapley, T. (2007). *Doing conversation, discourse and document analysis*. SAGE Publications, Inc.
- Richardson, E. L. (1992). The value of evaluation. *New Directions for Program Evaluation*, 1992(55), 15-20. <https://doi.org/10.1002/ev.1614>
- Riggin, L. J. C., Shipman, S. L., & York, R. L. (1995). Using a set-aside to encourage the evaluation of public health service programs. *Evaluation & The Health Professions*, 18(1), 3-12.

- Rizvi, F. (2004). Theorizing the global convergence of educational restructuring. In S. Lindblad & T. Popkewitz (Eds.), *Educational restructuring: International perspectives on traveling policies*. Information Age Publishing.
- Rossi, P. H., Lipsey, M. W., & Henry, G. T. (Eds.). (2019). *Evaluation: A systematic approach* (8th ed.). SAGE Publications, Inc.
- Schlauffer, C., Stucki, I., & Sager, F. (2018). The political use of evidence and its contribution to democratic discourse. *Public Administration Review*, 78, 645-649.  
<https://doi.org/10.1111/puar.12923>
- Schwandt, T. A. (2017). Professionalization, ethics, and fidelity to an evaluation ethos. *American Journal of Evaluation*, 38, 546-553.  
<https://doi.org/10.1177/1098214017728578>
- Schwandt, T. A. (2018a). Acting together in determining value: A professional ethical responsibility of evaluators. *Evaluation*, 24, 306-317.  
<https://doi.org/10.1177/1356389018781362>
- Schwandt, T. A. (2018b). Evaluative thinking as a collaborative social practice: The case of boundary judgment making. In A. T. Vo & T. Archibald (Eds.), *Evaluative thinking. New Directions for Evaluation*, 158, 125-137.  
<https://doi.org/10.1008/ev.20318>

- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (Vol. 39-83). Rand McNally.
- Scriven, M. (2008). The concept of a transdiscipline: And of evaluation as a transdiscipline. *Journal of MultiDisciplinary Evaluation*, 5(10), 65-66.
- Shadish, W. (1998). Evaluation theory is who we are. *American Journal of Evaluation*, 19(1), 1-19.
- Shah, N., Irani, Z., & Sharif, A. M. (2017). Big data in an HR context: Exploring organizational change readiness, employee attitudes and behaviors. *Journal of Business Research*, 70, 366-378. <https://doi.org/10.1016/j.busres.2016.08.010>
- Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, 22, 63-75.
- Shore, C., & Wright, S. (2011). Introduction. Conceptualising policy: Technologies of governance and the politics of visibility. In S. Shore, S. Wright, & D. Però (Eds.), *Policy worlds: Anthropology and the analysis of contemporary power* (pp. 1-25). Bergahn Books.
- Spall, S. (1998). Peer debriefing in qualitative research: Emerging operational models. *Qualitative Inquiry*, 4(2), 280-292.

- Stroufe, R. (2017). Integration and organizational change towards sustainability. *Journal of Cleaner Production*, 162, 315-329.  
<https://doi.org/10.1016/j.clepro.2017.05.180>
- Stack, K. (2018). The Office of Management and Budget: The quarterback of evidence-based policy in the federal government. *Annals of the American Academy of Political and Social Sciences*, 678, 112-123.  
<https://doi.org/10.1177/0002716218768440>
- Stevahn, L., & King, J. A. (2016). Facilitating interactive evaluation practice: Engaging stakeholders constructively. In R. S. Fierro, A. Schwartz, & D. H. Smart (Eds.), *Evaluation and Facilitation. New Directions for Evaluation*, 149, 67-80.  
<https://doi.org/10.1002/ev.20180>
- Stone, D. (2008). Global public policy, transnational policy communities, and their networks. *Policy Studies Journal*, 36(1), 19-38. <https://doi.org/10.1111/j.1541-0072-2007.00251.x>
- Stone, D. (2012). *Policy paradox: The art of political decision making* (3rd ed.). W. W. Norton & Company, Inc.
- Stone, V. A. (1969). A primacy effect in decision-making by jurors. *Journal of Communication*, 19(3), 239-247. <https://doi.org/10.1111/j.1460-2466-1969.tb00846.x>

- Strang, D., & Meyer, J. W. (1993). Institutional conditions for diffusion. *Theory and Society*, 22, 487-511.
- Strang, D., & Soule, S. A. (1998). Diffusion in organizations and social movements: From hybrid corn to poison pills. *Annual Review of Sociology*, 24, 265-290.
- Stufflebeam, D. L. (2001). Interdisciplinary PhD programming in evaluation. *American Journal of Evaluation*, 22, 445-455.  
<https://doi.org/10.1177/109821400102200323>
- Suddaby, R., & Greenwood, R. (1995). Rhetorical strategies of legitimacy. *Administrative Science Quarterly*, 50, 35-67.
- Tabuga, A. D. (2017). Knowledge utilization in policymaking: Evidence from congressional debates in the Philippines. *Journal of Asian Public Policy*, 10(3), 302-317. <https://doi.org/10.1080/17516234.2016.1219144>
- Taylor-Powell, E., & Boyd, H. H. (2008). Evaluation capacity building in complex organizations. In M. T. Braverman, M. Engle, M. E. Arnold, & R. A. Rennekamp (Eds.), *Program evaluation in a complex organizational system: Lessons from cooperative extension*. *New Directions for Evaluation*, 120, 55-69.  
<https://doi.org/10.1002/ev.276>
- Thames, R. C., & Webster, D. W. (2009). *Chasing change: Building organizational capacity in a turbulent environment*. John Wiley & Sons, Inc.



- The White House. (n.d.). *The Cabinet*. <https://www.whitehouse.gov/the-trump-administration/the-cabinet/>
- Tichy, N. M., Tushman, M. L., & Fombrun, C. (1979). Social network analysis for organizations. *The academy of Management Review*, 4(4), 507-519.
- Ting, M. M. (2009). Organizational capacity. *The Journal of Law, Economics, & Organization*, 27, 245-270. <https://doi.org/10.1093/jleo/ewp021>
- Tolman, E. C., Ritchie, B. F., & Kalish, D. (1946). Studies in spatial learning. *Journal of Experimental Psychology*, 36, 13-24.
- Trochim, W. M. K. (1989). An introduction to concept mapping for planning and evaluation. *Evaluation and Program Planning*, 12, 1-6.
- Trochim, W. M. K. (1998). An evaluation of Michael Scriven's "Minimalist theory: The least theory that practice requires". *American Journal of Evaluation*, 19(2), 243-249.
- Trochim, W. M. K. (2009). Evaluation policy and evaluation practice. In W. M. K. Trochim, M. M. Mark, & L. J. Cooksy (Eds.), *Evaluation policy and evaluation practice. New Directions for Evaluation*, 123, 13-32.  
<https://doi.org/10.1002/ev.303>

Trochim, W. M. K. (2017). Hindsight is 20/20: Reflections on the evolution of concept mapping. *Evaluation and Program Planning*, 60, 176-185.

<https://doi.org/10.1016/j.evalprogplan.2016.08.009>

Trochim, W. M. K., & McLinden, D. (2017). Introduction to a special issue on concept mapping. *Evaluation and Program Planning*, 60, 166-175.

<https://doi.org/10.1016/j.evalprogplan.2016.10.006>

U.S. General Accounting Office. (1990). *Prospective evaluation methods*. (GAO/PEMD-10.1.10). Washington, DC.

U.S. General Accounting Office. (2003a). *Ethnographic studies can inform agencies' actions*. (GAO-03-455). Washington, DC: U.S. General Accounting Office.

U.S. General Accounting Office. (2003b). *An evaluation culture and collaborative partnerships help build agency capacity*. (GAO-03-454). Washington, DC: U.S. General Accounting Office.

U.S. General Accounting Office. (2004). *OMB's performance rating tool presents opportunities and challenges for evaluating program performance*. (GAO-04-550T). Washington, DC.

U.S. Office of Management and Budget. (2019). *Phase 1 implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Learning agendas, personnel, and planning guidance*. (M-19-23). Washington, DC: Executive Office of the President.

- U.S. Office of Management and Budget. (n.d.). *Assessing Program Performance*.  
<https://georgewbush-whitehouse.archives.gov/omb/performance/index.html>
- usa.gov. (2019). *Branches of the U.S. Government*. U.S. Government.  
<https://www.usa.gov/branches-of-government>
- van Leeuwen, T. (2007). Legitimation in discourse and communication. *Discourse & Communication, 1*(1), 91-112. <https://doi.org/10.1177/1750481307071986>
- Vo, A. T., & Archibald, T. (2018). New directions for evaluative thinking. In A. T. Vo & T. Archibald (Eds.), *Evaluative thinking. New Directions for Evaluation, 158*, 139-147. <https://doi.org/10.1002/ev.20317>
- Vo, A. T., Schreiber, J. S., & Martin, A. (2018). Toward a conceptual understanding of evaluative thinking. In A. T. Vo & T. Archibald (Eds.), *Evaluative thinking. New Directions for Evaluation, 158*, 29-47. <https://doi.org/10.1002/ev.20324>
- Walk, M., & Handy, F. (2018). Job crafting as reaction to organizational change. *The Journal of Applied Behavioral Science, 54*, 349-370.  
<https://doi.org/10.1177/0021886318777227>
- Warrick, D. D. (2017). What leaders need to know about organizational culture. *Business Horizons, 60*, 395-404. <https://doi.org/10.1016/j.bushor.2017.01.011>
- Wedel, J. R. (2001). Some enchanted era. In *Collision and collusion: The strange case of western aid to Eastern Europe* (pp. 1-13). Palgrave.

- Wedel, J. R. (2004). Studying through a globalizing world: Building method through ethnographies. In J. Gould & H. Secher Marcusem (Eds.), *Ethnographies of aid-exploring development texts and encounters*. Roskilde University.
- Wedel, J. R. (2009). When privatization meets truthiness. In *Shadow elite* (pp. 23-45). Basic Books.
- Wedel, J. R. (2014). *Unaccountable: How the establishment corrupted our finances, freedom, and politics and created an outsider class*. Pegasus.
- Wedel, J. R. (2019). *The mapping method: A guide to charting corruption and influence processes* [Working paper forthcoming]. Global Integrity.
- Weiss, C. H. (1988). Evaluation for decisions: Is anybody there? Does anybody care? *Evaluation Practice*, 9(1), 5-20.
- Worthen, B. R. (1994). Is evaluation a mature profession that warrants the preparation of evaluation professionals? In J. W. Altschuld & M. Engle (Eds.), *The preparation of professional evaluators: Issues, perspectives, and programs. New Directions for Evaluation*, 62, 3-15.
- Worthen, B. R. (1999). Critical challenges confronting certification of evaluators. *American Journal of Evaluation*, 20(3), 533-555.
- Wright, S., & Reinhold, S. (2011). "Studying through": A strategy for studying political transformations. Or sex, lies and British politics. In C. Shore, S. Wright, & D.

Però (Eds.), *Policy worlds: Anthropology and the analysis of contemporary power* (pp. 86-104). Bergahn Books.

Wutzke, S., Roberts, N., Willis, C., Best, A., Wilson, A., & Trochim, W. M. K. (2017). Setting strategy for system change: using concept mapping to prioritise national action for chronic disease prevention. *Health Research Policy and Systems*, 15(69). <https://doi.org/10.1186/s12961-017-0231-7>

Zilber, T. B. (2002). Institutionalization as an interplay between actions, meanings, and actors: The case of a rape crisis center in Israel. *Academy of Management Journal*, 45(1), 234-254.

Zucker, L. G. (1977). The role of institutionalization in cultural persistence. *American Sociological Association*, 42(5), 726-743.

## Biography

Esther C. Nolton, PhD, is an experienced methodologist, evaluator, and facilitator, with specialized skills in quantitative measurement, survey, and qualitative methodologies. She has over 10 years of experience in research and evaluation with her primary areas of expertise related to sports medicine; social determinants of health and education; policy and program implementation; research and evaluation methodology; evidence-based policymaking; and organizational behavior. She has dedicated her career to discovering, studying, and informing issues related to organizational behavior; policy decision-making; democratization and utilization of evidence; strategic planning; and professionalizing evaluation.

Dr. Nolton (née Chou) has several peer-reviewed publications and over 50 local, regional, national, and international presentations. In 2019, she was selected as the winner of the Doctoral Oral Presentation Award by the National Athletic Trainers' Association Research and Education Foundation for a school health policy implementation evaluation funded by the Centers for Disease Control and Prevention and the Virginia Department of Health.

Dr. Nolton is an active committee member for Washington Evaluators, an AEA local affiliate, and also serves as Program Co-Chair for the AEA Evaluation Policy TIG. She was an AEA GEDI Scholar (2018-2019; GEDI15 "Tenacious 10" Cohort) placed at the National Science Foundation with the Evaluation and Assessment Capability section. She is currently a Program Officer in the Evaluation & Analysis department at the Patient-Centered Outcomes Research Institute.