William T. Grant
FOUNDATION

Studying the Use of Research Evidence:
A Review of Methods

Drew H. Gitomer
Rutgers University

Kevin Crouse
Rutgers University

The William T. Grant Foundation supports research to improve the lives of young people ages 5-25 in the United States. The Foundation's focus areas are reducing inequality in youth outcomes and improving the use of research evidence in ways that benefit young people. Its grantmaking centers on high-quality empirical research that has the potential to advance theory, build evidence, and inform policy and practice.

Acknowledgements

Table of Contents

Foreword by Adam Gamoran

This is a monograph by researchers and for researchers, but its message reverberates far beyond the reaches of the research community.

Social scientists undertake their work because they want to make a difference in the real world. Since the pioneering days of Max Weber (1904/1949), social scientists have acknowledged that choices about what to study reflect the values of those who carry out the work. Yet social science research can still be "objective" in the sense that once the object of study has been selected, the research can proceed without bias. What to make of the research findings, however, is ultimately up to the policymaker: while the social scientist can demonstrate how to get from point A to point B, the policymaker decides whether to strive for point B at all.

Even though they realize that policy decisions are not theirs to make, researchers are perennially frustrated when their research findings go unattended in policy and practice. As a funder of social science research, the William T. Grant Foundation shares this frustration. Our mission is to support research to improve the lives of young people; if those in authority fail even to consider the research findings, how can we possibly succeed?

Born of this frustration, for more than a decade the Foundation has supported research on the use of research evidence in policies and practices that affect young people (Tseng, 2012). Under what conditions is research evidence more likely to be used, and how can we create those conditions? Initially focused on understanding the use of research evidence, since December 2015 we have awarded funds for studies of ways to improve the use of evidence (William T. Grant Foundation, 2017). Yet the goals of the Foundation are broader than supporting individual studies that shed light on conditions that increase the likelihood that evidence will be used. We aim to build a robust field of research on knowledge utilization that encompasses a far wider range of studies than we can possibly fund ourselves. This book is an important new resource for that endeavor.

What is it that makes scientific research objective, despite its subjective roots? The distinguishing feature of objective scientific inquiry is its use of systematic methods and reliable, valid measures. Which methods are commonly used in research on the use of research evidence? What assumptions do they embody, what are their limitations, and how might they be improved? In this monograph, Drew H. Gitomer and Kevin Crouse respond to these questions. With examples from education, child welfare, and public health, the authors examine the methods deployed in studies of the use of research evidence. Their critical overview identifies holes in knowledge which should provide fruitful direction for new studies. In each area, they examine "threats to valid interpretation," which researchers would do well to consider as they design new studies. They also helpfully identify the distinct research questions that each methodological approach is best poised to answer. They place particular attention on measurement: how do we know research evidence is being used when we look for it? This vexing challenge

poses a significant obstacle to progress, and Gitomer and Crouse perform a valuable service by identifying the variety of approaches available and discussing their strengths and limitations.

This book provides guidance for social scientists across disciplines and domains—not only those who are already studying the use of research evidence. Scholars of education, for example, are experts in education, not knowledge utilization, but if they are motivated to improve the use of their research findings, this book can help them design studies to test ways to do so. The same holds for researchers in child welfare and public health, and likewise for researchers in other domains such as justice, housing, and workforce development, who may not be represented in the examples but for whom the lessons about the use of social science evidence may pertain equally well.

This book may also be useful to other funders who are involved in supporting research on knowledge utilization. In both public agencies and private philanthropy, leaders are asking how to increase the chances that policymaking is informed by evidence (Maynard, 2018; Granger, 2018). Increasingly, they realize that the key impediment to evidence-based policy is not getting credible evidence *produced*, but getting it *used* (Gamoran, 2018). That recognition is leading many to consider supporting research on ways to bring evidence into greater salience in the policy process. The lessons provided here can help funders across the spectrum assess potential approaches to meeting this challenge.

For our benefactor, William T. Grant, producing knowledge from research and using that knowledge to improve people's lives went hand in hand. "What I have in mind," he said when he established his namesake foundation in 1936, "is to assist, by some means, in helping people or peoples to live more contentedly and peacefully and well in body and mind through a better knowledge of how to use and enjoy all the good things that the world has to offer them" (William T. Grant Foundation, 1986). Over the decades, we have gained much of that knowledge, but seeing it used for social impact remains an ambition largely unfulfilled. This monograph helps us take the next step in our decades-long pursuit, from producing research evidence to improving its use, so that knowledge gained fulfills the promise that our founder intended.

Introduction

A primary motivation for conducting social science research is to inform policy and practice. And yet, the use of research evidence (URE) in these realms has been uneven and uncertain (National Research Council [NRC], 1978, 2012; Tseng, 2012). These circumstances have inspired a line of research, beginning more than two decades ago (Weiss & Bucuvalas, 1980), focused specifically on URE.

The relationship among social science researchers, policymakers, and practitioners is complicated by a number of well-documented factors. Policymakers face shifting priorities and the need to act in short-lived policy windows that do not align with the timescales of the research process, and practitioners require actionable practices that align with a particular configuration of constraints and affordances. Researchers tend to be focused on the integrity of the methods and whether the inferences made from the results are sound, given the sample for a particular study, while stakeholders are less concerned with internal validity than with whether conclusions from research will apply to their current or potential future contexts.

Such divisions between researchers and practitioners have been well described (NRC, 1978, 2012), but subsequent studies of the effective use of research[1] and the conditions to facilitate such use are still nascent. There has been some effort to build cross-disciplinary programs of study surrounding URE (DuMont, 2016; NRC, 2002; Tseng, 2012), but significant questions about what URE means and how to assess it in practice remain. Collectively, these studies reveal URE to be complex, nuanced, highly varied, and often unpredictable. The strength of these studies is that they cut across disciplines, theoretical frameworks, and social problems, and collectively provide a broad and flexible definition for how research may be used in practice.

A selective review of studies is included to illustrate particular examples of how URE is assessed. In this monograph we focus explicitly on research in the fields of education, child welfare, and public health. This triad of topics is broad and allows for cross-cutting inferences about the ways in which URE is conceptualized across relevant social sciences while also being sufficiently bounded to be explored in depth in a single report. We also find that the particular characteristics of each of these fields can, in some cases, be more or less amenable to particular methodological approaches.

---

1   This monograph focuses on research about the use of research evidence. To avoid confusion and ambiguity, we use the terms *research* and *research evidence* to refer to works that are used by practitioners, policymakers, and intermediaries and are the object of study in the field of URE. We use the terms *study* and *studies* to refer to the scholarly initiatives that investigate URE.

The studies were selected to represent a range of methodological approaches, domains, and facets of URE. We began by reviewing URE-focused projects conducted by researchers supported by the William T. Grant Foundation and related work by these scholars. We identified other relevant studies by reviewing the references cited in those works. Finally, we conducted independent literature searches that paired methodological approaches with terms that included *use of research evidence*, *evidence-based practice*, *evidence use*, and *research-based practice*.

Authors of each of the selected examples were invited to review the content of the summaries and discuss issues raised by their study. Some of their insights are included within this monograph.

These studies have shed light on the complexity of URE by asking fundamental questions about the existence, participants, processes, contextual factors, and outcomes of URE. Tseng (2012) identified a broad range of questions being asked:

- Who are the research users—including organizations, decision makers within the organizations, and intermediaries—who translate and package research for use by policymakers and practitioners?
- How is research defined? Distinctions are sometimes made between evidence and research, for example. Research sometimes refers to unique inquiries, syntheses, or products and tools that have some basis in research.
- How is research acquired? This question represents a significant focus on the users of research rather than on the dissemination of research.
- How is research interpreted? How do users interpret the quality, meaning, and implications of research?
- How is research used? Tseng (2012) built on the work of Weiss and Bucuvalas (1980) in distinguishing uses as instrumental, conceptual, symbolic, and political.
- How do human relationships and interactions influence the use of research?
- How do organizational, political, and policy contexts influence the use of research?
- How does research evidence compete with other sources of information to influence policy decisions and practices, and what are the factors that influence the relative attention to different sources of information? What conditions develop research that will or can be useful for policy and practice?
- What are the effects of URE on important social outcomes (e.g., Gormley, 2011; NRC, 2012)?

Addressing these questions and assessing URE has required the development and deployment of measures and methods that provide insight about URE. These measures and methods derive from different methodological traditions. Used alone, or in concert, they serve as particular lenses and tools for understanding different aspects of and questions about URE.

The purpose of this monograph is to provide a conceptual overview of how different measures and methods have been used to assess URE. We consider the relationship of different methods and measures to underlying conceptions and purposes of URE and how they contribute to an understanding of URE. This monograph also describes the particular affordances and limitations of specific methodological approaches.

In organizing this review, we debated how best to characterize the methodological approaches that have been taken to study URE. The terms *assessment/evaluation, method,* and *measure* all have multiple connotations that could inadvertently lead readers in unintended directions. The term *measure* often connotes a scaled quantitative assessment, but many of the studies reviewed take stock of URE through qualitative explorations. Therefore, this monograph differentiates these terms in the following ways:

- *Assessment/evaluation* is used to represent the overarching goal of these studies, to assess or make an evaluative interpretation (Kane, 2006; Messick, 1989) based on evidence. All studies that examine URE make an assessment of some aspect(s) of the phenomenon.
- *Methods* and *methodology* are used to represent the categories of inquiry processes used to explore URE. These categories vary widely and can include data collection methods, such as interviews, surveys, and observations, as well as research design and analysis methods, such as experimentation, case studies, and social network analysis (SNA), each of which makes use of different types of collected data.
- The term *measures* is used to represent the specific instruments, protocols, tools, and coding systems that are employed for carrying out an inquiry. While measures often have a connotation of a scaled quantitative assessment, they do not need to and we do not intend for them to be interpreted in such a context.

The remainder of the monograph is organized into three sections. The first section provides a set of core methodological issues and ideas that are central to the study of URE. All of these issues are addressed in great detail in methodology texts and reports, and our intent is not to reiterate those points. Instead, our goal is to highlight key issues as they apply to studies of URE. The second section is organized around specific methodological approaches. Each sub-section discusses how a selected methodology can support inferences about URE with respect to the framework. Examples are used to demonstrate the potential of particular methods to address certain aspects of URE as well as their limitations in addressing other aspects. Also discussed are critical considerations for each method that can contribute to the strength of URE interpretations.

The final section identifies several key opportunities for moving the field forward.

Assessing URE has required the development and deployment of measures and methods that provide insight about URE. These measures and methods derive from different methodological traditions. Used alone, or in concert, they serve as particular lenses and tools for understanding different aspects of and questions about URE.

Methodological Issues in the Study of URE

The methodological approaches reviewed in this monograph are all intended to contribute to an understanding of social phenomena associated with URE. Every methodology has guidelines for appropriate execution as well as standards for judging the validity, rigor, and quality of results and interpretations. Specific methodologies are designed to support certain kinds of inferences but not others, and these distinctions are highlighted in the treatment of the individual methods that follow.

Organization of Methods

Methods can be organized in different ways. Given the context of URE, we have found it most helpful to think about two general categories. The methods we focus on here represent the vast majority of those used to study URE. This does not imply that other methods could not be useful, only that they have not been pursued in the literature to date.

Within and across these categories, a range of approaches to scientific inquiry is included. Many approaches come out of interpretive traditions in which the goal is to "understand what people mean and intend by what they say and do and to locate those understandings within the historical, cultural, institutional, and immediate situational contexts that shape them" (Moss & Haertel, 2016, p. 133). These most frequently involve qualitative approaches in which researchers develop their claims through their own construction of understanding (e.g., Creswell, 2012).

Other approaches come from post-positivist traditions and typically use quantitative methods to examine specific variables, engage in explicit hypothesis and theory testing, and use structured instruments that yield statistical and/or relatively structured data (e.g., Creswell, 2012).

The first category of methods is associated with particular data collection methodologies. It includes interviews, surveys, observations, and document analysis. In order to study URE, researchers collect data on what people say about what they do, think, and believe (interviews and surveys); what they do in relevant contexts (observations); and what they produce (documents). Each of these data collection approaches also has associated procedures and processes for analyzing data and reporting results.

The second category of methods is associated with particular study designs and makes use of data collected via the first set of methods. Nothing illustrates this better than social network analysis (SNA); high-quality SNA studies have used surveys, transcripts, and interviews to construct the networks that are analyzed. In order to explore causal effects, experiments and quasi-experiments can use data from all of these data collection methods as independent variables, covariates, and outcome measures. Finally, mixed-method designs use multiple measures to investigate a particular case. Because URE crosses domains and contexts so pervasively, nearly every study of URE is a case study, and so we consider mixed-method approaches only in the context of case studies.

Issues related to the validity, rigor, and quality of specific methods as applied to URE also are described in the individual sections. This introductory piece considers issues that cut across methodological approaches and are germane to all studies of URE. How these issues are addressed, of course, is situated in the choice of methods. Nevertheless, every study needs to attend to these issues in order to ensure interpretations that are meaningful and useful for understanding URE.

These issues represent a distillation of ideas from three sources of evidence. First, we read widely both conceptual and empirical studies of URE. Second, we have benefitted from discussions with authors of URE reports cited throughout this document who shared their insights with us. Third, we had the benefit of reviews on earlier drafts from William T. Grant Foundation program staff and several anonymous external reviewers.

Clarifying the Object of Study

Core ideas about URE in policy and practice go back at least four decades (e.g., Caplan, 1979; Weiss & Bucuvalas, 1980) and have further evolved (e.g., Kingdon, 2011; Tseng, 2012). Studies of URE continue to embrace the ideas contained in seminal works and focus on different aspects of URE. Regardless of method, it is necessary for studies to locate themselves within the space of URE. Such studies are typically situated at the intersection of three broad elements: 1) influences on research use; 2) targets of research use; and 3) claims about research use.

Influences on Research Use

A critical question in studies of URE is: What factors influence the extent to which research is used by policymakers and practitioners? There are a host of contextual factors, including the political climate (e.g., policy mandates for the use of evidence-based practice), organizational practices, the presence and roles of intermediaries, and individual characteristics (e.g., knowledge of and attitudes toward research). To understand the influence of these factors on URE, studies have examined relationships between individuals and organizations, the interactions between and among relevant participants, documentary evidence, economic and other constraints and affordances, and the institutional capacity for change.

Targets of Research Use

Studies have examined how research has been used to study different facets of policy and practice development and implementation, why research is used, and what the impact of such research use has been. Facets include processes related to the access of research; the use of research in the development of policy and practices; and its enactment in code, statute, practice guidance (e.g., curriculum guides, professional standards), and implementation. Studies that have examined the purposes for which research is used have often employed Weiss and Bucuvalas' (1980) typology of types of research use (e.g., instrumental, conceptual, symbolic, and imposed uses) to frame the work. However, other conceptual frameworks have also been used. Studies of interventions have been used to examine their effects on particular targets of research use.

Claims About Research Use

Methodologies are selected to investigate URE because they can support particular kinds of claims. Case studies are designed to provide thick description of a particular setting or event. Methods such as observations, interviews, and document analysis are qualitative methods intended to support descriptive claims about how research is used, how people think about its use, and the mechanisms that lead to research use. A range of methods has been used to make claims about the strength and/or nature of relationships of URE to characteristics of individuals, organizations, contextual factors, and interactions.

Other studies attempt to make claims about the distribution of URE-related constructs across individuals and settings. Most often this is done through survey methodologies. Finally, causal claims about the effects of specific interventions on URE typically employ experimental and quasi-experimental methods.

Targets of Generalization

The vast majority of studies of URE is embedded in particular cases. Therefore, the traditional approach used in much research of statistically generalizing results from a sample to a population is not generally relevant to the study of URE. Instead, the most common concern of generalization is analytic and involves building theory that can be generalized to different cases and contexts. Yin (2014) describes the issue as follows:

> *Case studies, like experiments, are generalizable to theoretical propositions and not to populations or universes. In this sense, the case study, like the experiment, does not represent a "sample," and in doing case study research, your goal will be to expand and generalize theories (analytic generalizations) and not to extrapolate probabilities (statistical generalizations). (p. 21)*

Moss and Haertel (2016) (see also Firestone, 1993) identified several ways that analytic generalizations can be made. One is to apply the theoretical concepts and mechanisms to other cases to determine whether concepts developed through one case can be used to help understand and explain other cases. Certainly, this is being done in the field of URE as researchers from different domains draw on theoretical concepts developed from one case and apply them to others (e.g., Kingdon, 2011; Majone, 1989; Weiss & Bucuvalas, 1980).

Theoretical Underpinnings

Almost all studies contain an explicit theory of URE, which, as previously mentioned, is often built on theories developed in other studies. Reports are typically explicit about two dimensions of URE. First, it is incumbent on researchers to define what they mean when they refer to *research evidence*. Across studies, including the specific examples cited in this monograph, the term is used in different ways, from relatively bounded definitions limited to academic articles (e.g., Newman, Cherney, & Head, 2016) to much broader definitions that can include a range of research documents and ideas from external sources, as well as evidence that is produced within local institutions (see Tseng, 2012).

Second, most published studies of URE are explicit about the theoretical perspective in which the scholarly argument is grounded. In some cases, the study is informed by a particular theory of research. In other cases, researchers have synthesized multiple theories to conduct and report their work. Theory, of course, should be linked to and help address the particular research questions of the study.

A third dimension related to theoretical underpinnings is almost always tacit. Through our discussions with some researchers whose articles we have cited, it became obvious that theory development for a given study may be much more iterative than is obvious from a simple read of the research report. Theory can evolve and theoretical commitments can be made as data are collected and preliminarily analyzed. Researchers do not start with a blank slate. They bring significant knowledge of URE theory as well as deep expertise of the context they are studying. Though they may not start with commitment to a particular theory, they are highly disciplined in working toward such a commitment. They bring together empirical data and their deep and developing knowledge to hone in on a theoretical perspective.

Of course, many other studies are more linear in terms of theoretical commitments. The researchers begin with a strong theoretical commitment, and the full data collection and analytic plan is structured around that commitment.

Sampling

Every methodological approach has unique sampling considerations. When trying to generalize from a sample to a population, it is important to have a sample that is representative of the population. Other sampling issues are critically important and affect interpretations of results across all studies. The fundamental question is whether the data that have been collected and analyzed are sufficient to support the interpretations of the study. Aspects of studies that need to be considered include:

- Users: If the study is attempting to understand how research is used within a system, is there adequate sampling of individuals involved in the use of research? Are key people missing? Are the voices of particular people overemphasized?
- Occasions: If the study is examining deliberation processes by practitioners or policymakers, have data for all critical occasions been collected and analyzed? Is the study team aware of occasions that they may have missed? Is the study missing important kinds of meetings or interchanges (which may be informal or outside of scheduled meetings) where URE is important?
- Documentation: Policy documents are often used in studies of URE. Analogous issues pertain to documents. Have the appropriate documents been collected and analyzed to provide for a full understanding of URE to address the research questions of interest?

Lens

There are multiple lenses that researchers have used to explore URE. In reviewing these studies, we highlight two distinctions that have significant methodological and interpretive consequences. At the risk of simplification, the first distinction focuses on how studies are entered. One approach is to ask a general question: What is the relevant evidence about URE in a particular policy or practice context? A study like this might look for evidence of research use in places like school board deliberations, health policies, or legislative-related documents. The analytic lens would examine issues such as what research is being used, how it is being used, or by whom it is being used.

Another way of entering these studies is from a policy or practice perspective, asking a question such as the following: How are policy- or practice-related decisions made, and what is the role of URE in making those decisions? A study of this type is likely to examine a broad set of factors that influences policy and practice and then situate the use of research evidence within that problem space.

The former type of framing has the potential to address very targeted questions about research use. But such methods alone would not shed light on the relative role of URE within the policy or practice context vis-à-vis other considerations that are known to influence decisions (see Kingdon, 2011).

The second distinction concerns whether the focus is on practice or on a specific policy. Many studies focus on the kinds of URE practices that are used inside organizations and institutions. These studies focus on factors that are related to patterns of use that cut across specific practices and policies. Other studies, though, are focused on URE as applied to a particular issue. Several examples of these kinds of studies are presented in this monograph. McDonnell and Weatherford (2013) were interested in URE in the context of a specific educational policy initiative, the Common Core State Standards (CCSS; National Governors Association [NGA], Council of Chief State School Officers [CCSSO], & Achieve, Inc., 2008). Garces, Marin, and Horn (2017) focused on URE in the context of *amicus curiae* briefs related to a specific U.S. Supreme Court case. These kinds of studies are not trying to make claims about practices in general. They are bounded in terms of how URE was used in the shaping of a particular policy.

Pragmatism

Studies of URE are typically carried out in very complex contexts and can involve issues that are confidential, politicized, and otherwise sensitive for a host of reasons. Studies must rely on voluntary participation at institutional and personal levels. For efforts that require multiple years of data collection, it is important to note that turnover rates in some organizations and fields are very high. All of this is to say that studies often have to live with research implementations that are significantly compromised from original design plans. Samples may not have the representation initially expected, attrition can be high, and non-response is frequently a problem. Under these conditions, studies not only report these challenges but also attempt to give guidance on potential consequences for the interpretation of results, with appropriate caveats.

Appropriateness of Mixed Methods

As will be elaborated in the following sections, all methods have particular strengths and limitations. Any method can provide insights about part of the story. To understand URE, it is essential that multiple and complementary methods are used (Creswell, 2012). Fortunately, this is a perspective that is dominant within the community of URE researchers and funders. Researchers typically make use of multiple methods to understand a problem. Sometimes those methods are presented in a single study, while other times they are divided into multiple, related studies. In reading the literature, speaking with researchers, and participating in URE community gatherings, one particular observation that we have made is that there exists a deep respect for the affordances of different methods, an understanding of how they each contribute to the literature, and a lack of primacy of any single methodology.

Data Collection Methodologies

Interviews

Interviews represent a range of data collection and analytic approaches based on the self-report of respondents to questions asked by the researchers. The purpose of interviews ranges widely. Semi-structured instruments are intended to provide more elaborative detail than surveys can provide while qualitative interviews are open-ended and intended to elicit responses on how an individual makes sense of a particular context or issue (R. Edwards & Holland, 2013). In semi-structured interviews, exchanges begin with common questions for all respondents but then can involve more unique follow-ups and exchanges between interviewer and respondent.

Kvale and Brinkmann (2015) provide a rationale for qualitative interviews: "In general, when the research question can be formulated using the little word *how*, there is a good chance that qualitative interviews are relevant. *How* is something experienced? *How* is something done?" (p. 127).

Qualitative interviews derive from ethnographic traditions used to explore people's knowledge, interpretations, experiences, and understanding (see Mason, 2002) of any and all facets of URE within some specified context. Interviews are used to understand how people make meaning of URE and how those meanings shape, and are shaped by, the social contexts (e.g., Eisenhart, 2001). Context can refer to a broad setting (e.g., *How is research used within an organization that functions within a policy context?*) as well as a narrow context (e.g., *How was research used in the development of a specific policy that was enacted?*). Interviews can be used to inquire about the perceptions people have regarding how participants engage with URE, how they value research, and how particular policies and practices are developed. Interviews can also be used to reveal how respondents understand the relative salience of URE in advancing policy and practice in light of other influences.

In many URE studies, including some of the examples cited in this section, semi-structured interviews are used. Compared to surveys, semi-structured interviews allow for deeper probing of particular issues that can provide more insight, but the structured set of common questions still provides a common base for comparison among participants. However, the flexibility to probe and elaborate in interviews requires substantially more involvement and time commitment on the part of the research staff in conducting interviews and analyzing their results, so sampling is typically much smaller and more constrained. Thus, studies with the resources to engage in both surveys and interviews have the potential to jointly meet the goals of developing a deeper understanding of the meaning of the common responses through the interviews as well as a more consistent understanding of the incidence of different responses through the broader sampling of the surveys.

Methodology

Interview methodologies vary significantly, depending on their degree of structure. Post-positivist survey approaches tend to be more structured and have the goal of eliciting as much relevant thinking as possible through relatively consistent questioning. Therefore, norms for structured interviewing lead the research team to design questions to elicit relatively specific, low-inference responses and pose the questions in a predetermined sequence and with predetermined wording. An interview guide also provides boundaries for the kinds of responses and follow-up questions that the interviewer can pursue after the initial response. These interviews, nonetheless, are far from standardized or mechanical. In order to know when to ask follow-up questions and what questions to ask, the interviewer must have well-developed understandings of URE and the particular context being studied.

Once the data are collected, the interviews are transcribed and then analyzed, typically using a coding structure and coding software. Analytic processes for structured interviews may be based on a predefined codebook that is built from the research questions, prior literature, or theoretical background. The processes may also be more inductive, with the researcher using cross-cutting themes that emerge from the interview data to attempt to generalize areas of focus, response, or beliefs. The inductive approach to qualitative coding is a core component of many areas of qualitative inquiry, particularly that of the grounded theory methodology (Corbin & Strauss, 2008; Glaser & Strauss, 1967). At all steps in the process, quality-control procedures are implemented to ensure accuracy of transcription and reliability of coding.

Unstructured interviews are focused on a particular context and issues, but the process is much more fluid and conversational. These interviews are opportunities for individuals to share their experiences and understandings, and, as these come to light, the interview may take different turns. Interviews across individuals within a context are likely to vary not only in the participant response but also in the prompts that the interview provides, as they will emerge organically from the individual's responses. For example, within an organization, individuals in different roles might have different perspectives of and engagement with URE, leading to interviews that have different flavors.

Less structured interview formats almost universally rely on some form of inductive coding, as it is unlikely that themes and even topics of conversation can be identified ahead of time. In order to have these kinds of fluid conversations, the researcher must be extremely well versed in the domains under investigation, both URE in general and the specific context and situation that is being investigated. Researchers should be prepared for certain issues to arise and have plans for how to respond to them, as well be prepared to continue to elicit explanation when the conversation ranges outside of their familiarity.

For all forms of interviewing, researchers need to develop the skills to encourage the interviewee to share their understanding of URE while also ensuring that conversations remain focused within some boundaries. Because interviewing relies on self-report and requires a member of the research team to conduct the interview and follow up on responses, a level of rapport and trust must be established between interviewer and participant. This is not a trivial task and is an important topic to consider, especially for less structured interviews that seek to solicit more depth into the participant's beliefs or individual processes regarding the subject area. Implicit racial and gender bias (both for the interviewer and in the participant's perception of the interviewer) can affect responses, and the perception that the interviewer is a member or an outsider of the relevant community of focus can also impact how and whether the participant will respond (Creswell, 2012; Zinn, 1979).

The sample for interviews should be appropriate for the URE study questions and can vary widely for questions that are seeking to draw conclusions (e.g., *How is the U.S. Department of Education attempting to leverage recent trends in academic research since the new administration took office?*) compared to questions that are more exploratory in nature (e.g., *What new or novel approaches are emerging by which socially active individuals are employing research to inform their involvement in their local community?*). When the focus of the study is to move toward drawing conclusions about URE for a given context, participants need to represent the variety of roles within that context. It is especially important to ensure that key participants or groups are not missing from the sample.

For research questions that are more exploratory, it is possible to employ other sampling techniques that may not reach all groups as long as the analysis and discussion do not attempt to make claims based on the responses that were collected; in fact, it may be impossible to identify all populations and subgroups, as might be the case for an informal characteristic like "socially active individuals." Accordingly, for organizations and contexts with well-defined boundaries and divisions, the selection of the sample is typically defined at the beginning of data collection. However, for more exploratory interviews, the sample may be emergent as interviews bring to light roles or key individuals who should also be interviewed.

For introductory guidance on a range of interview methods, the following may be useful: R. Edwards and Holland (2013); F. J. Fowler and Mangione (1990); Galletta and Cross (2013); Gubrium, Holstein, Marvasti, and McKinney (2012); Kvale and Brinkmann (2015); Mason (2002); Rubin and Rubin (2012); and Seidman (2013).

Threats to Valid Interpretation

As with any self-report measure, interviews measure individuals' interpretations of a particular reality and their willingness to express it honestly and directly to the interviewer. To varying extents, these accounts may not always align with findings from other data sources and methods. Certainly, there are many URE studies, a number described in this monograph, in which there are inconsistencies between what individuals say and what is observed through other means. There are many ways in which the execution of an interview study can be problematic, including issues of trust, poor question design leading to a misunderstanding of the intended inquiry, inadequate or inappropriate follow-up questions, or weak analytic approaches. The general references provided in this section provide guidance on these issues. Three particular cautions are raised with respect to URE.

First, as with many interviews, there is always a threat of social desirability (A. L. Edwards, 1982): the tendency of interviewees to respond in a way that they perceive will be viewed positively by others. If interviewees believe that using research evidence is more favorably perceived than not using research, researchers need to be cognizant that responses may suggest a greater use of evidence than may actually be present. An effective way of addressing some of these concerns is to focus interviews on particular actions that have occurred; one method of doing this is to follow up on observed events of research use or previously produced documents (e.g., McDonnell & Weatherford, 2013). When researchers ground the questions in tangible events or artifacts, interviewees are less apt to respond in abstractions that are not based in specific experiences. A cognitive interview approach (e.g., Ericsson & Simon, 1993) that asks individuals to think aloud about their reasoning through some specific events or decisions about URE is another method that can be used to minimize socially desirable responding.

Second, it is important to ensure that the sample interviewed adequately represents the critical actors necessary to answer the research questions in the relevant URE context. Especially when the analysis seeks to draw conclusions about research use, there is a risk of both overestimating and underestimating URE depending on which individuals within the study context are actually interviewed.

Third, the analytic framework has a significant impact on the way that data are reported, understood, and interpreted. In open coding, which is the most common form of analysis we found in current URE research, cross-cutting themes are generalized about some aspect of URE on the basis of the collected interview data, and those themes are then analyzed to draw inferences to respond to the research questions. Particular excerpts are often reported in order to illustrate each broad theme. Of general concern regarding such analyses is the possibility that other researchers, given the same corpus of interview data, would construct substantially different generalized themes or feel that the chosen excerpts are not representative of the theme or the respondents. In order to ensure that the generalizations are defensible, steps should be taken to make the generalization process transparent, including ensuring that the research staff is well calibrated to each other, and engaging multiple researchers in verifying inferences against each other.

Questions Interviews Can Address

Interviews are powerful tools for understanding how practitioners and policymakers engage in and think about URE. They can be used to find out how individuals make sense of virtually all aspects of URE. Interviews can uncover specific examples and issues of URE that individuals perceive as important or relevant in some way. Interviews can highlight factors that practitioners and policymakers view as supporting or limiting URE. Specific questions that can be asked include:

- How did URE (and other factors) influence your thinking and/or decisions about policy or practice?
- What are the conditions, experiences, and/or interaction that influence how you or your institution thinks about and/or engage with URE?
- What are the ways in which your institution and individuals in your institution engage with URE?
- How has your URE or your institution's URE changed over time?

Questions Complementary Methods Can Address

Interview approaches can be very useful in generating hypotheses that can then be probed more systematically for purposes of both generalization and causation. Following an exploratory interviewing phase in which the research team establishes the major themes for URE and the range of likely responses, surveys can be designed and given to broader and more representative samples in order to determine the generalized patterns of the responses established through interviews or to discover if those patterns hold up across different contexts.

As self-report data, other methods can be used to support (or challenge) claims made on the basis of interview data. Methods such as observation, discourse analysis, and document analysis all provide evidence about specific manifestations of URE. These methods can all be used to triangulate and/or corroborate the interview data.

Example: Using Interviews to Study URE in Education
– Coburn and Talbert (2006).

As part of the No Child Left Behind Act (Elementary and Secondary Education Act of 2002), school districts were required to show that their school improvement choices were guided by research and data evidence to increase student learning outcomes. Coburn and Talbert (2006) used interviews to investigate how school administrators who occupied different roles within a district conceptualized the use of evidence and research to improve student learning.

The research inquiry was prompted by prior work in educational policy enactment that demonstrated how preexisting beliefs of individuals in the system influence how policies are taken up (Coburn, 2001; Spillane, Reiser, & Reimer, 2002). Coburn and Talbert (2006) looked across a district to better understand how individuals with different organizational roles, responsibilities, and educational reform histories made sense of four issues:

1.   What makes evidence of student learning valid;
2.   How evidence of student learning should be used;
3.   What makes research "high quality"; and
4.   How much faith should be placed in research as a guide to practice.

The data for this work came from 17 interviews of 10 top district administrators and 38 interviews of 14 frontline administrators who interfaced with school leaders in one large, urban school system. Coburn and Talbert (2006) also selected eight district schools in which they interviewed the principal and instructional leaders and held focus groups with teachers. They supplemented the interview data with observations and related documents at both the school and district levels.

Data analysis followed established qualitative methods in which interview data were iteratively coded about the participants' conceptions of valid evidence and its use as well as their opinions about research and research quality. Participants were then placed in a matrix to establish grounded typologies, and prior theory was used to analyze how participant characteristics and roles related to their placement in the typology. Consistent with qualitative research employing grounded theory (Corbin & Strauss, 2008), the researchers reported their findings through organizing their analyses by emergent themes and providing rich description to illustrate variation, complexity, and patterns. Top-level sections were ordered by the four overarching areas of focus, and subsections listed the different typological categories. This format allowed the researchers to describe and organize the considerable range of perspectives represented by the sample of individuals who participated in the interviews.

In the analysis, the researchers found that two key factors substantially influenced the patterns observed in the reported typologies. The nature of work roles appeared to shape individual conceptions, particularly with respect to how much of their responsibilities were associated with meeting accountability demands. District-level administrators were much more likely to have conceptions of high-quality research rooted in scientific rigor or the cumulative research base and faith in the value of that research, whereas school-level staff were less likely to have strongly rooted conceptions of research and more likely to express skepticism in its value. Top-level district administrators and school principals were also more likely to conceive of evidence validity in terms of measurement properties, whereas both frontline administrators and teachers were more likely to consider validity in terms of providing insight into student thinking that was rooted in teacher judgment. Additionally, the organizational structure, professional networks, and individuals' involvement in the district's reform history all contributed to shaping their conceptions of evidence and research use. In particular, those who had participated in prior reforms tended to continue to hold the views expressed by those

reform movements. To the extent that there were disagreements about evidence-based practice, they were associated with organizational unit and prior experiences with reform initiatives.

Combining these findings in the discussion, the authors suggested several ways that districts could support research and evidence use. They proposed that the relation of different conceptions of evidence to different work needs may suggest that districts develop a strategy that acknowledges these differences and creates mechanisms to support the use of different evidence types for different roles and also coordinates and combines evidence as necessary. Further, they found that the frontline central office administrators who interface directly with schools play a key role in mediating the different conceptions that top-level district administrators and school-based staff have regarding evidence, research, use, and validity. Finally, they found that participation in targeted reform movements did have lasting effects on the conceptions of research and evidence, especially when the participating subunits had sufficient interaction to develop norms that incorporated aspects of the reform initiative.

Example: Using Interviews to Study URE in Education – McDonnell and Weatherford (2013).

McDonnell and Weatherford (2013) provided an interesting contrast to Coburn and Talbert (2006) in that the focus of research and use of interviews were much more specific. The study explored how different stakeholders used research evidence through the process of promoting, developing, and adopting the Common Core State Standards (CCSS) (NGA et al., 2008), a major educational reform. Stakeholders were all deeply involved in the CCSS and were designated as *policy entrepreneurs,* individuals who advocate for particular policy positions in order to realize some type of perceived benefit (Kingdon, 2011). Of critical importance is the fact that the CCSS were founded on the principle that the standards would be grounded in evidence from research.

Given the particulars of the CCSS, there were multiple and contested perspectives throughout the process. McDonnell and Weatherford (2013) were interested in how research was martialed by these different key stakeholders, which included leaders of the CCSS movement, members of committees and work groups that worked on the CCSS, national and state education policymakers, as well as researchers and critics of the CCSS.

The study embedded research use within a broader frame of evidence use, recognizing that various types of evidence are used to inform and persuade as policy evolves. McDonnell and Weatherford (2013) also identified a policy cycle that included three discrete phases of policy development: *problem definition, policy design,* and *policy enactment.* The fundamental question guiding this research was how research was used by different stakeholders across these three phases.

The study involved 111 interviews with various stakeholders over two years during the different stages of CCSS policy development. Forty-nine of the interviews involved national stakeholders, and the remainder were done within four states. The interviews were structured and focused on "the politics and process of Common Core promotion, development, and adoption; why participants chose to use certain types of evidence; and what other types were either unavailable or not used" (McDonnell & Weatherford, 2013, p. 7).

As noted, one of the inherent risks of interpreting interviews is that self-reports can inaccurately represent what people actually do. One of the steps taken by McDonnell and Weatherford (2013) to address this threat to validity was to review relevant documents

prior to the interviews. Using the documents as reference points, the researchers queried interviewees about actions and reasoning around specific events in which they were involved and grounded the questions in particular documentary evidence.

While the authors did not provide detailed information about how these structured interviews were coded, they did highlight certain important aspects of their method. First, they engaged key organizations in identifying stakeholders to be interviewed. Second, the data analysis was structured in terms of stages of the policy cycle.

The interviews revealed that while most stakeholders accepted a large body of research that showed weaker-than-desired academic performance of U.S. students, there were important differences among stakeholders in ascribing causal factors to these assessment result patterns. Whereas the strongest advocates of the CCSS pointed to research that supported the need for stronger academic standards, the stakeholder group of researchers tended to focus on the lack of capacity within the educational system to implement new policies. Thus, while different constituencies acknowledged different strands of research and even held similar values in wanting to improve academic performance of all students, they also gave priority to these different strands in conceptualizing the problem that had to be solved.

Interview data that focused on policy design and implementation revealed the various ways that research was used, and not used, in developing educational policy. McDonnell and Weatherford (2013) acknowledged the commitment of all stakeholders to use research, and they concluded that much of the CCSS is firmly grounded in research on learning. Yet, they also identified significant gaps in research use because of political considerations as well as unavailability of appropriate research. Political considerations affected both the CCSS content and the process by which the CCSS were implemented.

This study is a useful example of how interviews can be highly targeted when exploring particular cases that have an explicit focus (the CCSS) and processes. That is, specific questions can be asked of different participants in the process, and their answers can be interpreted within a theoretical framework tuned to the particulars of the phenomenon under study.

Surveys

Surveys are self-report measures designed to assess knowledge, attitudes, values, and behaviors. Surveys are used to support claims about URE across some population(s) of interest. Those claims can be based on responses to single questions or on responses to multiple items that are part of a scale intended to measure a particular construct. Surveys are often given at multiple intervals to track changes in responses over time.

In this section, we also include self-described assessments, as they share many methodological characteristics of surveys (Moss & Haertel, 2016) and could be considered a subset of surveys that focuses on assessing knowledge or beliefs. In the case of URE, assessments are relatively indistinguishable from surveys in that they are used to estimate knowledge across groups of individuals. To date, assessments have not been used in the URE literature to provide individual scores or estimates of a theorized latent ability, which are common uses of assessments that arguably differentiate them from surveys.

Methodology

We focus on three methodological issues in survey design. The first issue is the design of the instrument itself. Regardless of the format or purpose of the survey, questions must be clear and avoid bias (Bradburn, Wansink, & Sudman, 2015; F. J. Fowler, 1995). An ambiguity in a focal concept can lead to results that are invalid; however, the researchers may not be aware of this ambiguity. This is especially true if the ambiguity or misunderstanding of a term is systematically based on a group characteristic of the survey takers. In Coburn and Talbert (2006), which is discussed in the section on interview methodology, district staff disproportionately viewed "evidence validity" in terms of the measurement properties and school-level staff viewed it in terms of how useful it was in the classroom. Thus, asking a question about "evidence" without context could have led to different answers based on interviewees' understanding of the term. Ultimately, "a good question is one that produces answers that are reliable and valid measures of something we want to describe" (Fowler, p. 2). Another important aspect of instrument design is that the survey is sufficiently concise so that respondents will be willing to complete the survey within a reasonable time span.

In our review of surveys of URE, we find more attention to clarity is needed regarding what questions are being asked about URE. Respondents need to be clear about what they are responding to when they are asked questions about potentially ambiguous concepts such as research, research use, evidence, and validity. Further, given the high levels of non-response to surveys in general, and URE surveys specifically, conciseness is also a highly relevant consideration.

Researchers should be clear on the purposes of the structure of the survey and its constituent parts. Sometimes, surveys (or parts) are designed to measure a latent trait (e.g., attitudes). In this case, it is important to provide analytic support that demonstrates that the items designed to measure a particular attitude can be empirically supported to function as a scale. Techniques such as factor analysis are often appropriate when there is a need to justify the validity of survey scales.

In many cases, though, the purpose of the survey is only to measure the frequency of behaviors. In this case, there is no need to explore relationships among items to defend the design of the survey. The only design issue is whether the questions adequately elicit from respondents the desired information.

A second aspect of URE survey design that we wish to highlight concerns the degree to which it is designed to be generalizable across contexts. Because URE is inherently interdisciplinary and the concept of research has so many interpretations in different contexts, many surveys of URE are quite specific to the particular case or cases under investigation and are not intended to be reusable in studies of other contexts; for instance, a survey of how nurses use research to inform patient care is not likely to be usable for teachers using research to inform lesson planning. The reporting of such surveys is rarely described in much detail, and the questions for these surveys are typically tuned to the specific issues associated with a particular study.

In contrast, there are surveys that are intended to be generalizable to different settings. Such surveys are highly useful because of their ability to meaningfully compare results across studies and research contexts and to potentially make inferences about differences in responses. Thus, the design and validation of these surveys are more likely to be published in scholarly journals.

The third methodological issue is sampling, which is always a primary consideration in survey-based research. In most uses of surveys in the social sciences, a sample of individuals is surveyed in order to make inferences about the broader population. In order for these inferences to be valid, the sample must be representative of any

systematic variations in the population(s), and there must be sufficient sample sizes and subsequent response rates to preserve that representativeness. There are multiple sampling strategies and highly sophisticated methodologies designed to evaluate the validity of population estimates (e.g., Lohr, 2010).

Recruiting and administering surveys to a representative sample is often difficult. Even if the researchers are confident that they are sending the survey out to a representative sample of all salient characteristics of the target population, participants often fail to respond for systematic reasons that distort the final set of results. Authors of URE studies often, and appropriately, raise caveats about the validity of inferences based on the sampling strategies and response rates of much of the existing literature in URE.

Accordingly, we cite the Penuel et al. (2017) example, which represents a very strong sampling design. Nonetheless, because of low response rates at organizational and individual levels, any generalizations of estimates to a larger population are still problematic.

Survey data are typically reported in terms of response frequencies for particular questions or score distributions on particular scales. The mean and standard deviation are usually reported. In addition, URE surveys are often disaggregated by groups of interest, such as by organization, hierarchical level, or role. Researchers may include multiple items to assess the same concept; doing so can help to establish that the participants are interpreting the questions as intended. Additionally, researchers may hypothesize that two items will be highly related: for example, are individuals who have particular past experiences with URE more or less likely to have certain attitudes toward using research for new policy contexts? Survey reporting also often includes bivariate distributions and/or correlations of pairs of questions to verify that similar questions are highly correlated or to investigate item relationships.

For introductory guidance on surveys, the following may be useful: Bradburn et al. (2015); F. J. Fowler (1995); Glasow (2005); Groves et al. (2009); and Visser, Krosnick, and Lavrakas (2000).

Threats to Valid Interpretation

As with interviews, surveys share general risks associated with interpreting self-report data. Responses about behaviors may not align with observations of those behaviors, and social desirability may influence responses about values and attitudes.

Sampling issues are always important. If the sample is not representative of the population of the research context, inferences may be inappropriate, even at the case level. Non-response to surveys is always an issue, especially when those who choose not to respond are different in some way, either known or unknown, from those who participate in the survey. Researchers should do what they can to evaluate differences between participants and non-participants. Researchers need to be clear not to generalize beyond the population that the study design can support. They also need to report how limitations in the sample might affect the study's interpretations.

Survey designs can be problematic in several ways. Scales may not have empirical support; thus, interpretations based on scales may be inappropriate. Questions may not be well formed and lead to biased responses or responses that are difficult to interpret.

Questions Surveys Can Address

The strength of surveys is that they can obtain specific responses from a large distribution of participants quickly and efficiently. Surveys are also useful in revealing the distribution of responses for particular questions or scales within the survey. They can cover a relatively broad set of questions that investigates beliefs, practices, and experiences associated with URE, including:

- What do you believe about some aspect of URE?
- What have you experienced with respect to some aspect of URE?
- What practices related to URE have you engaged with?
- What do you know or understand about some aspect of URE?

Questions Complementary Methods Can Address

Surveys are generally not adequate for purely exploratory study or to investigate poorly understood processes. Interviews and observations, as previously discussed, can reveal those aspects of URE in which surveys would fall short. Examples of complementary uses of surveys and interviews within single studies are presented in the mixed-method and case study section. Surveys are well suited to support other kinds of methodological approaches, including quasi-experimental and social network analyses; however, the survey must be designed explicitly to provide data to support such analyses.

Example: Using Surveys to Study URE in Education – Penuel et al. (2017).

Penuel et al. (2017) conducted a large-scale survey of school districts and leaders to understand how they accessed, perceived, and used research. The survey was designed to create a baseline of information of URE in education across the United States. The particular strengths of the study lie in its use of representative sampling and a theoretical framework of research use that builds on modern conceptions of URE.

The study was designed to address three questions:

1. How frequently do school and district leaders report that they use research—and for what purposes?;
2. Where do school and district leaders access research?; and
3. What individual and organizational characteristics are associated with research use? (Penuel et al., 2017, p. 2)

The conceptual model influencing the survey design was based on three lines of research. First, Weiss and Bucuvalas (1980) articulated that research is used by policymakers for multiple purposes:

- *Instrumental use*: directly informs the substance of particular decisions;
- *Conceptual use*: changes ways that an individual views a problem or problem space;
- *Symbolic (or political) use*: validates or justifies a preferred position; and
- *Imposed use*: mandated by law or policy.

Second, research use is influenced to a large extent by the individual characteristics of users, including their attitudes toward research and their individual preparation and training to use research (Landry, Lamari, & Amara, 2003). Finally, organizational characteristics and support for the use of research also affect individual URE (Penuel et al., 2017).

These findings about research use influenced both the design of the survey and the study's predictions as to what factors would influence how and why school district leaders use research. Eight scales were used to describe URE—three for the type of research use, three for attitudes toward research, one for the effort that individuals used to acquire research, and one to capture the organizational culture of research use.

The survey was then designed by the research team with input from education leaders and scholars, followed by cognitive interviews with 40 education leaders. All of these processes led to a revised instrument that was then piloted with 265 education leaders sampled from across the country. Following piloting, additional items were added to improve scale reliability, and further revisions were made to improve items.

The target population for the study was district office instructional leaders and school principals from mid- and large-sized urban districts in the United States (student population > 9,000). This resulted in an overall sampling frame of 1,000 districts and more than 41,000 school and district leaders. The research team then developed a stratified random sample that ensured representation of different occupational categories (e.g., principals, supervisors, coordinators, superintendent's office leadership) and district size. They also randomly sampled administrators within districts once a district was selected. They then sent the survey out to 733 administrators from 487 districts and had a total response rate of 51.5%.

The research team performed checks to evaluate differences between respondents and non-respondents and also included sampling weights to adjust for participation of individuals from districts of different sizes. They reported on reliability of each of the scales and then described patterns within the scales as well as correlations between scales.

Major findings included the fact that, as predicted, school leaders use research for a variety of purposes, most often instrumental (e.g., *How should a reading program be designed?*). They also found that individuals access research from many sources, most frequently through professional networks. Finally, they reported on relationships between individual characteristics and research use and even stronger relationships between organizational characteristics and research use.

While research findings were, in many ways, consistent with other studies of URE in education, there were also some departures. For example, apart from conceptual use, there was a lack of a relationship between attitudes regarding credibility of research and self-report of using research for instrumental, symbolic, or imposed uses. The researchers raised the possibility that the self-reports of surveys may not always accurately produce the same findings that are developed out of direct observation, for example. They also discussed other limitations with respect to interpreting how school leaders actually use research and strongly advocated the use of multiple methods to better understand and facilitate the use of research by school leaders.

Given a well-designed instrument and sampling approach, broad surveys like the one described by Penuel et al. (2017) can provide unique information about certain aspects of URE, most especially perceptions about URE, across large populations. However, they can never tell us about actual URE without triangulation with other methods, and they cannot provide the kind of detailed insights on URE that are provided through more intensive studies of particular cases.

Example: Using Assessments to Study URE in Health and Social Service Settings – Palinkas et al. (2016).

An assessment example is the *Structured Interview for Evidence Use* (SIEU; Palinkas et al., 2016), an instrument designed to be applicable to a broad range of social service settings. The work grew out of an interest in the extent to which research-supported treatments (RSTs) were used in clinical settings to address mental health and behavior problems among children and adolescents.

Palinkas et al. (2016) set out to develop an instrument that could be efficiently administered and scored to monitor the use of research evidence when implementing RSTs. To do this, and to develop their domain model, they surveyed the literature on URE and focused on three domains that included how research evidence is acquired (input), evaluated (processing), and used in context (output).

Assessment scales of 20 items each were developed to measure each of these domains. All items for each scale required responses on a 5-point Likert scale. For example, research input items asked participants to respond on a scale of 1 (not at all) to 5 (all the time) to indicate how much they relied on different sources (e.g., academic journals, intermediary organizations, the Internet) to obtain information about a particular program or intervention. The other two scales—how important each listed criterion is to evaluate the quality of research evidence (processing); how important each characteristic is when deciding whether or not to adopt a new program or practice (output)—had item responses anchored from 1 (not important at all) to 5 (very important). Within each of the three scales, subscales of items were hypothesized.

The SIEU was administered to three different samples across a range of child health and behavioral settings in order to evaluate the appropriateness of the instrument. The data were then aggregated across the samples. Internal reliability analyses were examined in two ways. First, a series of factor analyses was carried out. No clear factor structure emerged when all 60 items were included. Three additional factor analyses provided empirical support for the theoretical structure of the subscales of each scale. After revisions, the final instrument included 45 items across three scales. To estimate reliability for each scale (as well as subscales) of the revised measure, a measure of internal reliability (Cronbach's α) was calculated.

Finally, all participants in the study also completed two other measures. The first measured health service providers' attitudes toward adoption of evidence-based practices (EBPA). The researchers hypothesized that scores on the EBPA would be positively related to SIEU scores. A second instrument was a measure of the quality of organizational social context (OSC), but it did not examine anything about research-based practices. Palinkas et al. (2016) hypothesized that OSC scores would not be related to SIEU scores. Findings supported both hypotheses.

Thus, this study provided initial support for the use of the SIEU across a broad range of settings. The authors appropriately noted some limitations in their work and the need to pursue further validation efforts.

Example: Using Surveys to Study URE in Child Welfare – Wulczyn et al. (2015).

Wulczyn et al. (2015) used a survey to investigate the extent to which practitioners in child welfare agencies accessed research evidence in making decisions in their work with children and families. They then planned to use the survey results to examine two questions:

1. Do some individuals working in child welfare organizations report accessing research evidence more routinely than others? If so, what characteristics are associated with users of research evidence?
2. Do organizations with staff who report accessing research evidence achieve better outcomes? (p. 150)

The researchers began with a theoretical framework of URE describing interactions among three levels: person, agency, and operating context. The first, person-level, includes the people or practitioners in the organization and their personal skills, attitudes, beliefs, and participation with respect to URE. The second, agency-level, is conceived of as an aggregation of the person-level URE and the size and structure, climate, culture, and leadership within the organization. The third, operating-context, represents the larger organization that consists of multiple agencies and includes the business environment, poverty rates, system size, and urbanity. The model is used to help explain how individual URE practices are influenced by all three interacting levels. At each level, URE is considered in terms of three processes: acquiring, processing, and applying research evidence (e.g., Tseng, 2012).

The researchers sought participation from 49 agencies in the state; 26 agreed to participate. The sampling plan attempted to achieve a census as surveys were sent to all caseworkers, supervisors, mid-level managers, and executives (n = 947). Approximately half of these individuals responded, though the participation rate varied substantially across employee group and across agencies.

The survey instrument derived from two sources. First, the researchers included the aforementioned SIEU, a general measure of URE that has been studied and evaluated in multiple contexts (Palinkas et al., 2016). These questions were complemented by a set of questions designed to focus on the acquisition of URE at specific points during professional practice. In addition, one representative from each agency provided contextual information about agency and organizational factors.

The study first reported the means and distributions of specific research questions. Then the researchers examined person-level characteristics associated with acquisition processes of URE and found that workers with particular attitudes, experience, and know-how were differentially likely to engage in URE.

The study then used data in the survey to predict the likelihood of desired outcomes— in this case, having children and youth placed in families outside of foster care. They found that agencies with staff who reported higher URE had better outcomes. These improved outcomes were attributed to URE, even after controlling for individual characteristics of practitioners (attitudes, experience, and know-how) and characteristics of the agencies.

The researchers recognized certain limitations of the study, including response rates, limited information on what workers actually did with children and families, and the focus being only on the acquisition dimension of URE. Yet, this study demonstrates the potential of using survey data to explore a range of questions of interest relevant to URE.

Brennan et al. (2017) reported on an assessment instrument called SEER (Seeking, Engaging with and Evaluating Research), which was designed to support work in the health policy area but could also be applied in a much broader range of policy settings. The instrument focuses on the capacity of individuals and institutions to make use of research in policy development and implementation.

The study team developed a complex model of URE in health policy known as SPIRIT (Supporting Policy in Health with Research: An Intervention Trial), which includes as a prerequisite that organizations and staff have the capacity, or predisposing factors, to make use of research (Redman et al., 2015). The capacity domain has four dimensions: the value an individual places on using research; the confidence an individual has in his/her knowledge and skills for research engagement actions and use; the value the organization places on research; and the tools and systems the organization has to support research engagement actions and use. In addition to the capacity measure, SEER also includes scales that ask individuals to respond to levels of research engagement (e.g., *Did individuals access, appraise, and/or generate research?*, etc.) and nature of research use (e.g., *Was research used to advance conceptual understanding, persuade, or inform tactics and actions?*, etc.).

Items were presented using 4- or 5-point Likert scales ranging from 1 (not at all valuable) to 4 or 5 (very confident). The assessment was completed by 150 people across 12 agencies. A subset of these initial participants then took the assessment a second time to establish test-retest reliability.

Brennan et al. (2017) conducted reliability analyses and factor analyses to develop evidence in support of the SEER scores. They reported measures of consistency for each of the individual scales. Scores were relatively consistent across the two assessment windows for most, but not all, scales.

Finally, the researchers investigated the relationship between the SEER scores and another measure that was selected because it was thought to be substantively related to SEER. This measure, TPB (Theory of Planned Behavior), asks questions about individuals' plans and intentions for using research. Brennan et al. (2017) found positive relationships between these measures, though the relationships were not always as strong as they hypothesized.

The work of Brennan et al. (2017) presents an initial validation of an assessment designed to measure research capacity across a broad range of settings. Although this tool was developed in a health policy context, the structure of the domain and the items would seem to readily transfer to other contexts. As the authors noted, further validation efforts are necessary to fully understand the applicability and appropriateness of the instrument.

Observation and Discourse Analysis

An important method for studying URE is to observe policy or practice settings where URE would be relevant. In this section, we consider both observation and discourse analysis together because they fundamentally consider the observed (or recorded) behaviors and statements of individuals with a minimum of participation of the research staff. The most pervasive evidence attended to in observations is discourse in contexts related to policy and practice. Observations include consideration of all aspects available to perceive: the physical actions of individuals, the words they speak and the tone in which they speak, and the physical environment in which they operate. Discourse most often refers to oral or written communications but can include non-verbal gestures, graphical representations (e.g., a diagram or slide), or any other modality people use to communicate with each other.

Most observation methods come from ethnographic and anthropologic traditions (e.g., Eisenhart, 2001; Erickson, 1986; Geertz, 1973) in which an observer attempts to interpret actions and interactions of participants within some social or cultural context. There are other methods that come from more positivist traditions and are used less frequently in URE research.

Gee (2011) viewed discourse analysis as a method of understanding three uses of language: informing, action, and identity. Researchers will typically ask: What information about research is being communicated (informing)?; For what purpose is particular language about URE being used (action)?; In what roles and to which audiences is language being communicated (identity)? Thus, discourse analysis not only provides insight into the research being used in the development of policy and practice, but it also sheds light on the dynamics of URE within a particular social context. More critical forms of discourse analysis (see, for instance, Chouliaraki & Fairclough, 1999; Fairclough, 2013) can provide structured insight into power relationships and audience framing that surface through language dynamics among individuals and entities.

Methodology

First and foremost, observation is an interpretive act, not a simple recording of behaviors. Geertz (1973) cited Gilbert Ryle's distinction between behavior and action by contrasting the difference between a twitch and a wink. While the physical action is the same, the influence and intent of each behavior are quite different. The observer must consider information beyond the behavior itself in order to make an interpretation of the meaning of the physical action.

In the same way, observers studying URE are obligated not only to record specific comments made by policymakers or practitioners but also to make interpretations of their observations. In addition, the process(es) by which such interpretations are made must be clear and communicated in the reporting of the research. In some studies, formal methods of discourse analysis have been applied (e.g., Gee, 2011).

There is no single observational or discourse analysis method for making sense of a particular context. However, for important reasons, URE research tends to employ certain approaches more than others. Much observational research follows in the tradition of Geertz's (1973) conception of *thick description* in which the observer provides sufficient context to field observations in order for researchers and audiences to make sense of specific actions.

In using this kind of descriptive approach, the researcher comes to the context with a theoretical perspective to guide the observations—where to focus attention, important signals of evidence, guides for interpreting actions and interactions, etc. However, observing the context with a relatively open, descriptive approach is likely to raise new insights, new theoretical perspectives, and new ways of considering the data.

This approach can be juxtaposed against more structured approaches that use tools such as observation guides and/or checklists to note the incidence and frequency of particular predetermined evidence. These kinds of approaches are more appropriate when a theory, like URE, and its associated behavioral markers can be pre-specified very clearly. This can be challenging because there are so many contexts and definitions of research use.

The role of the researcher varies with observational methods as well. One common approach is for the researcher to act as a participant observer in which the researcher is immersed in the context for an extended period of time, building strong relationships and participating with others in the culture. While this is common for observational research in the social sciences, we have not found many examples in URE studies. As many contexts focus on policymaking or expert practice, URE studies tend to use a naturalistic observation approach in which the researcher is an external observer who is not a member of the observed context. Nonetheless, it is critical that strong relationships are developed in order for the researcher to have access to the kinds of interactions that are necessary to observe. This is particularly true for many of the social welfare contexts of URE research in which decisions about confidential and sensitive issues are often made.

Another consideration in observational data collection concerns sampling approaches. In many URE contexts, the instances in which research is procured, digested, and used to make decisions rarely take place in well-defined and scheduled events. A researcher, and/or members of the research team, cannot be physically present to observe all participants in order to wait for such instances, so identifying the times in which observation occurs is critical. For some studies of policymaking, for example, there are critical meetings and exchanges, possibly occurring over long periods of time or in concentrated periods, that need to be observed. The researcher needs to have a sufficient understanding of the context in order to know when the critical points of observation are likely to occur and the foresight to be present to capture them.

For other study questions, it may be important to develop some representative sampling of targeted practices. For example, a study may investigate the extent to which health care or child welfare providers demonstrate URE when making particular decisions about individuals in their care. It is unlikely that the research team could shadow all participants during all work hours. Therefore, a responsive design may require sampling of providers within a context as well as sampling of occasions for each provider. Accordingly, the research questions, the data analysis, and the inferences that are drawn all need to account for this sampling.

Analysis of observations is dependent on the lens adopted by the study. If observations are targeted at identifying instances of URE, then corresponding analyses will focus only on events in which URE surfaces. If, however, observations are targeted at understanding the larger process involved in decisions about policy or practice, then a much larger set of observational data must be considered.

The execution of observations requires detailed recording of the observation, typically through field notes. Observers need to be highly trained to appropriately carry out this task. Field notes then must be analyzed and interpreted following guidelines established and communicated by the researchers. In cases where more structured observation protocols are used, the guidelines for these protocols and their analysis must be adhered to as well. As with any structured coding system, it is important to establish the reliability of measurement across observers. Alternatively, qualitative methods of open coding may also be applicable to observational studies, especially ones that are more exploratory. Please refer to the comments in the prior section on interviews for a more in-depth discussion about open coding.

For introductory guidance on observations, the following may be useful: Angrosino (2008); Eisenhart (2001); Erickson (1986); Fetterman (2010); and Geertz (1973). For introductory guidance on discourse analysis, the following may be useful: Fairclough (2013); Gee (2011); Johnstone (2017); and Morgan (2010).

Threats to Valid Interpretation

In trying to understand URE, perhaps the biggest caution is interpreting the absence of evidence. Because most observations are sampling particular occasions, the observations themselves will not reveal URE that occurs outside of those occasions. It is possible, as argued by Zwolsman, van Dijk, and de Waard (2013), that individuals may engage with URE in ways that are not visible through observation and the discourse in which they engage. Thus, it is important to complement the observation with other evidence that may reveal URE not apparent through the observations themselves. The sampling plan should be as robust as possible in capturing occasions and actions in which URE could play a role.

A second caution in observation concerns the interpretation of actions, interactions, and decisions. There are quality control processes appropriate for different observational methodologies; all of these processes help to ensure that interpretations are transparent, defensible, and not idiosyncratic to the researcher. Without such processes in place, it is possible to misinterpret particular observables, resulting in claims that either overestimate or underestimate URE. In many observational studies of URE, there is limited detail reported about coding structures and quality control processes. This lack of detail most likely reflects space limitations that are part of the publication process. Nevertheless, without such detail it is possible that the same set of observations would be interpreted differently by others with expertise in the domain and in the methodological approach.

Questions Observations and Discourse Analysis Can Address

Observations have the virtue of not being self-reports, so inferences are based on what people actually do, not on what they say or think they do. Observations are powerful tools for capturing deliberative processes of policymakers and practitioners alike. Observations are also very helpful in providing evidence of how URE changes and develops over time as policies and practices unfold. Specific questions that can be addressed include:

1. What information related to URE in policy and practice is communicated?
2. Which individuals participate, and how do they interact, in targeted contexts about URE in policy and practice?
3. What roles regarding URE in policy and practice do individuals take on?
4. What actions and/or decisions about URE in policy and practice are evident?

Questions Complementary Methods Can Address

For the inherent limits of observations already discussed, observations are often complemented by other methods that can provide richer context and fuller information. Interviews and document analysis are two qualitative methodologies that are often used in partnership with observations to support claims about the intentions and contextual influences associated with URE. Interviews help to provide insight into the intent of an action or into a participant's interpretation of the events to which he/she responded. Artifact analysis may provide insight into work produced outside of the observable context.

Example: Using Observation and Discourse Analysis to Study URE in Education – Asen, Gurke, Conners, Solomon, and Gumm (2013).

Asen et al. (2013) used discourse analysis to study school board deliberations in three school districts across the full academic year. Researchers attended, recorded, and took field notes of 160 meetings that ranged from large public meetings of the board to small committee meetings. Their interest was to understand differences in URE as a function of district characteristics.

The study adopted theoretical perspectives on public policy deliberations that included considerations of how analysis must attend to legislative deliberations as well as issues that arise within the sphere of public debate (e.g., Goodnight, 2012). From this perspective, public debate can constrain and influence deliberations within the formal policy body.

In deciding which deliberations to focus on in these meetings, they limited inclusion to only policy deliberations linked to future actions. That is, the records of observations that were coded did not include simple reports of policy, the various matters unrelated to policy that come before school boards, or exploratory conversations that did not imply future action.

Their coding focused on research- and non-research-based forms of evidence used by board members during deliberations. Researchers first identified occurrences of evidence use and found that research evidence was used much less frequently than other sources of evidence such as examples, experience, data, and testimony. However, they did identify that some districts were more likely to use research evidence than others.

Asen et al. (2013) then developed a conceptual framework to analyze instances of URE. They considered four factors as part of their analysis of URE. First, they evaluated the presentation of research, differentiating between vague general references (e.g., "research has shown") and references to specific studies, noting that both kinds of references could be persuasive in this context.

The second factor they reported on referred to the background and interests of the individual or party sharing the research. They noted, for example, that district superintendents varied in their experience with research and, therefore, differentially referred to and presented research as they took on the role of advocate.

Asen et al.'s (2013) third factor was audience. They found that the language used in presenting the research was quite dependent on audience; when the domain of interest was well understood by a committee, for example, less background orientation was needed in the presentation and framing of research.

The final factor attended to was the surrounding context of the deliberations. Local and national contexts (e.g., No Child Left Behind [NCLB]) shaped the presentation of research during board deliberations. Research was more likely to be referenced when there was a perceived connection between the research and local policy issues. Further, the deliberations revealed how important the personal values held by participants were in influencing the deliberations and how research was used.

Example: Using Observation and Discourse Analysis to Study URE in Education – Farley-Ripple (2012).

In the URE literature, observation is a common method most frequently used in conjunction with other methods, such as interviews or document analysis. Farley-Ripple (2012) used observations together with interviews and documents to consider the following questions:

- How does research evidence inform central office curricular and instructional decisions in instrumental, conceptual, political, and symbolic ways?
- What research evidence is used and valued in central office decision-making?
- What factors shape research use in central office decision-making? (p. 790)

The study examined how research was used with respect to three instructionally-related policy decisions that the district was grappling with at the time. The first decision addressed the overhaul of professional development. The second concerned a high school textbook adoption. The final decision involved the revision of the school improvement planning process.

Farley-Ripple (2012) separated research use from other forms of evidence (such as data and working knowledge) that served as alternative sources for decision makers. Her conceptual framework referenced a typology of research use that employs four distinct purposes (Weiss & Bucuvalas, 1980):[2]

1. *Instrumental use*: directly informs the substance of particular decisions;
2. *Conceptual use*: changes ways that an individual views a problem or problem space;
3. *Symbolic (or political) use*: validates or justifies a preferred position; and
4. *Imposed use*: mandated by law or policy.

Farley-Ripple (2012) used an embedded case study design that included 34 observations of central office meetings resulting in extensive field notes, 20 semi-structured interviews, and supplemental document analysis to better understand the context of the school district. The study tried to understand research use by focusing on three factors that had been identified in the literature as being highly relevant to URE: the characteristics of evidence, the characteristics of organizational context, and the characteristics of decision makers.

To address the first research question, Farley-Ripple (2012) prioritized data that would help provide the most robust evidence of URE for each policy decision. She selected data for further analysis that best satisfied four criteria:

1. The timing of the decision and the ability to follow the decision from beginning to completion;
2. The ability to triangulate data about the decision across multiple participants and/or observations;
3. The decisions' potential to impact teaching and learning; and
4. The extent to which decisions were a response to accountability policy and status. (p. 791)

Following this selection, data were coded iteratively following the traditions of qualitative coding and cross-case synthesis (Yin, 2014).

---

2   See the survey example of Penuel et al. (2017), which uses the same conceptual framework to guide their research.

Observations took a prominent role in the reporting of findings and the analysis of the study as they allowed Farley-Ripple (2012) to demonstrate actions that were taken over the course of the decision initiatives. Regarding the analysis of the frequency of each type of research use, the article presents illustrative examples collected from the observations. She found instrumental use of research was generally absent: It was cited by one participant in one decision initiative, but it was not cited in the other two. In one instance, Farley-Ripple observed that the committee substantially disregarded research-related criteria that were part of its evaluation tool. Conceptual use was observed in only two instances—once when a participant discussed research on professional development in general and in another situation when a superintendent cited a particular study multiple times across several meetings.

In contrast, and as a valuable means of highlighting the complementary role of observations and self-report methods such as interviews, Farley-Ripple (2012) also provided substantially more examples of how participants responded in interviews when directly asked about how research changed their minds about an educational issue (conceptual use). Symbolic uses of research were primarily made by vendors and supportive administrators to promote their textbooks.

The second research question, which focused on the types of research resources that participants use or consider useful, led to a cataloging of types of research evidence and the frequency with which they were cited. The research used most often came from sources specifically oriented to professional practitioners. Academic publications intended for scholarly audiences were much less likely to be used.

The significant factors that influenced research use were the relevance of the research (i.e., whether or not the research focused on current problems of practice in the district); the time required to process the research compared to more efficient avenues of obtaining information; resource constraints for the decision; and the organizational culture, especially with regard to information-sharing among administrators.

Farley-Ripple (2012) concluded that, for this case at least, districts did not make instrumental use of research; research was more typically used symbolically and conceptually. This led to her suggesting that procedures for identifying URE need to be nuanced because the evidence of use in particular decisions often can be difficult to discern. Additionally, she challenged researchers to publish their work in ways that are more likely to be digested by practitioners and also raised questions about how to identify the malleable factors in district offices that provide the best opportunities for engaging in change and improvement efforts.

Example: Using Observation and Discourse Analysis to Study URE in Health Policy – Zwolsman, van Dijk, and de Waard (2013).

Zwolsman et al. (2013) explored whether the use of specific evidence-based practices (EBPs) was visible during general practitioners' (GPs) consultations with patients. EBP (Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996) suggests that physicians should engage with their patients during these consultations on three dimensions: use of evidence, use of experience, and use of the patient's situation (Dawes et al., 2005).

This study is one in which the model and associated actions of URE were very well defined and, thus, the observational process was highly structured. Therefore, the researchers chose to use a checklist approach for observing consultations (see Figure 1). They observed 147 consultations by 34 GPs, half of whom were trainees and half of whom were experienced. For each consultation they coded interactions in terms of what the practitioner cited and said with respect to use of evidence, use of experience, and use of the patient's situation. For each of these dimensions there were specific phrases used by the GPs that would be seen as positive evidence of EBP.

Following each consultation, the researchers interviewed the GPs to obtain self-reports about how physicians were using EBPs. The study found that evidence of EBP, as observed through this protocol, was very rare. Physicians, regardless of experience, did not commonly communicate with their patients using practices that are recommended based on the research. They also observed that GPs were internally consistent in their consultative interactions with patients, regardless of the kind of patient or health problem. Though the GPs reported using EBPs to some extent in their private deliberations, even that evidence was limited. For example, GPs might have made general assertions about using research without being able to identify specific studies or sources of research information.

Zwolsman et al. (2013) described just a few differences between experienced GPs and trainees. During interactions, experienced GPs referred to their historical relationships with specific patients as well as their own clinical experience. Trainees searched for evidence more often than did experienced GPs.

This study shows that, considering only observable physician-patient interactions, there is very limited evidence of EBP. However, it is less clear the extent to which these factors are considered in physicians' practice more broadly. The highly structured approach taken here allows for very clear judgments about specific actions. However, if URE is happening in other ways, a commitment to such an approach without fully addressing the broader context may not reveal the full story.

USES EVIDENCE

Looks for evidence
- ❏ GP guidelines
- ❏ Pharmacotherapeutic guidelines
- ❏ Internet ▭
- ❏ Aggregated evidence ▭

Says
- ❏ The guidelines say...
- ❏ The evidence says...
- ❏ We know that...
- ❏ Our policy is...
- ❏ Other ▭

USES EXPERIENCE

Says
- ❏ In my experience...

USES PATIENT SITUATION

Asks
- ❏ What would you like ▭
- ❏ Otherwise ▭

Says
- ❏ Given your situation
- ❏ Given your history
- ❏ Otherwise: ▭

Figure 1. Coding form from Zwolsman et al. (2013).

Document Analysis

Document analysis focuses on the analysis of texts to determine the explanations and processes that occur over a distinct period of time through the interpretation of "mute evidence" (Hodder, 2000). The strength of document analysis is that it interprets physical artifacts, often written for public consumption. Depending on the source and use of the original documents, they may have a strong persuasive component as seen in URE studies that consider political briefs and other documents prepared by advocacy organizations and vendors.

Document analysis is particularly well suited to historical analysis. Process tracing (Collier, 2011; George & Bennett, 2005) is one approach that is appropriate to the use of this methodology, as it emphasizes analyzing "diagnostic evidence," such as text artifacts, that is developed over time and can be used to determine the causal mechanisms leading to a social outcome. Grounded theory (Corbin & Strauss, 2008) is another methodological tradition that is well suited to document analysis, as it iteratively builds a theoretical basis from the systematic investigation of documents.

There are two broad classes of URE studies that include document analysis. The first focuses on specific institutional documents such as formal legislative policies and subsequent administrative memos regarding those policies. Researchers can look for evidence of URE in the documents themselves and can also trace the way in which the use of or reference to research changes over time within the institutional body or the specific document.

The second class of studies selects a much broader set of documents that enables a researcher to better understand social processes and interactions relevant to URE. The textual artifacts in these studies may include memos, emails, meeting minutes, administrative materials, and other informal sources; the researcher examines these artifacts to try to make sense of how research was considered and used in the context of policy and practice.

While document analysis is often a primary method for historical research in which evidence of social interactions is unavailable or not trustworthy, it is also used in contemporary qualitative methodologies, including studies of URE, as a complementary approach to triangulate or verify other evidence sources. In case study research, for example, document analysis often complements interviews, participant observations, and surveys to develop a fuller picture of how research is being used.

Methodology

O'Leary (2004) presents an accessible overview of the major steps for document analysis, which include planning, gathering, reviewing, interrogating, reflecting and refining, and analyzing data. Each step must be considered in light of the goals of a specific URE study.

In studying URE, the bulk of document analysis will focus on the textual content, and researchers will likely use several methodological perspectives, including process tracing (George & Bennett, 2005), qualitative comparative analysis (Rihoux, 2006), and grounded theory (Corbin & Strauss, 2008) for case studies or general qualitative analyses. In many cases, and especially in more exploratory studies in which there is not an established theoretical framework, the documents and content to be explored are classified and coded iteratively with a general theoretical model that incorporates elements of URE such as how arguments are made or how positions are framed. Relevant passages are identified and then organized into themes, which are then analyzed collectively to determine ways in which research is used in the texts and, when there is a longitudinal component to the analysis, how the texts demonstrate that the use of research changes over time.

For introductory guidance on document analysis, the following may be useful: Bowen (2009); Hodder (2000); and O'Leary (2004).

Threats to Valid Interpretation

Documents are artifacts that memorialize the act of some party recording the conditions, events, or outcomes of otherwise social phenomena. Such documents may only represent certain perspectives and, in the case of analysis of legislation or policy, are prone to losing the viewpoints that were not preserved in the final enactment of policy or regulation. Further, textual artifacts rarely include evidence or direct observation of the social processes that led to the production of the texts. The challenges of interpreting connotation may be less extreme than in interviews, observations, and discourse analysis but still provide a potential for misinterpretation that can lead to incorrect inference.

When the evidence focus for document analysis research is on informal texts, it may be difficult to retrieve and analyze all relevant documents. In these cases, researchers must ensure that they have reached saturation (Saunders et al., 2018) of the textual evidence to increase the likelihood that their analysis is not missing significant elements.

Questions Document Analysis Can Address

Document analysis is particularly useful to explore questions about URE such as:

- To what extent is URE represented in the process or outcomes of policy and/or practice development and implementation?
- To what extent does documentary evidence converge with other evidence about URE acquired through other qualitative methods?
- To what extent do artifacts demonstrate that URE changes over time in regard to one focal organization or context?

Questions Complementary Methods Can Address

Methods such as interviews and observations can provide valuable social context and complementary data about the text that provides deeper insight into context, motivations, and interactions that may have led to the way in which URE is memorialized in textual data. Interviews in which the document artifacts are presented can allow the researcher to elicit the participant's recollection or interpretation of the events leading to the production of the document or may trigger specific memories that the participant may not have remembered without additional help; such combinations can add valuable context that neither method would collect on its own. Similarly, observational data used alongside document analysis can help to understand the underlying social processes that led to the final decisions for the artifacts produced.

Example: Using Document Analysis to Study URE in the Justice System – Garces, Marin, and Horn (2017).

Garces et al. (2017) looked at the use of non-legal sources (e.g., news articles, journal articles, books) and *amicus curiae* briefs by the U.S. Supreme Court in deciding two related cases regarding the consideration of race in the college admission process (*Fisher* v. *University of Texas*). The University of Texas at Austin used race as one of several criteria for its admission process, and that use was challenged as a violation of the Equal Protection clause of the 14th Amendment of the U.S. Constitution. The Supreme Court agreed to hear the case in 2012 and, in a 7-1 vote, decided to remand the case back to the federal court to ensure that the lower court's decision satisfied particular Constitutional requirements for strict scrutiny. The lower court reaffirmed its decision, which the plaintiff appealed, and the Supreme Court agreed to rehear it. In 2016, the Court ruled in a 4-3 decision to uphold the admissions process. Both cases received a great deal of media attention and were the subject of intensive political discussion.

The authors argued that *amicus curiae* briefs often use social science research and findings to justify arguments in support of one of the parties in a case. They pointed out that the use of briefs in judicial opinions has been contested for a number of reasons, including that research findings can be inconsistent across studies or can evolve over time. Despite concerns about the reliability of research interpretations in *amicus curiae* briefs as well as other non-legal sources, the Court continues to rely on these briefs as it sorts out arguments, makes its decisions, and writes its opinions. The study looked at the Court's opinions in the two cases—the majority opinions (both written by Justice Anthony Kennedy) as well as concurring and dissenting opinions—as the context for evidence use. Garces et al. (2017) focused on the "use" of non-legal sources and *amicus curiae* briefs as "both explicit citation as well as instances in which the Justices' conclusions reflected the conclusions or findings of a non-legal source" (p. 172). They considered when the uses were consistent with and contrasted with conclusions from social science research. The study did not consider cases in which a brief or source was *not* used, as the methodology does not support such inferences. They noted that there were multiple avenues for use that the document analysis method could not capture. It is possible that sources are not explicitly cited because the evidence was reviewed and found to be similar to sources that were explicitly used; similarly, it is possible that the Justices were exposed to and considered briefs but did not include them in their written opinions.

Both cases elicited quite varying opinions, even in the initial case, which only had one dissent from the majority opinion. Reading the opinions, the researchers found that the Justices at times cited the same sources (non-legal and *amicus curiae)* but made very different interpretations of central arguments, including the manifestation of race and classifications on the basis of race. For example, they found that Justice Kennedy and

Justice Alito had contrasting understandings of the use of race, but they cited the same non-legal sources to reach their contradictory conclusions.

In other cases, Justices would simply attend to and cite different sources in developing their opinions. For example, the Justices had different interpretations of what the Equal Protection clause meant, which led them to cite substantively different non-legal sources to support their opinions. Justices Kennedy and Ginsburg drew on evidence related to whether the method of classification conducted in the admission policy constituted racial discrimination. In the majority opinion, Kennedy reviewed and analyzed briefs, university documentation, and details of the development of the admission policy to determine the reasons for using race in admissions decisions, how the policy related to existing evidence about different types of admission criteria, and the outcomes. The authors highlighted the briefs that Justice Kennedy used in his opinion regarding these topics and then analyzed how they related to dominant social science research. They found that the conclusions that Kennedy drew often reflected social science research findings, even when he did not directly cite them.

In contrast, Justices Thomas and Alito interpreted the Equal Protection clause to mean that any classification based on race equates to racial discrimination, regardless of any benefits to diversity that are achieved. Therefore, Thomas and Alito considered as irrelevant most of the non-legal sources that Justices Kennedy and Ginsburg cited. Thomas instead cited non-legal sources that provided contextual legal support for the interpretation of the Equal Protection clause consistent with his perspective. Additionally, Thomas cited sources that referred to the "mismatch theory" arguing that research consistently found students with lower test scores than the general admission criteria performed more poorly in school. In contrast, Garces et al. (2017) noted that this theory has been contradicted by numerous social science research studies and that several of these studies were cited in briefs in support of the University of Texas at Austin. Justice Alito indicated that he would like to see data and statistics about how the admission policy affected students in the school and noted that the lack of such items added to the evidence register was problematic. He further questioned the University's motives in pursuing a race-conscious admission policy, which was not based on submitted evidence.

The authors, therefore, found that the use of research and evidence cited in the *amicus curiae* briefs by the Supreme Court Justices in this case was not consistently interpreted or accepted at their face value. All of the Justices cited some briefs as evidence of consideration of research findings. However, in some cases, Justices used the lack of evidence to contradict the evidence cited or ignored briefs that cited studies that challenged the theories they wanted to support. In short, the authors contended that their study revealed that the Justices' ideological positions shaped whether research was deemed relevant for their opinions.

Yanovitzky and Weber proposed and applied a methodology to analyze the content of
documents used in the legislative process. The specific example they used focused on
childhood obesity, but the proposed methodology represented a generalizable approach.

Yanovitzky and Weber developed their content analysis methodology using a theoretical
framework of persuasion and argumentation, drawing on the work of Majone (1989),
Toulmin (2003), and others. A critical point they made is that research, in addition to
other information, is primarily used in public policymaking to argue for and persuade
others to support or reject particular positions. Thus, they were very interested in the
argument structures that are embodied in legislative documents. They also contended
that it is possible to reliably infer legislative motivations through such content analysis.
In this respect, how evidence is incorporated into arguments can be quite telling of how
evidence is used in each instance. In general, political actors use evidence (research and
non-research) to support three types of arguments: 1) documentation (using evidence to
establish the current state of affairs or what is happening); 2) analysis (using evidence
to offer plausible explanations for why those things are happening); and 3) prescription
(using evidence to suggest what should be done to address problems that have been
described and explained). Thus, their method did not simply seek to locate instances
of research use that surfaced during the legislative process, but also attempted to infer
the particular ways in which research (and non-research) evidence was used to support
specific purposes of argumentation and persuasion. To clarify use of evidence in
relationship to goals, they also adopted the aforementioned Weiss and Bucuvalas (1980)
typology of research use.

In contrast to many other studies we have cited, Yanovitzky and Weber used a
structured coding model in which they established the codebook and then used it to
analyze the documents. The coding structure they adopted is shared in their article.
One set of variables addresses the mechanics of URE in the policymaking process
that enable the mapping of research evidence flow among actors (see article appendix
for a fuller description):

● Who (e.g., legislator, lobbyist) in the legislative process supplied the evidence?
● What was the type of evidence (e.g., statistical fact, study findings)?
● What was the source of the evidence (e.g., academic research, advocacy research)?
● What was the policymaking context in which the research was used (e.g., hearings,
   policy negotiation)?
● When was the research used (e.g., to establish a timeline)?

The second set of variables was designed to examine the thematic/rhetorical uses of
research, or the purpose for using research evidence:

● What was the goal of evidence use (e.g., problem identification, preferred
   policy solution)?
● What was the valence of evidence use (e.g., pro, con, two-sided, neutral)?
● What was the motivation for evidence use (e.g., instrumental, conceptual)?

To evaluate the tool, the researchers reviewed 14 years' worth of policy documents
relevant to childhood obesity-related legislation. They specifically tracked official
public documents that contain texts of bills, committee hearings and reports, floor
debates, and executive actions. They began by retrieving 1,888 potentially relevant
documents and culling that to 786 after screening for relevance.

The research team then needed to ensure that the coding of documents was reliable
across multiple coders who would examine the documents, identify excerpts, and

then code each excerpt using the coding structure. The documents were coded by 14 undergraduates who went through a training and coding process using a small number of documents. When the agreement during the first round of coding was not at a satisfactory level, coders received additional training and the codebook was modified to improve reliability. The research team later judged the reliability to be sufficient (Krippendorff α = .86; Hayes & Krippendorff, 2007) and then went on to code the full set of documents that included 4,684 excerpts in 224 congressional bills and 190 congressional hearings.

A dominant finding was that evidence cited for bills was quite different from evidence cited for hearings. While the evidence cited for bills was mostly limited to statistical facts, descriptions of research studies and expert opinions were commonly infused into hearings. While generic and government research most commonly appeared in bills, anecdotes were the most frequent type of evidence cited in hearings.

The motivation for using research in bills was almost always conceptual (i.e., attempting to influence how others think about the issue). In hearings, though, evidence was used to serve instrumental (i.e., informing choice among alternative courses of action) and tactical (or political) purposes. Finally, the goal of using evidence in bills was primarily to objectively describe the status of the problem and its cause (documentation); the goal of possible solutions (prescription) was also evident but to a much lesser degree. Hearings included possible solutions and preferred policy responses along with problem status and cause (i.e., analysis and prescription).

Yanovitzky and Weber also examined patterns longitudinally across different presidential administrations and different party make-up of Congress. Significant differences in both the number of documents and the type of use coincided with these different administrations. The content analysis tool was sensitive to capturing changes in the political landscape over this time period.

Study Design Methodologies

Social Network Analysis

A fundamental question explored in URE studies concerns the ways in which research evidence is conveyed from its source of production to the policymakers and practitioners who use it. Social network analysis (SNA) is a method used to identify and map relationship and information networks of entities. In most URE research that uses SNA, these entities are human actors or organizations (e.g., policymakers, administrators, schools, intermediary research institutes), and the relationships are tracked in a social landscape, which can be physical, conceptual, or virtual. The relationship connections ("ties") that are tracked in these studies range from the presence of a general social connection (e.g., a nurse indicates that she regularly talks with former classmates about new medical research) to the exchange of specific information or a specific tool (e.g., a principal receives information on a specific observation protocol from another principal).

Methodology

Social network analysis describes the structure of interaction networks and sometimes can provide additional descriptors about the network as a whole, about specific connection patterns, or about trajectories of information.

SNA is not a complete methodology so much as it is a method of analyzing data collected through other methods. To determine the ways in which research flows from producer to user, custom surveys are often used to determine self-reported connections.

SNA can also be used to analyze data from interviews, observations, and communication data such as email connections or raw data from virtual social networks such as Facebook or Twitter (which can be confusing, as SNA predates the existence of all of these online social networks).

For all of these sources of raw data, the SNA research extracts nodes (i.e., people, organizations, or other entities) and ties (i.e., evidence of a relationship, connection, or interaction). Some SNA also includes the strength of ties, which may be a raw count of the number of interactions from communication data, a reported frequency of interaction on a survey, or a value of an interaction stated by an interviewee. The nodes and ties create the structural network, and judgments can then be based on the entire network or on characteristics of individual nodes. Inferences at the network level, especially when the researcher seeks to compare characteristics of two networks, require that information about all related nodes is collected to ensure that metrics are commensurate. Therefore, the researcher must create well-defined boundaries of the network in order to identify the full set or sample of individuals to be included in the judgments.

Metrics

There are numerous measures that SNA researchers use to describe the nodes and the network. Analyses can focus on an entire network or on comparisons between different networks, or they can focus on characteristics of the ties to specific nodes. Analyses of sampled nodes are particularly useful when data are available only for a subset of individuals; the researcher wishes to compare characteristics of relationships between people who are in different networks; the network in focus does not have well-defined boundaries; or the research questions are focused on connections of individuals rather than on overall network characteristics.

Many measures focus on the centrality of particular nodes, which is a set of different measures describing the connections that a particular node has with other nodes in the network. In the context of URE, centrality frequently refers to the entities that other nodes in the network turn to for URE-related interactions.

Other SNA metrics describe broader characteristics. Density measures indicate how well connected members are to each other, clustering measures suggest the degree to which there are social cliques that are structurally isolated, and cohesion measures describe how easily the interconnection of the entire network might collapse if several highly central nodes were removed (e.g., if an assistant principal who assumed a role of managing all of the informational and procedural needs of a school quit to become a principal at another school).

In terms of URE, such methods help identify which actors and groups of actors interact, which member(s) in a network are viewed as central resources, the extent to which different subgroups within the network are interacting, and the extent to which the network changes over time.

For introductory guidance on SNA, the following may be useful: Frank (1996); Hanneman and Riddle (2005); Knoke and Yang (2008); and Wasserman and Faust (1994).

Threats to Valid Interpretation

SNA is especially susceptible to problems if the network of interest is not adequately sampled. Researchers need to be careful to include all relevant participants in the data collection. Even when considering measures looking at smaller samples of nodes, it is critical that all relevant connections to the selected nodes have been assessed. Each data source in SNA presents a particular risk to the accurate description of the network. For example, survey collection is particularly susceptible to participants deciding to skip to the next question before submitting all of their connections, leaving those participants' networks to appear invalidly small. For qualitative approaches, the research team may not have sufficient time to ask questions that exhaustively query about potential relationships, or participants may not feel inclined to list all of their connections to the survey administrator. The use of raw data or communication artifacts does not have these limitations but often is much more limited in terms of obtaining estimates of tie quality and magnitude.

Questions Social Network Analysis Can Address

SNA is particularly useful to explore questions about URE such as:

- Who are the primary brokers of research within a network, as defined by those who are the primary connections between people?
- What pathways connect policymakers to researchers investigating relevant policy contexts?
- To increase URE, which positions would be the best to target because they have the most direct connections to potential research users?
- How do network characteristics correlate with the frequency with which research is cited in policymaking meetings?

Questions Complementary Methods Can Address

SNA is fundamentally structural in nature; it provides information about the presence of a connection and little more. In URE research, SNA provides little information about how research is used and often cannot capture whether research evidence is actually shared across social connections. Other methods, including interviews and observations, are better suited to address such questions.

Example: Using Social Network Analysis to Study URE in Education – Finnigan, Daly, and Che (2013).

In education, Finnigan et al. (2013) examined how low-performing schools in one district made use of research evidence as they endeavored to implement school improvement reforms as part of a multi-year, multi-district study. They surveyed almost 300 teachers, school administrators, and district leaders by asking them to rate their level of interaction with all other individuals in the district on a 1 (no interaction) to 5 (1-2 times per week) scale. One set of questions focused on who connected them to research evidence they use in their practice while another set focused on affective relationships. The study focused on three different schools and included school and district leadership. Survey questions not only asked for information necessary to provide data to conduct the social network analysis but also asked for additional details about research-based evidence, including issues of access, research availability, research use, and opinions on its credibility. Finally, selected participants were interviewed, resulting in triangulation of the findings that substantial use of research evidence and commitment to improvement were correlated with individuals interacting around the use of research evidence.

The three schools revealed quite different levels of interaction around research use. In the first school, the resulting network was extremely sparse; there were very few connections between nodes. In the second school, the principal was the central source of research evidence, but there were relatively few interactions among teachers. In the final school, the principal again was the central source of information, but interactions among teachers were much richer, resulting in a high-density network with many connections among educators. The results regarding research use indicated that educators reported varying levels and types of research use, with a generally large percentage of respondents indicating that they use data and research evidence in some form. However, there was quite a bit of disagreement regarding what types of evidence were most credible: Sixty percent of school staff believed practitioner experience was more valuable than data, and 48% relied on strategies they knew to be effective without searching for evidence to support them.

At the district level, the SNA revealed several additional patterns. Central office staff tended to have the richest interconnections with each other, with small numbers of connections between the central office staff and the principals. There was a small number of central actors at the district level, most of whom were central office leaders. Principals were much more likely to interact with each other in terms of affective relationships than they were in terms of research use. The authors also found that the affective relationships, or the social bonds, between central administration and principals were even sparser than the work-related relationships. What this meant for the district was that the reform effort was relatively superficial and that ties between the district central office and the schools were very limited (Finnigan et al., 2013).

Example: Using Social Network Analysis to Study URE in Health Policy – Shearer, Dion, and Lavis (2014).

Shearer et al. (2014) studied how health policy networks exchanged evidence as part of health policy development in Burkina Faso in sub-Saharan Africa. Policies focused on community-integrated management of childhood illness, home management of malaria, and removal of user fees for antiretroviral treatment for human immunodeficiency virus (HIV). Researchers began by identifying two actors for each policy issue and then asking each to generate a list of others with whom they interacted during policy formation. Those listed were then approached and the same procedure ensued, resulting in a list of 101 unique actors. A subset (69) across the three policy initiatives was then interviewed. During the interviews individuals were asked whether they provided research to each of the other members on the list or whether these members requested that they provide research evidence. Separate networks for providing and requesting evidence were created.

Shearer et al. (2014) found that network density varied by policy initiative—there was more evidence exchange for childhood health, for example, than for HIV. Networks were also denser with regard to providing information rather than requesting information—providers shared information with multiple parties whereas requests for information tended to be directed at a very small set of network members.

Additionally, there were significant findings about evidence exchange at the individual level. Actors were more likely to provide information if they had already had information requested of them, and actors were more likely to share information if they already had a connection in common. Ties were also unidirectional: It was not common for actors to reciprocate either an evidence request or provision, which reflects the hierarchy of the organizations. In the malaria network, actors were more likely to request information from outside of their own organization, which differed from the other two policy areas. Finally, actors who actively provided evidence were also more likely to report actually using research evidence in their policymaking; however, actors who received requests for evidence were only statistically more likely to use evidence in the malaria policy area.

Experiments and Quasi-experiments

As with many areas of social science, one of the central questions in URE relates to the establishment of cause and effect: What factors cause an increase in URE, and does an increase in URE cause improvements on subsequent outcomes of interest? This section focuses on methods that are designed to support causal inference.

The optimal research method to establish a causal effect is an experiment (Shadish, Cook, & Campbell, 2001). In prior sections, we discussed how a range of qualitative methods is most appropriate to investigate the processes through which outcomes occur. In an experiment, the goal is to develop a research design that allows for measuring (in quantifiable terms) the average impact of an intervention, treatment, or event on an outcome in a way that can be compared to the impacts of alternative options. Doing this requires the ability to methodologically exclude (or account for) the potential impact of other factors. When we identify the effects of a phenomenon on an outcome and understand the mechanisms by which the effects occur, we can establish better guidelines for practices and interventions.

The fundamental method for ruling out alternative causal factors is through the process of random assignment to treatment and control conditions. In many URE contexts, however, it is simply not feasible to conduct true experiments because often there is either no reasonable way to randomly assign people or institutions to intervention conditions, or researchers want to use existing non-experimental data sets to evaluate causal relationships (e.g., using large-scale surveys). In addition, often, there are not sufficient numbers of individuals or organizations that can be randomly assigned to treatment and control groups in order to reliably interpret effects. For circumstances in which true experimentation is not impossible, researchers have developed a range of methodologies to support causal inference that are collectively identified as quasi-experiments (Shadish et al., 2001).

Quasi-experiments are used to explore the same kinds of questions as do experiments. As with experiments, the goal is to develop a research design that allows for measuring the average impact of a treatment or intervention on an outcome in a way that can be compared to other options and which excludes other factors that could be responsible for the observed outcomes.

Causal inferences on the basis of quasi-experimental studies are typically not as strong as they are for experiments. Quasi-experimental methods attempt to control for factors unrelated to the intervention, but the possibility still remains that alternative factors are not fully accounted for. However, as argued by Wagner,

Soumerai, Zhang, and Ross-Degnan (2002), certain well-executed quasi-experimental approaches can both overcome almost all threats to validity and provide for certain insights that most experiments do not.

Methodology

Because experiments and quasi-experiments focus on estimating causal effects, the outcomes need to be explicit and quantifiable. These studies are designed to support claims about a population even though only a sample of the population actually participates in the study. Research questions, hypotheses, treatments, and outcomes should all be informed by a theoretical framework.

The treatment, intervention, or condition under investigation needs to be sufficiently explicit that the boundaries are clear. Studies looking at causal effects should be easily replicable by other researchers to consider whether it applies to other contexts or to verify the effects. This is especially true in URE research because it crosses so many domains and contexts. Also, it is worthwhile to note that while we discuss experiments in terms of a treatment and control group, analyses that compare multiple conditions or interventions can be conducted, as Dobbins et al. (2009) does in one of the examples summarized at the end of this section.

It is also critical to identify the unit of study. In most studies of URE, individuals are members of a larger unit, such as school districts, departments, or institutions, that is fundamentally the unit of interest. In this case, sampling may be done at multiple levels, first developing a random or representative sample of the unit of interest and then sampling individuals within that unit. Appropriately considering the unit of analysis has implications for appropriately determining standard errors of measurement and the interpretation of results.

It is not coincidental that experimental and quasi-experimental studies of URE come from research contexts where the outcomes are quantifiable; both the research findings that are the basis of the intervention and the valued outcomes are relatively uncontested, and there is a strong theoretical link between research evidence and outcomes. URE studies of causal inference are more likely to be found in health policy and child welfare contexts than they are in education.

Experiments

In order to make causal inferences about a population, experiments include two procedures that need to be carefully executed. The first is drawing a sample that represents the population. This can be done by randomly selecting from the population of interest but is typically done by attempting to draw from as representative a sample of the population as possible. The larger the sample size, the better any estimates of population effects will be. For social science research, there have been significant concerns raised that this assumption can be satisfied. The samples selected for study are often unlikely to be representative of the population to which results are often generalized (e.g., Ercikan & Roth, 2014). With respect to URE contexts, it is almost certain that the studied sample is relatively unique because we find that processes through which research is used varies based on the field as well as the level of responsibility of the participants.

The second step is to randomly assign units of interest into treatment conditions (and control conditions). By randomizing the assignment, it is assumed that any differences among individuals or units will be randomly distributed across treatment groups and, therefore, will not be associated with any observed patterns in the experimental data. Again, the larger the number of units in a group, the more likely that differences between groups will even out and be accounted for.

Even with random assignment, though, there still may be differences in groups that occur by chance. This is especially a concern if there are high levels of attrition, as the participants who drop out of the study may have similar characteristics. Therefore, in comparing groups, researchers may also include covariates in the analysis that are observable and that might plausibly be related to outcomes. Including such covariates can lead to more precise estimates of treatment effects.

Quasi-experiments

Because quasi-experiments do not include random assignment, a broad range of methodological approaches has been developed in order to approximate the make-up of groups that would be achieved by randomization. Only with randomization is the assumption that groups are equivalent on both observable and unobservable variables viable. A number of quasi-experimental approaches attempt to match comparison groups on as many observable variables as possible, particularly on variables that might potentially affect the outcome.

One approach attempts to approximate random assignment by using covariates to statistically control for factors that can have a potential influence on outcomes and that are not equally distributed across treatment groups. The most common analytic approach is to use multiple regression models to estimate the treatment effect, taking into account any outcome variance associated with the covariates. Covariates may include initial assessments before the intervention begins as well as demographic and institutional factors. Even with large sets of covariates, however, covariate-based designs are still subject to significant concerns regarding the strength of causal inferences.

The second set of methods focuses on the construction of comparison groups from a larger population that are as equivalent as possible. Often, this entails determining the non-relevant variables that may still have independent effects on the outcome variables, matching subjects based on the difference in the focal characteristic but that are the same in all of the other variables, and then conducting the analysis on this matched sample. In comparing institutional response to research evidence, for example, URE researchers have attempted to create groups that have similar institutional profiles in terms of size, location, demographics of those affected by policy or practice, as well as scores on the outcome measure(s) of interest. More sophisticated approaches to matching, such as propensity score matching (Rosenbaum & Rubin, 1983), are not typically seen in studies of URE likely because of the relatively small number of units that are compared.

The third set of methods takes advantage of external rules or systems that lead to potentially random assignments. For example, if one child welfare agency in a city adopts a new method of case management and other agencies use a different approach, comparing the outcomes of children who are assigned to the first agency against the other(s) would approximate random assignment. Other examples include looking at children who score similarly but are assigned to interventions based on different local cutoffs, or examining students with the same academic profile who opt in to one elective compared to another.

Another example involves using time series approaches whereby data on the same group are collected over an extended period of time and in which the intervention occurs in the middle, which serves as a discontinuity between the status of the condition of interest before and after the intervention. Such studies have the problem that the group itself may change over time for reasons independent of the intervention, but there are analytic approaches that attempt to control for these instances (for example, see Wagner et al., 2002).

For introductory guidance on experiments and quasi-experiments, the following may be useful: Morgan and Winship (2007); Moss and Haertel (2016); Murnane and Willett (2011); National Research Council (2002); Schneider, Carnoy, Kilpatrick, Schmidt, and Shavelson (2007); Shadish et al. (2001); and White and Sabarwal (2014).

Threats to Valid Interpretation

Threats that exist for all methods designed to estimate causal effects that are most likely to pertain to URE studies include:

- Inadequate sample size that can both introduce error and preclude the detection of a statistically significant effect of a treatment that would have been apparent had a larger sample been used;
- Lack of clarity about the treatment conditions such that they are not readily replicated;
- Challenges to ensuring that any treatment is implemented as intended;
- Inadequate assignment to treatments that makes interpretation of effects uncertain;
- Attrition of the sample during the study, which may lead to comparisons of groups that no longer share the same characteristics of the groups determined by randomization; and
- Heterogeneity of results: While these methods estimate the average causal effect for an entire population, the same treatment may not work equally well across subgroups or institutions that have different characteristics (see Ercikan & Roth, 2014).

Additional threats for quasi-experimental studies exist and revolve around the likelihood that the attempt to approximate random assignment or to control for group differences is not sufficient. Replication studies with other samples and triangulation using alternative methods can all strengthen the confidence in causal claims.

Questions Experiments Can Address

Experiments and quasi-experiments are particularly useful to explore two kinds of questions about URE:

1. To what extent are specific interventions designed to increase the likelihood of URE having an effect?
2. Does URE lead to improved outcomes?

Questions Complementary Methods Can Address

While experiments and quasi-experiments can identify and quantify the effect of a particular treatment on URE outcomes, they cannot by themselves help explain the reasons for outcome effects. In addition to interpreting the results within the theory that led to the study design, methods that target the processes of URE can help to explain any observed effects.

Consequently, other methods are more suited to providing insight into causal mechanisms. While experimental and quasi-experimental methods may provide evidence that a particular treatment led to an improvement in specific metrics associated with URE, they cannot provide insight into why such improvement occurred without other evidence attained by using complementary methods. Qualitative methods such as interviews, observations, document analysis, and artifact analysis are likely to be useful complements to examine processes and mechanisms.

Example: Using Experiments to Study URE in Public Health – Dobbins et al. (2009).

Dobbins et al. (2009) conducted an experiment to try to understand which knowledge translation and exchange (KTE) strategies would be most effective in promoting URE by public health institutions. They focused specifically on the issue of childhood obesity.

The population of interest included all public health departments in Canada (n = 141). All were invited to participate; 108 agreed to do so. The fact that the sample represented a significant proportion (77%) of the population suggested the high likelihood (though not certainty) that any findings would generalize to the population. Within each health department, a single individual was selected as the responsible decision maker who would participate in the study and represent the health department.

The participating institutions were randomly assigned to one of three treatment groups that varied in the amount of interactivity between the research team and the health departments. The first group (health evidence; [HE]) was, in effect, the control group and was simply provided access to the already available government website that was a repository of all systematic reviews evaluating public health interventions. The second group (targeted message; [TM]) had access to the HE website but also received emails that included titles, full references, abstracts, and Internet links associated with systematic reviews in the area of promoting healthy body weight in children. Each week they received a TM that included a link to a specific review article, including an abstract and summary of the review. The third group included both the HE and TM information but also included a knowledge broker (KB) who worked with each department. Knowledge brokering activities included face-to-face and virtual interactions, site visits, and work on department-specific plans to develop individual and organizational capacity for using evidence in setting policy and practice.

At the beginning of the project, each participant completed a survey of organizational, environmental, and individual characteristics in order to establish a measure of research use culture prior to any intervention. The first outcome variable was *global evidence-informed decision-making* (GEID). Respondents were given a set of Likert scale questions asking about the extent to which research evidence was considered in making program decisions over the last year. The second outcome variable, *health policies and programs* (HPP), was used to characterize the proportion of research-based programs and policies being implemented by each department that were identified as relevant by the research team. Measures were given before the intervention began and several months after the intervention was completed.

For the GEID measure there was no significant main effect of treatment group assignment following the intervention. Results for the HPP measure were significant, however, as the TM group showed greater gains in URE on these specific policies. When the researchers accounted for initial capacity to make use of research evidence, they were able to develop a more nuanced understanding of what was happening. Departments that entered the study with a low research culture benefitted from both the TM and KB interventions on the HPP measure. However, departments that entered the study with a high research culture only benefitted from the TM intervention and actually showed a decline on the HPP measure when the KB treatment was introduced.

The non-uniform effect of knowledge brokering was not anticipated and led to consideration of potential factors that deserved further examination. First, the researchers also acknowledged potential limitations with respect to the sample and the decision that only one individual would represent each department. Second, within a large health department, any given individual might have deep familiarity with only a segment of the institution. Thus, the selected individual often had significant direct knowledge of some issues and much more limited knowledge of others. Third,

individuals in different roles often vary in the social desirability of their responses (Mosley, personal communication). Also, many individuals who represented each department changed over the course of the study, making judgments of change more challenging. The researchers also raised questions about their ability to account for the fidelity of the interventions, as well as the validity of outcome measures based on self-report. These are the kinds of issues that deserve consideration and evaluation in similar experimental studies.

Example: Using Experiments to Study URE in Public Health – Masset, Gaarder, Beynon, and Chapoy (2013).

Masset et al. (2013) asked whether the content of a policy brief has an effect on how people in developing countries respond to critical health issues. In this case, the policy issue presented involved the impact of food-based agricultural interventions on children's nutritional status.

The population of interest was adults in an unidentified group of developing countries. Invitations were sent to 75,000 individuals. The actual sample was comprised of 807 individuals who signed up and completed the baseline survey. Thus, it is not clear how representative the sample was of the entire population.

The initial group of participants was randomly assigned to one of four groups, each of which received a policy brief with different content. The control group received a policy brief that was similar in length and style to the other briefs but had no content relevant to agricultural interventions and children's nutrition. The three treatment groups all were presented policy briefs that discussed four agricultural intervention issues: biofortification, dairy development, home gardens, and small fisheries.

Masset et al. (2013) described the three treatment groups:

The first version was a basic policy brief of three pages. The second version was identical to the first version with the addition of a concluding opinion piece written by a sector expert and director of the institution conducting the review (five pages total). The third version was identical to the second version, but the final commentary was credited to an unnamed research fellow rather than to the director of the institute. (p. 51)

All participants were given a baseline survey that sought demographic information and also collected evidence about knowledge and beliefs regarding the interventions and their effectiveness. The two outcome measures were derived from answers to the following questions:

1. Do you believe the intervention is effective for improving the nutritional status of children?; and
2. To the best of your knowledge, how strong is the evidence on the effectiveness of the intervention?

A third outcome asked respondents about the likelihood that they would engage in follow-up actions that ranged from forwarding the brief or telling someone about it to actions that were more involved (e.g., reviewing current policies/practices, sourcing more information related to the topic of the policy brief).

The outcome measures were administered immediately after participants read the policy brief, then administered again one week later and again three months later.

The researchers had to contend with a very large attrition rate. Approximately half of the volunteers did not respond to the immediate survey. Another 17% of participants did not respond after one week. Only 29% of the original sample participated in the three-month follow-up.

The researchers conducted a set of analyses to test whether there were any differences on baseline measures between individuals who remained in the study and those who left. Several variables were related to attrition, but overall there did not seem to be substantial differences between the groups. Attrition rates did not differ across treatment groups.

Masset et al. (2013) found that, compared to the control group, all of the policy briefs increased the number of people who had an opinion on the strength of evidence about the agricultural interventions. However, ratings of effectiveness were not affected by the treatments. Further, there did not seem to be differences between the treatment conditions on either metric. Overall patterns did vary somewhat across the four agricultural interventions described in each document.

With respect to intended follow-up actions, the treatment groups, considered in aggregate, were more likely to pursue actions that required little effort such as rereading the brief, sending the brief to someone else, or telling someone about the key message of the brief. For more involved actions, there was very little evidence of any treatment effect.

Although the results did not confirm the researchers' initial hypotheses, the work demonstrates an approach to using experimental design to investigate the effects that format may have on the delivery of information through policy documents, and such designs may be useful for identifying effects in other contexts. The Masset et al. (2013) study further highlights issues of small sample and attrition that arise in experimental design studies, and the authors recommend that studies using similar methods attempt to address these issues in their initial design. To revise the methodology to better approach the research questions, the researchers suggest that lab studies that closely examine what individuals take away from different policy document manipulations may be more illuminating about the function of policy messaging than that which could be obtained by attempting to identify causal effects through experimentation in natural contexts.

Example: Using Quasi-Experiments to Study URE in Public Health – Olayo, Wafula, Aseyo, Loum, and Kaseje, (2014).

Olayo et al. (2014) studied the effectiveness of a research-based community health strategy on health outcomes in Kenya. This comprehensive strategy was an attempt to increase the use of health practices known, through research, to improve health and life outcomes. The study consisted of a two-year implementation of a health strategy policy in different socio-demographic regions of the country. The intervention consisted of a series of training sessions with health management teams, service providers, and the community. The three socio-demographic regions were nomadic, rural, and peri-urban (urban fringe) (Tacoli, McGranahan, & Satterthwaite, 2015).

In each region, communities had volunteered to participate in the strategy. Four communities across the three regions were selected as intervention sites. The researchers then identified matched communities that served as control sites. These control sites shared many characteristics with the intervention sites save a very important one—they did not volunteer to participate in the intervention.

The study collected two types of data. The first data set consisted of information about the extent to which intervention sites implemented community health strategy elements. This included such indicators as the establishment of committees and

committee meetings; implementation of and support for training; visits by community health workers; registration of community members; and holding of various community events and dialogues. These data were collected for the intervention sites several times over the course of the study.

The second set of data came from surveys of households in both the intervention and control sites. A baseline survey was given prior to the implementation, and a second survey was given following the intervention. Indicators included health facility delivery, antenatal care, presence of clinic cards in the household, immunization coverage, vitamin A supplement use, use of modern family planning methods, use of treated nets, water treatment, latrine use, and food availability—all critical to health and life outcomes.

The researchers' support for matching came from selecting sites in each socio-demographic region with seemingly similar profiles. They then compared the treatment and control groups on the survey indicators and concluded that they did not differ across groups.

Olayo et al. (2014) found that, for the most part, the elements of the health strategy policy were implemented in the intervention sites. They also found that health care outcomes generally improved more in the intervention sites, though that tended to differ both by health outcome and socio-demographic region. Their analyses consisted of testing for significant differences within the households for each treatment and intervention site in the study. No covariates were used in the analyses.

Although the researchers found that the research-based intervention was generally successful, they did recognize and discuss the limits on interpretation due to the self-selection into treatment and control groups. Olayo et al. (2014) summarized the study limitations in the following way:

> *We attempted to reduce bias by matching the control to the intervention districts by geographical location, ethnicity, infrastructure and socio-cultural characteristics. However, it was not possible to control for such factors as leadership and management effectiveness that would influence both the intervention and the health outcomes. (p. 10)*

The matching strategy and analytic approach are fairly weak, even in terms of quasi-experimental studies. However, given how health service support is delivered in Kenya and the context of the study, including resources and sites available, it may not have been possible to carry out more sophisticated designs. What can strengthen the confidence in the results is the authors' identification of similar patterns of results in multiple studies in other countries.

Example: Using Quasi-Experiments to Study URE in Public Health – Wagner , Soumerai, Zhang, and Ross-Degnan (2002).

Wagner et al. (2002) described a general approach, segmented regression analysis of interrupted time series designs that could be used to study the longitudinal effects of interventions in health policy. The example they drew from was the use of specific medicines in response to policy initiatives, which could be based on medical as well as economic research. While their focus was methodological, we include this example because of its potential to study URE across a range of contexts, including education and child welfare.

The most basic model is to take a series of measures (e.g., the average number of patient prescriptions given every month over a two-year period) for some time period before a policy intervention. Then following the intervention, the same measures would be taken over the next two years. The strength of this quasi-experimental approach is that the pre-intervention condition serves as a controlled comparison with the post-intervention condition.

Two parameters are central to estimating the impact of an intervention. One is *level,* or the intercept of the two time periods. Is there a change in the outcome measure after the policy change (e.g., *Are prescription rates lower after the intervention than they were before the policy intervention?*)? Analogously, to examine the impact of URE, studies of child welfare could examine differences in foster care placements and in education changes in disciplinary practices after the institution of a relevant policy that is based on the use of research evidence. A second parameter involves *trend.* Do trends (i.e., the slope) in particular outcomes change from before to after the intervention? Trend is important when policy changes result in changes that occur over time rather than having an impact that is observed immediately after the intervention.

Regression methods were used to estimate the statistical significance and strength of either of these changes. Wagner et al. (2002) then described applications of this general approach to situations in which there may have been multiple policy-based interventions sequenced over time or when the impact of an intervention took some time (lag) to take effect.

While segmented regression analysis of interrupted time series analysis can support causal claims about an intervention, there remain some important threats to interpretation that must be accounted for. The most significant threat is that some other factor could be introduced contemporaneously with the intervention that may lead to changes in the outcome of interest. Including a control group in the research design can be used to rule out such a competing hypothesis. Wagner et al. (2002) offered several other statistical and design approaches that reduced the likelihood of any fallacious causal attribution.

They also identified two potential limitations to using this approach. First, the models assumed linear trends in the outcomes within the pre- and post-segments. Non-linear patterns would not be correctly estimated with these methods. Second, because data were typically aggregated across individuals at each data point (e.g., each month), the models did not accommodate the inclusion of covariates at the individual level. Such covariates might have further explained any observed impact of the policy intervention. However, both of these limitations have been successfully addressed in subsequent research (Garabedian, Ross-Degnan, Ratanawijitrasin, Stephens, & Wagner, 2012; Wagner, personal communication; Weinberg et al., 2001).

This approach, as well as regression discontinuity models (Jacob, Zhu, Somers, & Bloom, 2012), offers the potential for strong URE causal designs in situations where both the outcome and research-based intervention can be clearly defined. The method offers an advantage over most experimental studies in that trend trajectories of impact can be estimated over time.

Case Studies

Case study is the dominant research paradigm for studies of URE. Although we often focus on single methods used in many of the studies cited throughout this monograph, there is more often than not one method used in a larger case study comprised of several methods. Yin (2014) defined case studies as "[a]n empirical inquiry about a contemporary phenomenon (e.g., a case), set within its real-world context—especially when the boundaries between phenomenon and context are not clearly evident" (p. 18).

Given a conception of the phenomenon of URE as "unfolding within a social ecology of relationships, organizational settings, and political and policy contexts" (Tseng, 2012, p. 7), it is not surprising that researchers have turned to cases as the preferred methodology to understand URE. The use of research evidence and the real-world context of its use are inseparable.

Yin (2012) claimed that case studies are appropriate to address the central questions of many URE studies: What is happening or has happened? (descriptive); and How or why did something happen? (explanatory). Case studies generally do not pursue questions of causal effect, such as with experimental and quasi-experimental design. Case studies are more frequently focused on understanding the causal mechanisms and processes by which contextual factors affect URE or URE leads to observed outcomes. Cases also are not designed to support the statistical generalization of findings. The focus of these studies is the case, in contrast to more quantitative methods that focus on relationships among specific variables.

The case itself can take a number of different forms and needs to be chosen based on the research questions of the study. A single case may, for example, focus on an organization (e.g., a school district or social service agency), the development of a policy (e.g., a law), or the implementation of practice (e.g., changes in health or social service practice). Many studies employ a multiple case design, including a deliberate sample of cases that would provide meaningful insight into the topic of interest, though the actual choice of cases should be guided by the theoretical framework of the researcher. Cases should be representative of salient characteristics of the population, though the analyses can be very different. Some multiple-case designs may look at the separate cases together, attempting to determine the cross-cutting processes that occur regardless of different contexts. Other multiple-case designs may look at the cases in contrast to each other in order to establish the mechanisms that lead to different outcomes.

Methodology

Design of case studies first requires the definition of the case. For URE studies, bounding the case is critical for a number of reasons. First, defining the case is a part of theory building and problem formulation. For some cases, the researcher may have a relatively clear sense of the problem and the theoretical framework that will drive the study prior to a study's inception. However, for many cases, theory building and problem formulation are iterative processes as the researcher begins to observe and better understand the problem context. Both the theory building and problem formulation assist the researcher in developing an argument for why the study of a particular case is important. As a case is defined it also contributes to shaping the collection of data, both in terms of how broad the data collection needs to be as well as providing guidance for how to bound the data collection and not go beyond saturation (Saunders et al., 2018).

Case studies, especially in studies of URE, almost always rely on multiple- or mixed-method research. In analyzing the data, researchers triangulate data from among the various sources of evidence. Their task is to seek consistencies as well as inconsistencies and to try to make sense of the multiple sources of evidence. For example, as previously discussed, in many URE studies there are inconsistencies between self-report data and observed data, and those inconsistencies may vary across individuals who have different roles in policy or practice. Those inconsistencies, and the reasons for them, become part of the case. In comparative case studies, characteristics of the contexts of each of the individual cases are used to make sense of inconsistencies. As researchers develop their cases, they consider alternative or rival explanations and seek evidence that either challenges the explanation or pushes the researchers to further develop their case in an appropriate manner or even accept rival explanations.

For introductory guidance on case studies, the following may be useful: Angrosino (2008); George and Bennett (2005); and Yin (2012, 2014).

Threats to Valid Interpretation

The primary threats to any case study are that the case is not adequately specified, that insufficient or inappropriate data are collected, or that interpretations at any point of case development are not warranted. Especially because case studies use multiple methods, questions can arise about how to integrate and analyze the data from the underlying methods. While this is generally true, it is especially true when the data collected points in different directions. There may not be sufficient contextual data to reliably interpret such situations, and it may also extend outside of established theory. This is especially true of URE research because so much of the theory building is still exploratory. It is important for the inferences made from the multiple sources of data to be consistent with the inferences that other researchers would make when examining the same set of evidence. The research teams can address this through internal inter-rater calibration, by checking with participants about whether the inferences comport with their understanding of their own context, and by sharing evidence with others and allowing them to also make independent interpretations.

Questions Case Studies Can Address

Case studies are well suited to address and offer explanations for most questions related to URE. Using a given theoretical framework, cases can describe the role and process of URE within the case, either contemporaneously or historically. Case studies can subsume all the research questions associated with particular methodologies. Because of the use of multiple methods, with data typically collected over an extended time frame, case studies can help address questions that extend beyond particular methods such as:

- How do relationships among individuals and entities within an institution influence URE?
- To what extent are self-reports of URE reflected in policymaking and practice that is observed or evident in artifacts such as policy documents?
- How does URE change within a case over time and context?
- How and why does URE vary across cases?

Questions Complementary Methods Can Address

Cases contribute to theories of URE. For some cases, at least, the theory may lead to testable hypotheses of factors relevant to URE that extend beyond the specific context examined. Generalized causal inference methods can then be used to test such case study hypotheses and to estimate the effects in a way that can be compared with other conditions. It is also of interest to understand whether the patterns uncovered in a case study apply broadly to other contexts or more narrowly to situations that have similar characteristics to the specific cases examined. Surveys administered broadly or developed to apply to other domains can be useful in determining the frequency of particular behaviors or attitudes across a broader set of contexts.

Example: Using Case Studies to Study URE in Education – Honig, Venkateswaran, and McNeil (2017).

Honig et al. (2017) looked at how school district central offices use research to provide evidence-based supports for teaching and learning and to inform decisions about school improvement. Their work focused on three fundamental questions:

1. How do central office administrators engage with research when they intend to use it to shift their practice, especially when the research fundamentally challenges the status quo?;
2. To what extent do they actually shift their practice?; and
3. What conditions support the use of research for significant shifts in practice? (p. 942)

For their conceptual framework, they employed sociocultural learning theory, describing the trajectories that novices and non-participants take to become experts (Lave & Wenger, 1991), and activity theory (Wertsch, 1993), which describes how individuals appropriate particular pedagogical tools as they engage in fundamental changes in practice (Grossman, Smagorinsky, & Valencia, 1999).

Honig et al. (2017) approached this work by mixing qualitative methods to arrive at a cohesive picture of central office practices and outcomes. They employed a multi-case design (Yin, 2014), examining six districts (of seven that were initially selected) in which the superintendents focused on using research to reform their central offices in ways they believed would lead to fundamental change in their current practice.

The primary method used was 500 hours of participant observations of 23 central office administrators who were most involved in the use of research and who received coaching from an intermediary organization. They combined this data source with 124 semi-structured interviews of these administrators, staff with whom they interacted, school principals, and coaches from the intermediary organization. Finally, they conducted a document analysis of more than 300 documents related to the core participants' research use.

Research staff combined and coded all data in multiple phases and iterations to ensure consistency. The researchers found that participants stated during interviews that they used research more than the observational and document evidence suggested, which aligns with past research showing that self-report suggests more evidence use than is observed. To account for over-reporting, the researchers focused only on examples of research use that participants said they were performing and that were also supported by observations or artifacts. Honig et al. (2017) used these data to determine the likelihood of increased research use over time. They used multiple approaches to analyze and report their findings, consistent with mixed-methods approaches of legitimization (Onwuegbuzie & Johnson, 2006). Honig et al. (2017) initially demonstrated aggregate patterns of research use across the entire participant sample and then provided more open, rich description to address the research questions and to describe differences across the six districts.

The researchers found four patterns of research use: 1) participants consistently did not use research during the study period (*low stasis*); 2) participants were already consistently using research at the start of the study period and continued to do so (*high stasis*); 3) participants started with a low level of research use and showed a small increase over the study period (*low growth*); and 4) participants showed high growth in their use of research over the study period (*high growth*). Throughout their presentation of results, they introduced snippets of interviews to illustrate how these different patterns manifest themselves.

Honig et al. (2017) found that high-growth districts had superintendents who were heavily engaged in leading the instructional sessions even while they were learning about the new practices themselves. Interestingly, and somewhat surprisingly, given prior findings on research-practice intermediary organizations, the researchers found that intermediary coaches who consistently engaged in the teaching practices in line with situated learning theory were associated with consistently low research use or low growth and that coaches who were inconsistent in engaging these teaching practices were involved in the high-growth cases. Together, these findings suggest the critical importance of internal district leadership in shifting practice related to research use.

Example: Using Case Studies to Study URE in Child Welfare – Mosley and Gibson (2017).

Mosley and Gibson (2017) used a case study approach as a way of understanding how evidence was used to enact a policy that extended foster care support in California to youth ages 18–21. The researchers adopted two theoretical frameworks to explain the policy process and expanded on them to theorize that different types of evidence are more influential at various points of the legislative process. They used a variety of research methods, including interviews, observations, and document analysis, to develop this case.

Mosley and Gibson (2017) adopted the very common perspective that research is but one form of evidence that is used in developing and enacting policy. Thus, in building their case, they wanted to understand the role of research evidence and its interplay with other sources of evidence during the legislative process. Their thinking was shaped by the Advocacy Coalition Framework (ACF; Sabatier, 1988), which argues that policymaking involves advocacy groups arguing for their particular positions based on belief systems. From this perspective, advocacy efforts are influenced by constraints and opportunities that surface through the policymaking effort, often leading to the use of different forms of evidence to address these issues. They complemented the ACF with insights from the Narrative Policy Framework (NPF), which focuses on the power of narratives to influence how policymaking takes shape (Jones & McBeth, 2010). Narratives can produce powerful stories about individuals and settings that can sway policymakers in ways that research and data cannot.

Mosley and Gibson (2017) focused on three research questions:

1. Why might certain types of evidence be more effective at certain points in the legislative process than others?;
2. What impact does budgetary constraint have on the types of evidence needed?; and
3. What is the relationship between evidence and narratives in the policymaking process? (p. 698)

California, along with other states, was in a position to respond to 2008 federal legislation offering matching funds to states that introduced programs extending foster care to older youth (ages 18–21). This legislation was passed at a time of economic crisis; consequently, many states did not introduce such programs because of additional costs they felt they could not support. Yet, California did pass a law (AB12) that was expensive ($70–80 million) in the midst of a recession. Mosley and Gibson (2017) wanted to understand how evidence influenced California to pursue the relatively unique direction it did.

The research team studied the legislative development over two years, conducting 38 in-depth interviews with key stakeholders (legislators, state and county administrators, judges and judicial staff, advocates, and funders). These interviews served as the primary data for developing the case. To corroborate and triangulate with these data, the research

team observed stakeholder and implementation planning meetings as well as open conference calls. They also conducted document reviews of legislative history, court rulings, press releases, meeting agendas, and various communications from advocates.

In establishing the case, the research team needed to identify key stakeholders within the legislative process, with particular focus on a coalition of advocates for AB12. Mosley and Gibson (2017) identified three discrete phases of the legislative process. For each, they identified the goals of the stage, the evidence that was needed to move the legislation forward, and the narrative elements that were critical to influencing policy.

The first stage involved drafting AB12 and moving it through committee. Legislative co-sponsors of the bill considered this phase as "laying the groundwork" and "maintaining credibility." The most important evidence brought to bear at this point was several research reports from highly regarded institutions that documented improved social outcomes (e.g., a reduction in homelessness and early pregnancy) with extended foster care. This research was complemented by narratives that helped to build the case that these youth were victims who could move from helplessness to control with the additional years of foster care.

The second phase was focused on addressing the issue of cost, which raised understandable concern by many legislators, particularly amid the economic crisis faced by California. Key evidence that was marshalled during this phase included a rigorous cost-benefit analysis to highlight short- and long-term benefits of enacting this policy. Narratives about the positive impact of other large social programs were also brought to the argument. Evidence in the form of data and narratives that led to very different interpretations was also introduced, ultimately resulting in significant compromises to the bill.

The final phase involved the bill's release from committee so that it could be voted on by the full legislature. It was at this point that narratives took center stage. For example, youth who had been successful through the use of extended foster care offered testimony. The narratives moved from youth as victims to youth as heroes who benefitted from the type of legislation being proposed.

In trying to fully understand this case and the enactment of legislation, the researchers opened themselves up to understanding how research evidence is used within a larger system of policymaking. From cases like this, it is possible to better understand the contexts in which URE is of primary evidence and when it may play a more secondary role. In focusing on the case rather than on particular instrumental variables (e.g., *What was the evidence of URE in the enactment of AB12?*), we are able to get a more textured and nuanced view of how, when, and why research is used within policymaking.

Example: Using Case Studies to Study URE in Education – Goertz, Barnes, and Massell (2013).

Goertz et al. (2013) conducted a comprehensive study of URE in state education agencies (SEAs) engaged with school improvement strategies. The study examined three state agencies through interviews, surveys, social network analysis, and document analysis. In this example, we focus on the combined use of surveys and interviews.

The study began with a conceptual framework of knowledge use that attends to how different types of knowledge—research, other evidence, and practice—enter into the system from different sources and, within a complex social and organizational context, influence school improvement policies and programs. This framework guided the design of all research methods and is presented in Figure 2.

A first round of interviews was conducted with 49 senior SEA staff members across the three states. Interviews were tailored to the particular roles of staff in the organizations in terms of their leadership role and extent of involvement with school improvement initiatives. Questions focused on each of the components of the conceptual framework in order to obtain a full understanding of how knowledge use occurred within complex systems. A second round of interviews was carried out with 11 of the individuals most directly involved with school improvement.

Surveys were sent to 610 SEA staff across the three states. The researchers attempted to connect with all staff who might be involved with school improvement. Of the 450 who responded, 305 answered affirmatively to a question about their involvement with school improvement efforts. Only the responses from these individuals were used to inform the study results. The survey focused on all aspects of the conceptual framework as well as items that would support the subsequent social network analysis.

Goertz et al. (2013) analyzed and reported their data in the policy brief by discussing all evidence related to each framework concept rather than by separating them by the method used to collect the data. Thus, the survey and interview data were combined with the social network analysis to create an integrated set of findings for each state on each of the following four issues:

- SEA context, organization, and school improvement strategies and delivery systems;
- The flow and organization of information in SEAs' work and knowledge networks;
- External organizations; and
- The use of knowledge in school improvement policy.

The scope of the findings exceeds what we can summarize here, but the study provides a strong example of how complementary methods used across multiple sites can illuminate URE in ways that are more comprehensive than single methodological approaches. At the same time, the strength of the conceptual framework was necessary to organize all facets of the study, from sampling to instrument design and through analysis and reporting. It would be much more difficult to integrate the wide variety of data sources included in this study without such a framework to anchor the analysis.
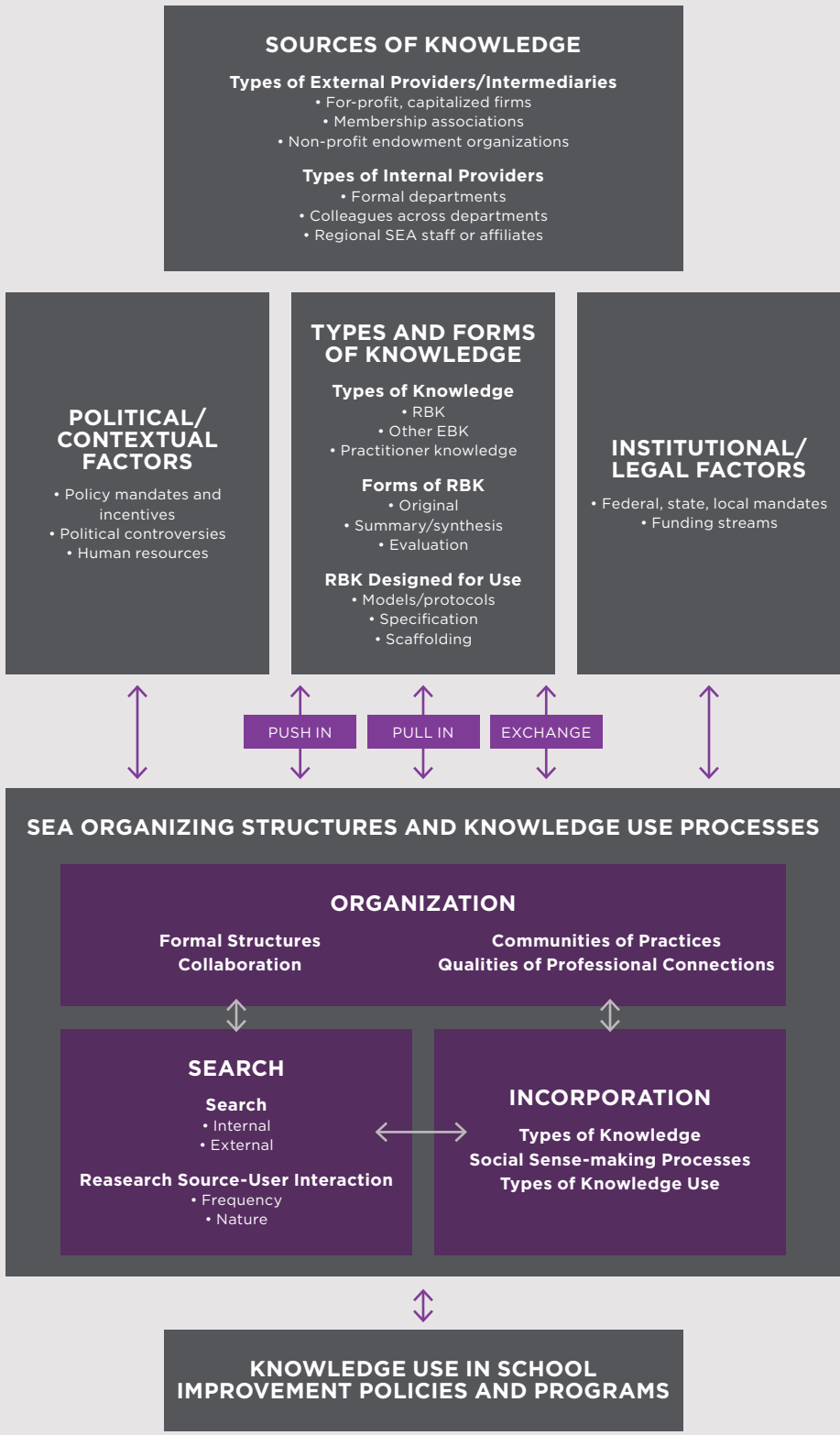
SOURCES OF KNOWLEDGE

Types of External Providers/Intermediaries
• For-profit, capitalized firms
• Membership associations
• Non-profit endowment organizations

Types of Internal Providers
• Formal departments
• Colleagues across departments
• Regional SEA staff or affiliates

POLITICAL/ CONTEXTUAL FACTORS

• Policy mandates and incentives
• Political controversies
• Human resources

TYPES AND FORMS OF KNOWLEDGE

Types of Knowledge
• RBK
• Other EBK
• Practitioner knowledge

Forms of RBK
• Original
• Summary/synthesis
• Evaluation

RBK Designed for Use
• Models/protocols
• Specification
• Scaffolding

INSTITUTIONAL/ LEGAL FACTORS

• Federal, state, local mandates
• Funding streams

PUSH IN   PULL IN   EXCHANGE

SEA ORGANIZING STRUCTURES AND KNOWLEDGE USE PROCESSES

ORGANIZATION

Formal Structures
Collaboration

Communities of Practices
Qualities of Professional Connections

SEARCH

Search
• Internal
• External

Reasearch Source-User Interaction
• Frequency
• Nature

INCORPORATION

Types of Knowledge
Social Sense-making Processes
Types of Knowledge Use

KNOWLEDGE USE IN SCHOOL IMPROVEMENT POLICIES AND PROGRAMS

Figure 2. Conceptual framework from Goertz et al. (2013).

Example: Using Case Studies to Study URE in General Policymaking – Newman, Cherney, and Head (2016).

Newman et al. (2016) used surveys and interviews to investigate the legitimacy of the claim that policymakers and academics are two separate communities, wherein "social scientists and policymakers live in separate worlds with different and often conflicting values, different reward systems, and different languages" (Caplan, 1979, p. 459). Newman et al. (2016) reviewed the literature on URE and contended that while the influence of research on policy was limited, the barriers were not as stark as represented in the two-community theory.

The researchers received survey responses from 2,084 public servants (policymakers) at state and federal levels in Australia. They acknowledged that there was not a statistically representative probability sample and, thus, there would be some risk of self-selection biasing the results. The survey made use of questions from other studies and focused on questions of how and the extent to which policymakers accessed and made use of academic research in their work to understand policy and programs or as information to cite as they prepared policy documents. The project took a more circumscribed view of research than many other URE studies, focusing on academic articles and reports rather than also including other sources of research-related information.

Newman et al. (2016) found that the majority of policymakers accessed and read academic research in their work, though there was great variability and most use was occasional. However, they also found that using research instrumentally in policymaking was less frequent, often because of factors such as budgetary constraints and time pressures. They argued that the survey findings did not support a two-community view but rather one that was more complicated and nuanced with a wide variety of research/policy interactions.

They then followed up the survey with semi-structured interviews of 126 individuals who were selected because they held more senior roles in their organizations. Some, but not all, had taken the survey. The interviewees were selected through a variety of channels and, thus, the sampling here was not systematic or necessarily representative.

Responses from the interviews were also far from homogenous. While there was a substantial number of interviewees who had little connection to or respect for academic research, there were others who highly valued research. The majority of respondents fell somewhere in between. The researchers identified from the interview responses four major themes that influenced URE: academic contextual factors such as incentive structures and institutional priorities; research dissemination; knowledge brokering; and competing pressures on policy formulation, such as those coming up from stakeholders or down from the political executive (Newman et al., 2016, p. 28).

Newman et al. (2016) used the collective findings to highlight the complexity of the policymaker-academic relationship and the need to find metaphors more useful than the two-community concept. They argued that their findings are more consistent with the model of URE as being strengthened by knowledge brokers and intermediaries who can support policymakers' substantial inclinations to use research, albeit within highly constrained contexts.

Closing Comments and Moving the Field Forward

Scholarship on the use of research evidence now has several decades behind it and continues to build on seminal works that are still having a major influence on work today (e.g., Caplan, 1979; Kingdon, 2011; Weiss & Bucuvalas, 1980). More recently, and particularly in light of increasing policy pressure for research to inform practice, a significant and vibrant body of research on URE has continued to grow. In the last few years, the field has continued to expand, in part with significant support from the William T. Grant Foundation, which also has funded this monograph.

One of the most striking things that we discovered in the course of doing this work is the collaborative nature of the URE research community and its pragmatic and pluralistic approach to methodology. On the one hand, URE research spans many traditional academic disciplines in the focus of its work, crossing boundaries among such diverse fields as education, social work, nursing, information sciences, and medicine. Yet, we have not seen the entrenched methodological divides and silos that characterize many fields of inquiry in the social sciences. Thus, the usual call for more cross-talk among social science researchers who have different methodological commitments is quite unnecessary here.

We have five suggestions that we believe will strengthen the research and move the field forward as it continues to mature and evolve. Each of these suggestions addresses aspects of issues introduced at the beginning of this monograph. The suggestions represent kernels of ideas that will require broader consideration by established URE scholars and the inclusion of new researchers to further address these areas of growth and produce new initiatives for the field. The first two suggestions focus on substantive questions that have not been the subject of sustained investigation. The other three suggestions focus on tools and processes to enhance the methodological work of URE researchers.

Bringing Critical Perspectives to URE

Studies of URE almost always have a strong theoretical perspective that frames the research, informs methodological choices, and guides interpretations. While the examples cited in this monograph represent a broad range of theoretical perspectives, notably missing are critical perspectives.

Critical social theory encompasses theoretical perspectives that include critical race theory and critical pedagogy (D. A. Bell, 1980; Giroux, 1997; Freire, 2000; R. Fowler, Hodge, Kress, & Trew, 1979; Ledesma & Calderón, 2015; Tate, 1997), and together these theoretical traditions have brought to the fore issues of structural inequality that are essential to understanding our social institutions. Critical researchers look at problems through various lenses of race, language, gender, sexuality, and disability, often involving social and historical context to investigate the implicit power structures in society and the underlying social dynamics that reproduce them. Critical theorists also demonstrate that traditional academic research and methods often overlook or fail to capture these mechanisms of social inequity. They have been building more systematic approaches to studying these issues as well as developing practical applications such as critical discourse methods and culturally responsive proposed evaluation models (Hammersley, 1997; Hood, Hopson, & Kirkhart, 2015; Scheurich & Young, 1997).

The incorporation of principles of critical social theory and the subsequent methods will likely expand the field of URE in ways that are not currently salient in this scholarly community. While much of URE scholarship relies on interpretive methods to address the complicated and contested definitions of research and evidence use and to explore the social contexts in which it is embedded, it seems appropriate and important to also bring to the field critical lenses that attend to the underlying social structure, implicit social inequality, and ways in which inequality is reproduced. Critical interpretations are likely to broaden understandings of how power dynamics intersect with URE within institutions and will consider the contexts in which URE functions to replicate the existing power hierarchy or silence voices as well as times when it is used to disrupt it. We suspect further investigation would also identify contextual factors in the structure of organizations and within the control of decision makers that can remedy or inadvertently increase such social disparities. In all cases, critical theories are an important set of theoretical perspectives that will enhance the current field of URE and provide more nuanced understanding of its effects and effectiveness.

Critical interpretations are likely to broaden understandings of how power dynamics intersect with URE within institutions and will consider the contexts in which URE functions to replicate the existing power hierarchy or silence voices as well as times when it is used to disrupt it.

Addressing Research Quality

The object of study for most research of URE is situated at the intersection of three broad elements: 1) influences of research use; 2) targets of research use; and 3) claims about research use. The political nature of URE in policy development has been well documented (e.g., National Research Council, 2012) and illustrated in many of the examples cited. Many of the current studies examine different ways in which URE exists in different stages and how characteristics of individuals and organizations relate to whether and how research is used. For the most part, however, there is very little systematic study of which research is selected and which is ignored, how research is selected, and by what approaches practitioners gauge the quality of the research with which they are interacting.

It is quite apparent that focusing on different research evidence will lead to very different implications for policy and practice. This is true, even if the different evidence sources are all well regarded. It is often the case that different researchers explore similar problems from different vantage points and explore different aspects of the problem.

Take, for example, the research area of teaching quality in K–12 classrooms. Economists have built up a significant literature base that focuses on teacher quality in terms of impact on student test scores and has attempted to identify factors that are predictive in these terms, which often leads to discussion about using teacher characteristics for hiring and retention strategies (e.g., Chetty, Friedman, & Rockoff, 2014; Fuller & Ladd, 2012; Jacob, 2011).

In contrast, teacher educators have a complementary literature that focuses on developing and sustaining better instruction through changes in teacher preparation, curriculum development, and professional learning initiatives (e.g., Allen & Penuel, 2015; C. A. Bell, Wilson, Higgins, & McCoach, 2010; Cochran-Smith et al., 2016).

While both of these research perspectives are well established and include highly regarded studies, decisions that practitioners may make would look very different if they considered only one of these perspectives. At present, studies of URE do not attempt to account for the degree to which the research that is used in decision-making is germane to the topic or whether all areas of established and relevant research focus are considered. We mentioned at the beginning of this monograph that a strength in the URE community is that it is highly collaborative and often crosses methodological and theoretical boundaries. At the same time, there are few indications that the field has identified the differences within the research communities that produce the relevant research or has found meaningful ways to incorporate disparate research areas for the same topic of interest. We understand that this will not be an easy task for the URE community because we are suggesting, in effect, that it develop methods to account for the silos and contested terrain of the underlying research areas in its assessment of research use and what research is considered for use. However, the fact that there are silos and contested terrain that have not yet been resolved suggests that innovative approaches to account for this in URE research will be necessary.

In a similar vein, another characteristic of the research used that we believe needs further attention is the question of quality. Published and disseminated research studies vary greatly in the methodological rigor of study design, in the quality of data collection, analysis, and discussion, and in the likelihood that the results will generalize to other contexts. For URE to be effective in improving outcomes, practitioners need to rely primarily on high-quality and contextually appropriate sources; however, we have not found examples in the existing literature that attend to the question of quality.

Building a Library of Methods and Measures

Because of the conceptual complexity of the meaning of "research" and "evidence" and that "use" itself is often not a directly observable phenomenon, we find tremendous variation in the specific data collection protocols regarding these concepts. In many cases, these reflect context-specific differences in the way URE exists in different fields and the current work of the individual organization. For instance, in some studies described in this monograph, URE is operationalized by a self-reported account of what the research "says" but without any reference; elsewhere, URE is operationalized by the explicit citation of a published source, and in others, the use of statistics that arose from a reputable agency constitutes a form of URE. Some of the studies further explore the meaning and extent of what constitutes research evidence itself and how it interplays with the use of other reputable sources of non-research.

Further, the predominant use of case studies and other qualitative methodologies means that generalizations from studies will be theoretical rather than statistical. Stronger connections between studies are possible when they make use of common methods and measures, and this can help users understand the degrees to which findings in one context apply in other contexts and whether interventions related to URE can be used broadly. As one of the underlying goals of URE research is to encourage and evaluate whether using research for decision-making leads to improvements in social outcomes, employing common methods to establish stronger grounds for generalization is an important next step in the field.

Unfortunately, much of the detail about study methods that would be necessary to fully understand the methodological bounds is not included in scholarly publications. In large part, we find that space limitations that are part of the publication process are a hindrance. Therefore, it is relatively rare to see specific instruments, data collection protocols, or coding systems as part of the publications.

We suggest that a more systematic approach for the sharing of research instruments, analytic schemes, operationalization of core components, and other tools be made available in a repository specifically created for URE research. All contributions would be tagged with metadata so that a user could search the data set for research tools that had particular characteristics. An interested researcher might ask to see interviews that have been used to gather self-report evidence to examine how individuals within an organization interact with each other with respect to URE or might be looking for ways that prior researchers examined research in policymaking decisions surrounding child welfare. Even though contexts and situations differ, the nature of questions and issues addressed might have utility for the interested researcher, and this may allow him/her to more directly connect to past work and expand the understanding of how similar aspects of URE apply in the new domain or situation.

As one of the underlying goals of URE research is to encourage and evaluate whether using research for decision-making leads to improvements in social outcomes, employing common methods to establish stronger grounds for generalization is an important next step in the field.

To build such a library, institutional, conceptual, technical, and logistical aspects need to be addressed, especially as such an undertaking will undoubtedly require iterative development and revision to make it both robust enough to support the entire URE research community and simple enough to be easily accessed. Institutional issues include governance questions about ownership and copyrights of the contributions; establishing the institution, organization, or leadership body that will oversee its development; maintenance; whether security of data or metadata needs to be regulated; and the permissibility and protocol for changes and edits to be made following initial submission.

Conceptual issues include what artifacts should be part of the repository, which metadata are maintained, and how to engage the URE community to contribute and maintain their contributions. While we do find that the URE community is very collaborative in working together and discussing multiple methodological approaches, translating that into the work of contributing methodological detail to a new, external repository may not necessarily follow without additional guidance or community support. The complexity of developing a technically sound, broadly usable interface for input and retrieval, as well as a robust underlying data structure that can easily be revised and extended, then implementing necessary security protocols is enough to warrant significant attention.

Creating Methodological Templates

The methods and measures used in current URE studies tend to be relatively idiosyncratic to the projects. In some circumstances, the manner of operationalizing URE in one study may be mutually exclusive or even explicitly outside the bounds of the method used to operationalize it in another study. Because of the wide variation of fields and domains in which URE is investigated, we do not believe it would be appropriate to seek a set of common measures that can be used across broad sets of studies.

However, in order to further support theoretical connections among studies, it would be useful to develop a set of methodological templates that can be adapted to similar contexts. This would benefit from connecting to the proposed library of measures and metadata, which itself could be expanded to play a role in the development of these templates. If, for example, one could look across studies and locate interview questions that focus on the instrumental use of research, would it be possible to create templates that could readily be adapted for particular contexts?

As an example, in the method sections we discussed the utility of open and iterative coding for exploratory research, the value of more structured coding schemes to validate and extend existing theory, and that we primarily see open coding frameworks in the literature to date. As the field of URE further develops, connecting to existing templates of coding structures would be valuable for this second area of study and could give cues to new researchers for aspects of deliberations to which they could attend.

To further support theoretical connections among studies, it would be useful to develop a set of methodological templates that can be adapted to similar contexts.

Consequently, designing such templates represents a significant conceptual challenge, especially because of the extensive variation of meaning of URE in different contexts, studies, and uses. There are also technical challenges to ensure that using and customizing templates is a straightforward and manageable process. How can the template usefully connect researchers and study designs while respecting the real variation in forms of evidence, uses of research, and meaning of research itself? What is involved in designing templates that can be used to support new initiatives and explore new terrains without being overly constraining? This is a design activity that may itself be suited to a coordinated community involvement, beginning with acquiring input from many URE researchers, designing templates and studying their use by researchers, and conducting an evaluation of the utility of the approach.

Developing Research Reporting Guidelines

The National Institutes of Health maintain a repository of guidelines for the reporting of research in various areas of specialization (see https://www.nlm.nih.gov/services/research_report_guide.html). The site states that these guidelines "provide advice for reporting research methods and findings" and "usually specify a minimum set of items required for a clear and transparent account of what was done and what was found in a research study, reflecting, in particular, issues that might introduce bias into the research."

What would a set of guidelines for the reporting of URE studies look like? In terms of both content and process, this is a task that requires many voices and contributions, and significant negotiation. While such guidelines cannot be overly prescriptive, it is possible to imagine guidance that, at least for specific methodological approaches, specifies expectations for what needs to be included in describing study design, instrumentation, sampling issues, analytic procedures, etc.

Conclusion

The research that has been reviewed has been performed during a time in which research evidence that predominantly draws consistent conclusions about even the most actionable questions is often ignored or dismissed for a variety of political and economic reasons. This is frequently coupled with challenges to the merits and integrity of the scientific community at large, especially by politicians that have alternative policy perspectives than the direction in which the research points. To be fair, and as we have discussed, there are variations in the methodological rigor and quality of research that at proposed for use in decision-making, and there are times in which other sources of evidence are appropriate to consider in competition with research findings for use by practitioners.

Yet, using research to inform practice and policymaking remains a valued goal of both researchers across the social sciences and the practitioners and policymakers who are making decisions. The current URE research community has found that research use continues to be both mandated by decision-making bodies and valued across institutions that are concerned with societal welfare. However, it also finds that the processes and understanding surrounding URE have significant areas yet to be explored.

Research and evidence use are complex concepts that vary substantially across different fields and decision-making contexts, and for specific situations. URE methods, accordingly, show significant innovation and flexibility to address these variations. We hope that this monograph is helpful in illuminating the different aspects of URE, the methodological terrain that is currently involved, and areas for further investigation and elaboration. We are eager to see the ways in which future research improves on current methods and develops tools that help the URE community to provide more systematic insight into complex social organizations, their deliberative processes, and their use of research evidence. By continuing to explore URE through methods that illuminate the phenomena, deeper understandings of the connections among research, policy, practice, and outcomes will continue to evolve.

References

Allen, C. D., & Penuel, W. R. (2015). Studying teachers' sensemaking to investigate teachers' responses to professional development focused on new standards. *Journal of Teacher Education, 66*(2), 136–149. doi: 10.1177/0022487114560646

Angrosino, M. V. (2008). *Doing ethnographic and observational research.* London, UK: Sage Publications, Ltd.

Asen, R., Gurke, D., Conners, P., Solomon, R., & Gumm, E. (2013). Research evidence and school board deliberations: Lessons from three Wisconsin school districts. *Educational Policy, 27*(1), 33–63. doi: 10.1177/0895904811429291

Bell, C. A., Wilson, S. M., Higgins, T., McCoach, D. B. (2010). Measuring the effects of professional development on teacher knowledge: The case of developing mathematical ideas. *Journal for Research in Mathematics Education, 41*(5), 479–512. Retrieved from https://www.jstor.org/stable/41110411

Bell, D. A. (1980). *Brown* v. *Board of Education* and the interest-convergence dilemma. *Harvard Law Review, 93*(3), 518–533. doi: 10.2307/1340546

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal, 9*(2), 27–40. doi: 10.3316/QRJ0902027

Bradburn, N. M., Wansink, B., & Sudman, S. (2015). *Asking questions: The definitive guide to questionnaire design—for market research, political polls, and social and health questionnaires.* San Francisco, CA: Jossey-Bass.

Brennan, S. E., McKenzie, J. E., Turner, T., Redman, S., Makkar, S., Williamson, A., … Green, S. E. (2017). Development and validation of SEER (Seeking, Engaging with and Evaluating Research): A measure of policymakers' capacity to engage with and use research. *Health Research Policy and Systems, 15*(1). doi: 10.1186/s12961-016-0162-8

Caplan, N. (1979). The two-communities theory and knowledge utilization. *American Behavioral Scientist, 22*(3), 459–470. doi: 10.1177/000276427902200308

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review, 104*(9), 2633–2679. doi: 10.1257/aer.104.9.2633

Chouliaraki, L., & Fairclough, N. (1999). *Discourse in late modernity: rethinking critical discourse analysis.* Edinburgh, Scotland: Edinburgh University Press.

Coburn, C. E. (2001). Collective sensemaking about reading: How teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis, 23*(2), 145–170. doi: 10.3102/01623737023002145

Coburn, C. E., & Talbert, J. E. (2006). Conceptions of evidence use in school districts: Mapping the terrain. *American Journal of Education, 112*(4), 469–495. doi: 10.1086/505056

Cochran-Smith, M., Villegas, A. M., Abrams, L. W., Chavez-Moreno, L. C., Mills, T., & Stern, R. (2016). Research on teacher preparation: Charting the landscape of a sprawling field. In D. H. Gitomer & C. A. Bell (Eds.), *Handbook of research on teaching* (5th ed., pp. 439–547). American Educational Research Association. doi: 10.3102/978-0-935302-48-6_7

Collier, D. (2011). Understanding process tracing. *Political Science and Politics, 44*(4), 823-830. doi: 10.1017/S1049096511001429

Corbin, J. M., & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory.* Thousand Oaks, CA: Sage Publications, Inc.

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Boston, MA: Pearson.

Dawes, M., Summerskill, W., Glasziou, P., Cartabellotta, A., Martin, J., Hopayian, K., ... Osborne, J. (2005). Sicily statement on evidence-based practice. *BMC Medical Education, 5*(1). doi: 10.1186/1472-6920-5-1

Dobbins, M., Hanna, S. E., Ciliska, D., Manske, S., Cameron, R., Mercer, S. L., ... Robeson, P. (2009). A randomized controlled trial evaluating the impact of knowledge translation and exchange strategies. *Implementation Science, 4*(1). doi: 10.1186/1748-5908-4-61

DuMont, K. (2016). *Studying ways to improve the use of research evidence: Presenting a strong rationale in your application.* New York, NY: William T. Grant Foundation. Retrieved from http://wtgrantfoundation.org/studying-ways-improve-use-research-evidence-presenting-strong-rationale-study-application

Edwards, A. L. (1982). *The social desirability variable in personality assessment and research.* Westport, CT: Greenwood Press.

Edwards, R., & Holland, J. (2013). *What is qualitative interviewing?* London, UK: Bloomsbury Academic.

Eisenhart, M. (2001). Changing conceptions of culture and ethnographic methodology: Recent thematic shifts and their implications for research on teaching. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 209–225). Washington, DC: American Educational Research Association.

Elementary and Secondary Education Act of 2002, Pub. L. No. 107-110.

Ercikan, K., & Roth, W.-M. (2014). Limits of generalizing in education research: Why criteria for research generalization should include population heterogeneity and uses of knowledge claims. *Teachers College Record, 115*(5). Retrieved from http://www.tcrecord.org/content.asp?contentid=17429

Ericcson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Revised ed.). Cambridge, MA: MIT Press.

Erickson, F. (1986). Qualitative methods in research on teaching. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 119–161). New York, NY: Macmillan.

Fairclough, N. (2013). *Critical discourse analysis: The critical study of language* (2nd ed.). London, UK: Routledge.

Farley-Ripple, E. N. (2012). Research use in school district central office decision making: A case study. *Educational Management Administration & Leadership, 40*(6), 786–806. doi: 10.1177/1741143212456912

Fetterman, D. M. (2010). *Ethnography: Step-by-step* (3rd ed.). Los Angeles, CA: Sage Publications, Inc.

Finnigan, K. S., Daly, A. J., & Che, J. (2013). Systemwide reform in districts under pressure: The role of social networks in defining, acquiring, using, and diffusing research evidence. *Journal of Educational Administration, 51*(4), 476–497. doi: 10.1108/09578231311325668

Firestone, W. A. (1993). Alternative arguments for generalizing from data as applied to qualitative research. *Educational Researcher, 22*(4), 16–23. doi: 10.3102/0013189X022004016

Fowler, F. J. (1995). *Improving survey questions: design and evaluation.* Thousand Oaks, CA: Sage Publications, Inc.

Fowler, F. J., & Mangione, T. W. (1990). *Standardized survey interviewing: Minimizing interviewer-related error.* Newbury Park, CA: Sage Publications, Inc.

Fowler, R., Hodge, B., Kress, G., & Trew, T. (1979). *Language and control.* London, UK: Routledge and Kegan Paul, Ltd.

Frank, K. A. (1996). Mapping interactions within and between cohesive subgroups. *Social Networks, 18*(2), 93–119. doi: 10.1016/0378-8733(95)00257-X

Freire, P. (2000). *Pedagogy of the oppressed* (30th anniversary ed.). New York, NY: Continuum International Publishing Group.

Fuller, S. C., & Ladd, H. F. (2012). *School based accountability and the distribution of teacher quality among grades in elementary school* (CALDER Working Paper No. 75). Washington, DC: Center for the Analysis of Longitudinal Data in Education Research, American Institutes for Research. Retrieved from http://files.eric.ed.gov/fulltext/ED532767.pdf

Galletta, A., & Cross, W. E. (2013). *Mastering the semi-structured interview and beyond.* New York, NY: New York University Press.

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

Garabedian, L. F., Ross-Degnan, D., Ratanawijitrasin, S., Stephens, P., & Wagner, A. K. (2012). Impact of universal health insurance coverage in Thailand on sales and market share of medicines for non-communicable diseases: An interrupted time series study. *BMJ Open, 2*(6), e001686. doi: 10.1136/bmjopen-2012-001686

Garces, L. M., Marin, P., & Horn, C. L. (2017). The U.S. Supreme Court's use of non-legal sources and *amicus curiae* briefs in *Fishe*r v. *University of Texas*. *Journal of College and University Law, 43*(2), 167–194. Retrieved from https://heinonline.org/HOL/Page?collection=journals&handle=hein.journals/jcolunly43&id=183

Gee, J. P. (2011). *An introduction to discourse analysis: Theory and method* (3rd ed.). New York, NY: Routledge.

Geertz, C. (1973). *The interpretation of cultures: Selected essays*. New York, NY: Basic Books.

George, A. L., & Bennett, A. (2005). *Case studies and theory development in the social sciences*. Cambridge, MA: The MIT Press.

Giroux, H. A. (1997). *Pedagogy and the politics of hope: Theory, culture, and schooling: A critical reader*. Boulder, CO: Westview Press.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative theory.* New York, NY: Aldine De Gruyter.

Glasow, P. (2005). *Fundamentals of survey research methodology* (MP 05W0000077). McLean, VA: MITRE.

Goertz, M., Barnes, C., & Massell, D. (2013). *How state education agencies acquire and use research in school improvement strategies* (CPRE Policy Brief #RB-55). Philadelphia, PA: Consortium for Policy Research in Education. doi: 10.12698/cpre.2013.rb55

Goodnight, G. T. (2012). The personal, technical, and public spheres of argument: A speculative inquiry into the art of public deliberation. *Argumentation and Advocacy*, *48*(4), 198–210. doi: 10.1080/00028533.2012.11821771

Gormley, W. T. (2011). From science to policy in early childhood education. *Science, 333*(6045), 978–981. doi: 10.1126/science.1206150

Granger, R.C. (2018). The roles foundations are playing in the evidence-based policy movement. *The Annals of the American Academy of Political and Social Science*, *678*, 145-154.

Grossman, P. L., Smagorinsky, P., & Valencia, S. (1999). Appropriating tools for teaching English: A theoretical framework for research on learning to teach. *American Journal of Education, 108*(1), 1–29. Retrieved from https://www.jstor.org/stable/1085633

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Gubrium, J. F., Holstein, J. A., Marvasti, A. B., & McKinney, K. D. (Eds.). (2012). *The sage handbook of interview research: The complexity of the craft* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Hammersley, M. (1997). On the foundations of critical discourse analysis. *Language & Communication, 17*(3), 237–248. doi: 10.1016/S0271-5309(97)00013-X

Hanneman, R. A., & Riddle, M. (2005). *Introduction to social network methods*. Riverside, CA: University of California, Riverside. Retrieved from http://faculty.ucr.edu/~hanneman/

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures, 1*(1), 77–89. doi: 10.1080/19312450709336664

Hodder, I. (2000). The interpretation of documents and material culture. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 110–129). Thousand Oaks, CA: Sage Publications, Inc.

Honig, M. I., Venkateswaran, N., & McNeil, P. (2017). Research use as learning: The case of fundamental change in school district central offices. *American Educational Research Journal*, *54*(5), 938–971. doi: 10.3102/0002831217712466

Hood, S., Hopson, R. K., & Kirkhart, K. E. (2015). Culturally responsive evaluation. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of practical program evaluation* (4th ed.). San Francisco, CA: Jossey-Bass.

Jacob, B. A. (2011). Do principals fire the worst teachers? *Educational Evaluation and Policy Analysis, 33*(4), 403–434. doi: 10.3102/0162373711414704

Jacob, R., Zhu, P., Somers, M.-A., & Bloom, H. (2012). *A practical guide to regression discontinuity*. Washington, DC: MDRC. Retrieved from https://www.mdrc.org/publication/practical-guide-regression-discontinuity

Johnstone, B. (2018). *Discourse analysis* (3rd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Jones, M. D., & McBeth, M. K. (2010). A narrative policy framework: Clear enough to be wrong? *Policy Studies Journal*, *38*(2), 329–353. doi: 10.1111/j.1541-0072.2010.00364.x

Kane, M. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). New York, NY: American Council on Education.

Kingdon, J. W. (2011). *Agendas, alternatives, and public policies* (Updated 2nd ed.). Boston, MA: Longman.

Knoke, D., & Yang, S. (2008). *Social network analysis* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Kvale, S., & Brinkmann, S. (2015). *InterViews: Learning the craft of qualitative research interviewing* (3rd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Landry, R., Lamari, M., & Amara, N. (2003). The extent and determinants of the utilization of university research in government agencies. *Public Administration Review, 63*(2), 192–205. doi: 10.1111/1540-6210.00279

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.

Ledesma, M. C., & Calderón, D. (2015). Critical Race Theory in education: A review of past literature and a look to the future. *Qualitative Inquiry, 21*(3), 206–222. doi: 10.1177/1077800414557825

Lohr, S. L. (2010). *Sampling: Design and analysis* (2nd ed.). Boston, MA: Brooks/Cole.

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

Mason, J. (2002). *Qualitative researching* (2nd ed.). London, UK: Sage Publications, Ltd.

Masset, E., Gaarder, M., Beynon, P., & Chapoy, C. (2013). What is the impact of a policy brief? Results of an experiment in research dissemination. *Journal of Development Effectiveness, 5*(1), 50–63. doi: 10.1080/19439342.2012.759257

Maynard, R. (2018). The role of federal agencies in creating and administering evidence-based policies. *Annals of the American Academy of Political and Social Science, 678*, 134–144.

McDonnell, L. M., & Weatherford, M. S. (2013). Evidence use and the Common Core State Standards movement: From problem definition to policy adoption. *American Journal of Education, 120*(1), 1–25. doi: 10.1086/673163

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–100). Washington, DC: American Council on Education.

Morgan, A. (2010). Discourse analysis: An overview for the neophyte researcher. *Journal of Health and Social Care Improvement, 5*, 1–7.

Morgan, S. L., & Winship, C. (2007). *Counterfactuals and causal inference: Methods and principles for social research*. New York, NY: Cambridge University Press.

Mosley, J. E., & Gibson, K. (2017). Strategic use of evidence in state-level policymaking: matching evidence type to legislative stage. *Policy Sciences, 50*(4), 697–719. doi: 10.1007/s11077-017-9289-x

Moss, P. A., & Haertel, E. H. (2016). Engaging methodological pluralism. In D. H. Gitomer & C. A. Bell (Eds.), *Handbook of research on teaching* (5th ed., pp. 127–147). Washington, DC: American Educational Research Association.

Murnane, R. J., & Willett, J. B. (2011). *Methods matter: Improving causal inference in educational and social science research*. New York, NY: Oxford University Press.

National Governors Association (NGA), Council of Chief State School Officers (CCSSO), & Achieve, Inc. (2008). *Benchmarking for success: Ensuring U.S. students receive a world-class education*. Washington, DC: Authors. Retrieved from https://eric.ed.gov/?id=ED504084

National Research Council. (1978). *Knowledge and policy: The uncertain connection*. Washington, DC: The National Academies Press. doi: 10.17226/19941

National Research Council. (2002). *Scientific research in education*. Washington, DC: The National Academies Press. doi: 10.17226/10236.

National Research Council. (2012). Using science as evidence in public policy. Washington, DC: The National Academies Press. doi: 10.17226/13460.

Newman, J., Cherney, A., & Head, B. W. (2016). Do policy makers use academic research? Reexamining the "two communities" theory of research utilization. *Public Administration Review, 76*(1), 24–32. doi: 10.1111/puar.12464

Olayo, R., Wafula, C., Aseyo, E., Loum, C., & Kaseje, D. (2014). A quasi-experimental assessment of the effectiveness of the Community Health Strategy on health outcomes in Kenya. *BMC Health Services Research, 14*(Suppl 1), S3. doi: 10.1186/1472-6963-14-S1-S3

O'Leary, Z. (2004). *The essential guide to doing research*. London, UK: Sage Publications, Ltd.

Onwuegbuzie, A., & Johnson, R. (2006). The validity issue in mixed research. *Research in the Schools*, *13*(1), 48–63. Retrieved from videolectures.net/site/normal_dl/tag=48066/MixedMethodsaandValidity.RITS.pdf

Palinkas, L. A., Garcia, A. R., Aarons, G. A., Finno-Velasquez, M., Holloway, I. W., Mackie, T. I., ... Chamberlain, P. (2016). Measuring use of research evidence: The structured interview for evidence use. *Research on Social Work Practice*, *26*(5), 550–564. doi: 10.1177/1049731514560413

Penuel, W. R., Briggs, D. C., Davidson, K. L., Herlihy, C., Sherer, D., Hill, H. C., ... Allen, A.-R. (2017). How school and district leaders access, perceive, and use research. *AERA Open*, *3*(2), 1–17. doi: 10.1177/2332858417705370

Redman, S., Turner, T., Davies, H., Williamson, A., Haynes, A., Brennan, S., ... Green, S. (2015). The SPIRIT Action Framework: A structured approach to selecting and testing strategies to increase the use of research in policy. *Social Science & Medicine, 136–137,* 147–155. doi: 10.1016/j.socscimed.2015.05.009

Rihoux, B. (2006). Qualitative Comparative Analysis (QCA) and related systematic comparative methods: Recent advances and remaining challenges for social science research. *International Sociology, 21*(5), 679–706. doi: 10.1177/0268580906067836

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55. doi: 10.1093/biomet/70.1.41

Rubin, H. J., & Rubin, I. (2012). *Qualitative interviewing: The art of hearing data* (3rd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Sabatier, P. A. (1988). An advocacy coalition framework of policy change and the role of policy-oriented learning therein. *Policy Sciences*, *21*(2-3), 129–168. Retrieved from https://www.jstor.org/stable/4532139

Sackett, D. L., Rosenberg, W. M. C., Gray, J. A., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't. It's about integrating individual clinical expertise and the best external evidence. *British Medical Journal, 312*(7023), 71–72. doi: 10.1136/bmj.312.7023.71

Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., ... Jinks, C. (2018). Saturation in qualitative research: Exploring its conceptualization and operationalization. *Quality & Quantity, 52*(4), 1893–1907. doi: 10.1007/s11135-017-0574-8

Scheurich, J. J., & Young, M. D. (1997). Coloring epistemologies: Are our research epistemologies racially biased? *Educational Researcher, 26*(4), 4–16. doi: 10.2307/1176879

Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W. H., & Shavelson, R. J. (2007). *Estimating causal effects using experimental and observational designs* (Report from the Governing Board of the American Educational Research Association Grants Program). Washington, DC: American Educational Research Association. Retrieved from https://pdfs.semanticscholar.org/d735/151b5ab7fdb4e7b2f776f3d32c4d63b24b9a.pdf

Seidman, I. (2013). *Interviewing as qualitative research: a guide for researchers in education and the social sciences* (4th ed.). New York, NY: Teachers College Press.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Shearer, J. C., Dion, M., & Lavis, J. N. (2014). Exchanging and using research evidence in health policy networks: A statistical network analysis. *Implementation Science, 9*(1). doi: 10.1186/s13012-014-0126-8

Spillane, J. P., Reiser, B. J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research, 72*(3), 387–431. doi: 10.3102/00346543072003387

Tacoli, C., McGranahan, G., & Satterthwaite, D. (2015). *Urbanisation, rural–urban migration and urban poverty* (IIED Working Paper). London, UK: IIED. Retrieved from http://pubs.iied.org/10725IIED/

Tate, W. F. (1997). Chapter 4: Critical Race Theory and education: History, theory, and implications. *Review of Research in Education, 22*(1), 195–247. doi: 10.3102/0091732X022001195

Toulmin, S. (2003). *The uses of argument* (Updated ed.). New York, NY: Cambridge University Press.

Tseng, V. (2012). The uses of research in policy and practice. *Social Policy Report, 26* (2), 1-16.

Tseng, V. (2012). *The uses of research in policy and practice*. Washington, DC: Society for Research in Child Development. Retrieved from https://files.eric.ed.gov/fulltext/ED536954.pdf

Visser, P., Krosnick, J. A., & Lavrakas, P. J. (2000). Survey research. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 223–252). New York, NY: Cambridge University Press.

Wagner, A. K., Soumerai, S. B., Zhang, F., & Ross-Degnan, D. (2002). Segmented regression analysis of interrupted time series studies in medication use research. *Journal of Clinical Pharmacy and Therapeutics*, *27*(4), 299–309. Retrieved from https://www.alnap.org/help-library/segmented-regression-analysis-of-interrupted-time-series-studies-in-medication-use

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge, UK: Cambridge University Press.

Weber, M. (1904/1949). Objectivity in social science and social policy. In *The methodology of the social sciences*, translated and edited by E. A. Shils and H. A. Finch. New York, NY: Free Press.

Weinberg, M., Fuentes, J. M., Ruiz, A. I., Lozano, F. W., Angel, E., Gaitan, H., … Huskins, W. C. (2001). Reducing infections among women undergoing Cesarean section in Colombia by means of continuous quality improvement methods. *Archives of Internal Medicine*, *161*(19), 2357. doi: 10.1001/archinte.161.19.2357

Weiss, C. H., & Bucuvalas, M. J. (1980). *Social science research and decision-making*. New York, NY: Columbia University Press.

Wertsch, J. V. (1993). *Voices of the mind: A sociocultural approach to mediated action* (1st paperback ed.). Cambridge, MA: Harvard University Press.

White, H., & Sabarwal, S. (2014). *Quasi-experimental design and methods* (Methodological Briefs: Impact Evaluation 8). Florence, Italy: UNICEF Office of Research. Retrieved from https://www.unicef-irc.org/publications/753-quasi-experimental-design-and-methods-methodological-briefs-impact-evaluation-no.html

William T. Grant Foundation. (1986). *The William T. Grant Foundation: The first fifty years*. New York: William T. Grant Foundation.

William T. Grant Foundation. (2017). *Improving the use of research evidence: Supplemental guidance for applicants*. New York: William T. Grant Foundation.

Wulczyn, F., Alpert, L., Monahan-Price, K., Huhr, S., Palinkas, L. A., & Pinsoneault, L. (2015). Research evidence use in the child welfare system. *Child Welfare*, *94*(2), 141–165. Retrieved from https://eds.b.ebscohost.com/eds/pdfviewer/pdfviewer?vid=1&sid=fb2f043b-15ee-4249-ae04-e9292abb77f5%40pdc-v-sessmgr03

Yanovitzky, I., & Weber, M. (2018). *Analysing use of evidence in public policymaking processes: a theory-grounded content analysis methodology. Evidence & Policy* (advanced online publication). https://www.ingentaconnect.com/content/tpp/ep/pre-prints/content-ppevidpold1700095

Yin, R. K. (2012). *Applications of case study research* (3rd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Yin, R. K. (2014). *Case study research: Design and methods* (5th ed.). Thousand Oaks, CA: Sage Publications, Inc.

Zinn, M. B. (1979). Field research in minority communities: Ethical, methodological and political observations by an insider. *Social Problems, 27*(2), 209-219. doi: 10.2307/800369

Zwolsman, S. E., van Dijk, N., & de Waard, M. W. (2013). Observations of evidence-based medicine in general practice. *Perspectives on Medical Education, 2*(4), 196–208. doi: 10.1007/s40037-013-0078-8

Appendix:
Additional Resources

This appendix was developed to include additional URE studies that could not be included in the monograph, but that readers might find useful as examples of particular methodological approaches.

Interviews

Hyde, J. K., Mackie, T. I., Palinkas, L. A., Niemi, E., & Leslie, L. K. (2016). Evidence use in mental health policy making for children in foster care. *Administration and Policy in Mental Health and Mental Health Services Research*, *43*(1), 52–66. doi: 10.1007/s10488-015-0633-1

Reime, M. (2016). Discourses in residential child care and possibilities for evidence-based practice. *Professions and Professionalism*, *6*(1). doi: 10.7577/pp.1434

van de Goor, I., Hämäläinen, R.-M., Syed, A., Juel Lau, C., Sandu, P., Spitters, H., ... Aro, A. R. (2017). Determinants of evidence use in public health policy making: Results from a study across six EU countries. *Health Policy*, *121*(3), 273–281. doi: 10.1016/j.healthpol.2017.01.003

Surveys

Amara, N., Ouimet, M., & Landry, R. (2004). New evidence on instrumental, conceptual, and symbolic utilization of university research in government agencies. *Science Communication*, *26*(1), 75–106. doi: 10.1177/1075547004267491

Cooper, A., & Levin, B. (2013). Research use by leaders in Canadian school districts. *International Journal of Education Policy and Leadership*, *8*(7). doi: 10.22230/ijepl.2013v8n7a449

Landry, R., Lamari, M., & Amara, N. (2003). The extent and determinants of the utilization of university research in government agencies. *Public Administration Review*, *63*(2), 192–205. doi: 10.1111/1540-6210.00279

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bi-factor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*(2), 137–150. doi: 10.1037/met0000045

Williams, D., & Coles, L. (2007). Teachers' approaches to finding and using research evidence: An information literacy perspective. *Educational Research*, *49*(2), 185–206. doi: 10.1080/00131880701369719

Observation and Discourse

Huguet, A., Allen, A.-R., Coburn, C. E., Farrell, C. C., Kim, D. H., & Penuel, W. R. (2017). Locating data use in the microprocesses of district-level deliberations. *Nordic Journal of Studies in Educational Policy*, *3*(1), 21–28. doi: 10.1080/20020317.2017.1314743

Document Analysis

Godfrey, D. (2017). What is the proposed role of research evidence in England's 'self-improving' school system? *Oxford Review of Education*, *43*(4), 433–446. doi: 10.1080/03054985.2017.1329718

Social Network Analysis

Arling, P. A., Doebbeling, B. N., & Fox, R. L. (2011). Leveraging social network analysis to improve the implementation of evidence-based practices and systems in healthcare. In *Proceedings of the 2011 44th Hawaii international conference on system sciences* (pp. 1–10). Washington, DC: IEEE Computer Society. doi: 10.1109/HICSS.2011.285

Daly, A. J., & Finnigan, K. S. (2011). The ebb and flow of social network ties between district leaders under high-stakes accountability. *American Educational Research Journal*, *48*(1), 39–79. doi: 10.3102/0002831210368990

Jessani, N. S., Boulay, M. G., & Bennett, S. C. (2016). Do academic knowledge brokers exist? Using social network analysis to explore academic research-to-policy networks from six schools of public health in Kenya. *Health Policy and Planning*, *31*(5), 600–611. doi: 10.1093/heapol/czv107

Palinkas, L. A., Holloway, I. W., Rice, E., Fuentes, D., Wu, Q., & Chamberlain, P. (2011). Social networks and implementation of evidence-based practices in public youth-serving systems: A mixed-methods study. *Implementation Science*, *6*(1). doi: 10.1186/1748-5908-6-113

Yousefi-Nooraie, R., Dobbins, M., Brouwers, M., & Wakefield, P. (2012). Information seeking for making evidence-informed decisions: A social network analysis on the staff of a public health department in Canada. *BMC Health Services Research*, *12*(1). doi: 10.1186/1472-6963-12-118

Experiments and Quasi-Experiments

Lord, P., Rabiasz, A., Roy, P., Harland, J., Styles, B., & Fowler, K. (2017). *Evidence-based literacy support: The 'Literacy Octopus' trial*. Millbank, UK: Education Endowment Foundation. Retrieved from https://files.eric.ed.gov/fulltext/ED581248.pdf

Murphy, D. J., Lyu, P. F., Gregg, S. R., Martin, G. S., Hockenberry, J. M., Coopersmith, C. M., … Sevransky, J. (2016). Using incentives to improve resource utilization: A quasi-experimental evaluation of an ICU quality improvement program. *Critical Care Medicine*, *44*(1), 162–170. doi: 10.1097/CCM.0000000000001395

Palinkas, L. A., Saldana, L., Chou, C.-P., & Chamberlain, P. (2017). Use of research evidence and implementation of evidence-based practices in youth-serving systems. *Children and Youth Services Review*, *83*, 242–247. doi: 10.1016/j.childyouth.2017.11.005

Weinberg, M., Fuentes, J. M., Ruiz, A. I., Lozano, F. W., Angel, E., Gaitan, H., … Huskins, W. C. (2001). Reducing infections among women undergoing Cesarean section in Colombia by means of continuous quality improvement methods. *Archives of Internal Medicine*, *161*(19), 2357–2365. doi: 10.1001/archinte.161.19.2357

Case Studies-Mixed Methods

Allen-Meares, P., Hudgins, C. A., Engberg, M. E., & Lessnau, B. (2005). Using a collaboratory model to translate social work research into practice and policy. *Research on Social Work Practice*, *15*(1), 29–40. doi: 10.1177/1049731504272345

DeBray, E., Scott, J., Lubienski, C., & Jabbar, H. (2014). Intermediary organizations in charter school policy coalitions: Evidence from New Orleans. *Educational Policy*, *28*(2), 175–206. doi: 10.1177/0895904813514132

Elliott, H., & Popay, J. (2000). How are policy makers using evidence? Models of research utilisation and local NHS policy making. *Journal of Epidemiology and Community Health*, *54*(6), 461–468. doi: 10.1136/jech.54.6.461

Farley-Ripple, E., & Jones, A. R. (2015). Educational contracting and the translation of research into practice: The case of data coach vendors in Delaware. *International Journal of Educational Policy & Leadership*, *10*(2). doi: 10.22230/ijepl.2015v10n2a617

Judkins, M., Stacey, O., McCrone, T., & Inniss, M. (2014). *Teachers' use of research evidence: A case study of United Learning schools.* Slough, Berkshire: National Foundation for Educational Research.

Leslie, L. K., Maciolek, S., Biebel, K., Debordes-Jackson, G., & Nicholson, J. (2014). Exploring knowledge exchange at the research–policy–practice interface in children's behavioral health services. *Administration and Policy in Mental Health and Mental Health Services Research*, *41*(6), 822–834. doi: /10.1007/s10488-014-0535-7

Makkar, S. R., Brennan, S., Turner, T., Williamson, A., Redman, S., & Green, S. (2016). The development of sage: A tool to evaluate how policymakers' engage with and use research in health policymaking. *Research Evaluation*, *25*(3), 315–328. doi: 10.1093/reseval/rvv044

Penuel, W. R., Farrell, C. C., Allen, A.-R., Toyama, Y., & Coburn, C. E. (2018). What research district leaders find useful. *Educational Policy*, 32(4), 540–568. doi: 10.1177/0895904816673580

William T. Grant

FOUNDATION