OXFORD

Article

A Reviewer's Guide to Qualitative Rigor

Branda Nowell[*], Kate Albrecht[†]

*Professor, and †Doctoral Candidate, North Carolina State University, School of Public and International Affairs

Address correspondence to the author at blnowell@ncsu.edu

ABSTRACT

Institutions are useful for advancing methodologies within disciplines. Through required coursework, doctoral students are indoctrinated into basic guidelines and frameworks that provide a common foundation for scholars to interact with one another. Lacking such forums in many of our doctoral granting institutions (Stout 2013), the field of public management continues to struggle with an ambivalence toward qualitative approaches. Lack of shared understanding concerning basic tenets of qualitative methodology abounds. This article is intended for qualitative consumers, those not formally trained in qualitative methods but who serve as peer reviewers, content experts, and advisors in arenas where qualitative methods are encountered. Adopting a postpositivistic stance dominant in the field, we seek to offer a pragmatic perspective on qualitative methods with regards to some basic tenets of rigor appropriate (and inappropriate) for assessing the contribution of qualitative research. We argue that the first step in this effort is to stop conflating data type (qualitative versus quantitative) with inductive versus deductive modes of inquiry. Using deductive modes as the basis for comparison, we discuss both common, as well as, diverging criteria of quality and rigor for inductive modes of inquiry. We conclude with a discussion of rigor in emerging methods which utilize qualitative data but from within a deductive, mixed, or hybrid mode of inquiry.

Introduction

The field of public management continues to have a rocky relationship with qualitative methods. Like most methods, qualitative research has both its champions and its critics in the field. However, it is our sense that the majority of the field sits somewhere a bit right of center, open to a discussion but still suspect of what to do with findings from any study consisting of a small unrepresentative sample and no standard error. Much of this stems from fundamental misunderstandings about what qualitative inquiry is, and is not, designed to do. The cost of this to our discipline is significant. In a recent review, Ospina and colleagues (2017) reported only 7.5% of the articles published in top PA journals over the past 5 years relied solely on qualitative methods. This is not particularly surprising as our doctoral training institutions allow graduates to remain largely uninformed about qualitative approaches (Stout

2013). However, there are many questions germane to our discipline that are best suited to qualitative inquiry (for discussion, see Brower, Abolafia, and Carr 2000; Milward forthcoming, Ospina et al. 2017). In order to advance the contribution qualitative methods can make to the field, some foundational understanding about qualitative rigor is needed.

In embarking on this effort, we join an esteemed cadre of scholars who have grappled with the issue of qualitative rigor in public management (e.g., Brower et al. 2000; Dodge et al. 2005; Lowery and Evans 2004; Ospina et al. 2017). However, we seek a very specific audience. This is not an article written for the initiated qualitative scholar; we are not seeking to offer advancements in qualitative techniques or further the discourse on the precepts of qualitative inquiry. Nor is this an article particularly aimed at the edification of the novice qualitative scholar looking to embark upon

qualitative inquiry for the first time; there are many excellent texts out there that deal with the issues contained in this article in a much more thorough manner. Rather, this article was conceptualized and written primarily for the *qualitative consumer* who, at present, represents the over-whelming majority in the field of public management.

As we are envisioning our intended audience, three general categories of consumers come to mind. First, this article is for the quantitatively trained peer reviewer who finds themselves asked to assess the quality and contribution of a qualitative study brought to them for review. These folks serve as the gatekeepers and a quality assurance mechanism critical to the advancement of the discipline. Second, this article is for the scholar reviewing the literature within a content domain populated by both qualitative and quantitative studies. If we want qualitative research to have a greater substantive impact on the discipline, we need to give non-qualitatively trained scholars the tools to assess the contribution of qualitative research within their own research paradigm. Otherwise, citations will inevitably trend into methodological silos. Finally, this article is written for the quantitatively trained professor who finds themselves on a committee trying to support a student pursuing a qualitative or mixed method dissertation. We have a beloved colleague who routinely asks students whether their dissertations are going to be empirical or qualitative. Her intent is not to be pejorative; she simply has no frame of reference for how to think about quotations as data.

A Brief Note on Epistemology

We recognize that the writing of this article requires the adoption of some normative stances linked to the philosophy of science; namely, an epistemological stance that is primarily postpositivist in nature. We have intentionally deviated from normative practice in qualitative scholarship in minimizing our discussion of epistemology (for further discussions, see Creswell 2018; Creswell and Miller 2000; Raadschelders 2011; Riccucci 2010). This is not because we do not appreciate the value and relevance of alternative epistemological stances for the field of public management. However, many methods associated with qualitative rigor can be applied across different epistemological stances, varying in intention and orientation rather than practical execution.[1] For example, Lincoln and Guba's (1986) criteria of trustworthiness are useful. This is true regardless of whether you are utilizing those practices because you believe in the postpositivistic limitations

of humans to fully comprehend social processes present in natural settings, or because you believe these social processes are co-constructed in an inseparable relationship between the researcher and the participant. In a similar way, reflexivity[2] is relevant to both the postpositivist as well as the interpretivists regardless of whether you embrace the inseparability between the knower and knowledge (constructivism) or just view humans as fallible in part because they cannot fully and objectively separate who they are from the questions they ask and the answers they find (postpositivism; Guba 1990).

In this paper, we seek to offer a pragmatic perspective on qualitative inquiry with a focus on how to conceptualize and assess quality and rigor within a postpositivistic framework; the dominant philosophical stance of most qualitative consumers within public management. We do this with the aim of widening the pathways through which qualitative studies might influence scholarship in public management. We recognize that such an endeavor may be highly controversial within some branches of the qualitative community which maintain a strong allegiance to advancing a constructivist philosophy of science (e.g., Carter and Little 2007; Rolfe 2006). However, we argue it is neither reasonable nor necessary for qualitative consumers to suspend their fundamental view of reality in order to appreciate and assess the contribution of qualitative work to the broader field. There is a rich history of the integration of qualitative research within postpositivism (e.g., Clark 1998; Glaser and Strauss 2017; Prasad 2015; Yin 2017), particularly in the organizational sciences (Eisenhardt and Graebner 2007).

We do not foresee a reconciliation between constructivism and postpositivist philosophies occurring any time soon. However, we do see sizable opportunity for naturalistic, inductive qualitative inquiry to have a broader impact in the field of public management if we start from the perspective that both qualitative and quantitative methods are compatible and complementary buckets of tools within social science. Different tools are best suited for different jobs and there is

---

1 The method, methodology, epistemology coupling is a topic of considerable debate and concern in the field of qualitative methods (Corbin and Strauss 2014; Haverland and Yanow 2012; Ospina et al 2017).

Whether certain methods can or should be implemented by scholars embracing diverging epistemological stances is a topic warranting further discourse in the field of public management.

2 Reflexivity refers to the practice of being intentionally reflective about who you are both as a person situated within society and as a scholar professionally socialized within a cultural and institutional milieu. Specifically, reflexive practice calls upon scholars to consider how the totality of who they are as individuals influences the manner in which they approach scholarship, the questions they ask, the way the subjects of one's inquiry may react/respond, and how one interprets what they observe. This is done with an eye toward critically examining how these factors may shape and constrain what one "finds" (for discussion, see Pillow 2003).

almost as much variation within each bucket as there is between them. Regardless, the world is an increasingly complex place. As a discipline that routinely trudges off into some really messy domains of inquiry and holds itself accountable to informing practice as well as advancing theory (Brooks 2002; Denhardt 2001; Gill and Meier 2000; Head 2010; Weber and Khademian 2008), we need every tool we can get.

To address building this toolbox for qualitative consumers, we present first an overview of critical domains of inquiry in the field of public management where we see qualitative methods as being particularly well suited to advancing scholarship. This review highlights some of the most cited and celebrated theories of our field that have been initially shaped or meaningfully re-imagined from qualitative approaches. Next, we argue for a reframing of the question of qualitative rigor, asserting the more productive distinction lies in differentiating inductive versus deductive modes of inquiry. Leveraging this perspective, we discuss both commonalities and points of departure in appropriate criteria of quality and rigor between deductive versus inductive models. Finally, we discuss issues of rigor in three emerging methods in public management that use qualitative data in deductive, mixed and hybrid models of inquiry.

Where Qualitative Methods Shine

If qualitative methods are viewed as a category of tools, it is relevant to next consider some of the functionality one maximizes through the use of such tools. Although this list is not exhaustive, it is intended to provide a general grounding into the types of situations where qualitative approaches are particularly well equipped to make a contribution to the field of public management.

Advancing New Theory and Discovering Nuance in Existing Theory

Quantitative hypothesis testing requires *a priori* theory. Arbitrarily searching for significant correlations between variables in a dataset without a theoretically grounded hypothesis to direct the analysis is infamously problematic for well-documented reasons (Kuhn 1996; Steiner 1988). Theory is a combination of a premise as well as a well-explicated mechanism that explains the why behind the premise or proposition.

Cross-sectional quantitative designs can test the strength and nature of association between two or more constructs. Longitudinal quantitative designs can examine the patterning of association over time, and experimental designs can even narrow in on causality. These are powerful tools, but none are well equipped to discover the mechanisms by which these observed

patterns are operating or identifying intervening factors that explain inconsistencies across cases. We use existing theory to infer the mechanism associated with an observed pattern but this is generally not an empirical exercise, it is a conceptual one. Further, it often requires the extrapolation of theoretical mechanisms conceptualized in one organizational context (e.g., private firms) to be applied in a completely different organizational context (e.g., public organizations). When the hypothesized association holds, we generally conclude that the same mechanisms are in operation in the same manner. How critically do we look at this assumption? What else might be going on? Qualitative methods offer tools specifically designed to empirically shed light on these questions.

Qualitative methods are particularly useful in the theory development process because they are able to provide detailed description of a phenomenon as it occurs in context. These methods do not require the scholar to guess in advance the most important factors and their relationship to each other. Mechanisms associated with the co-occurrence of two phenomena can be observed in real time or described by first hand informants who experienced it. For example, Feldman's (e.g. Feldman 2000; Feldman and Pentland 2003) seminal work on the role of routines as sources of change and innovation in organizations was based on organizational ethnography. Some other classic examples of theory development in public management that began as qualitative research can be found in organizational culture and sense making case studies (Schein 2003; Weick 1993). Toepler's (2005) case study of a CEO in crisis, the phenomena of iron triangles (Freeman 1965), and the social construction of target populations (Schneider and Ingram 1993) are also illustrations of theoretical advances through qualitative inquiry. Additionally, a major contribution to theory of both formal and informal accountability in the public sector and multi-sector collaboration was a direct result of a grounded theory qualitative approach (Romzek and Dubnick 1987; Romzek, LeRoux, and Blackmar 2012; Romzek et al. 2014). All of these examples leverage a qualitative researcher's ability to harness an inductive approach that allows for the emergence of our understanding of the nature of phenomena from those organizations and people who experienced it.

Beyond advancing new theories, qualitative methods have a strong tradition of clarifying and expanding upon existing theory. Underpinning many public management research areas is the ever-present politics-administration dichotomy. Maynard-Moody and Kelly's (1993) foundational piece used a phenomenological approach to present the views of public workers who must navigate their administrative and

political responsibilities every day. Agency and stewardship theories have also been examined and further delineated using qualitative methods (Schillemans 2013). Theories of goal-directed networks and managerial tensions around unity and diversity have been expanded through qualitative studies (Saz-Carranza and Ospina 2010). Finally, the nature of public participation has been theorized and statistically tested, but along the way the notion of authentic engagement—described as "deep and continuous involvement…with the potential for all involved to have an effect on the situation" (p. 320) was introduced to clarify theories, in part as a result of King et al.'s (1998) qualitative study.

Developing New Constructs, Frameworks, and Typologies

Quantitative hypothesis testing and construct validation requires the conceptualization and suggested operationalization of a construct. The development or usage of a new measure is aptly treated with skepticism if it is not empirically and theoretically grounded. In this way, many variables that we quantitatively leverage could not exist without prior development through qualitative research. For example, a foundational idea, and the basis for subsequent quantitative considerations of the differences between managers and front-line workers, is rooted in Lipsky's (1971) examination and discussion of the street-level bureaucrat. Drawing from case studies and interviews, Lipsky highlights the nature of front-line worker discretion and challenges public management scholars to include this important context in future research.

Public Service Motivation (PSM), public management's very own and home-grown construct, was born from Perry and Wise's (1990) discussion citing both cases and quotes from public servants. Their argument for PSM to be more fully operationalized and then measured is rooted in their content analysis. Although they do not explicitly state the qualitative nature of their article, their argument for, and legacy of PSM scale measures, is drawn directly from the words and actions of public servants themselves.

Defining Mechanisms Underlying Statistical Associations

Although some quantitative articles do include mechanisms in their discussion sections, many simply rehash results and what hypotheses were or were not supported. Indeed, quantitative research in public management gives considerable weight to well-documented statistical association, even when the causal mechanism is ambiguous. In this world, how then do mechanisms get clarified when an association is found? This is an area where qualitative researchers have been working

with less recognition of the importance of their research striving to answer "how" and "why" questions. The literature mentioned here again is not an exhaustive list, but emblematic of some prime examples of how our field's understanding of a statistical result has been given more texture and a much richer application to both theory and practice through qualitative methods.

In the area of government contracting, Dias and Maynard-Moody (2007) further examine past quantitative findings that turn on Transaction Cost Economics (TCE) (Williamson 1981) by explicating how and why implementing competing contracting philosophies of agencies and service providers underpins the nature of the transaction itself. Another qualitative piece examining the deeper mechanisms behind TCE is Van Slyke's (2003) discussion of the "mythology" of contracting. In his research, data from semi-structured interviews suggests competition is not a simple construct in testing TCE interactions between governments and service providers because of the nature of environmental constraints, actions by nonprofit organizations, networked relationships, and government-enacted barriers have important dynamics. Honig (2018) offers another apt example in a mixed method study in which he demonstrates how comparative case study designs can reveal insights about the role of the environment in moderating the relationship between managerial control and success that were not possible to capture through quantitative modeling.

Forget Qualitative Versus Quantitative

We have observed many scholars get conceptually hung up on the numbers versus text dichotomy associated with qualitative versus quantitative traditions. Although it is true that qualitative methods generally involve the analysis of some form of text and quantitative methods always involve the analysis of numbers, this focus on data type is largely a distraction from the more important distinction of inductive versus deductive forms of inquiry (Eisenhardt and Graebner 2007). Deductive approaches to inquiry start with a general premise or proposition and then investigate whether this premise holds within a specific sample intended to represent a broader population. Inductive approaches start with a specific case or set of cases of theoretical importance and seek to describe a phenomenon of interest within that case in such a manner as to draw rich insight into that phenomenon (for discussion, see Eisenhardt and Graebner 2007; McNabb 2014). Although there are a handful of qualitative and/or hybrid qualitative/quantitative methods intended for deductive inquiry (more on this below), the bulk of tools in the qualitative bucket are intended for inductive inquiry.

Overarching Principles of Quality Between Inductive and Deductive Inquiry

Before we get into differences, it is important to first consider similarities. Although inductive and deductive traditions of scholarship differ in many important respects, they also share some commonalities that form the mutual basis of considerations of quality in terms of assessing their contribution to the literature. In our exuberance to elaborate their differences, we can forget what these forms of inquiry can hold in common. We argue that inductive and deductive approaches share in common three core values that are foundational to the notion of quality scholarship in public management: 1) the importance of scholarship that advances theory, 2) the principle of inquiry-driven design, and 3) the criticality of gap-driven inquiry.

Relevance of Scholarship for Advancing Theory

In public management, our focus is to inform practice as well as advance theory (Kettl 2000). As a result, we give the greatest currency to knowledge that has relevance beyond the boundaries of the specific case, individual, or instance. Thus, within our field, the degree to which findings can have relevance beyond the study case or sample is foundational to conceptualizations of quality regardless of inductive or deductive approach (Dubnick 1999). Inductive scholarship, different from most deductive studies, allows for a plurality of truths and an equifinality of pathways to the same outcome (Eisenhardt, Graebner, and Sonenshein 2016), but the same standards of quality still apply. In other words, in inductive approaches, one need not argue an observed finding is the only explanation for a given outcome observed in another space or time, but it must be a plausible explanation for a similar outcome given a similar set of circumstances (Holland 1986; Lewis 1973).

As such, both inductive and deductive studies are in the same boat of trying to figure out the extent to which and ways in which their limited study has broader implications for the field. The criteria and processes used to establish this element of quality certainly differs, but the precept that findings must have relevance beyond the scope of the data analyzed is common to both qualitative and quantitative scholarship in the field of public management (McNabb 2015).

Inquiry-Driven Design

Both inductive and deductive traditions are inquiry driven. This means that evaluating the quality of any design—qualitative or quantitative—is inseparable from understanding the research question the study is designed to address. It is possible to hammer a nail with a screwdriver, but it is not considered good practice as you are likely to just make a mess of it. In the same way, different research questions are more or less appropriate to different designs. Thus, while it is possible to attempt to describe the different ways in which people experience transformational leadership with an exploratory survey or use a series of focus groups to examine the relative prevalence of public service motivation among different groups, it is not a good practice as you are likely to just make a mess of it.

A common misconception is that inductive qualitative methods seek to ask and answer the same questions as quantitative methods, just using different types of data and standards of rigor. This is not the case. Inductive approaches are designed to pose and address *fundamentally* different kinds of questions that necessitate different types of data and criteria of rigor. However, methodological appropriateness (Haverland and Yanow 2012), or using the right tool for the job, is a value common to both inductive and deductive traditions and a key element of quality for all public management scholarship.

Gap-Driven Inquiry

Both inductive and deductive traditions recognize that knowledge does not advance in isolation—it takes a community of scholars to build a body of knowledge (Kuhn 1996; Gill and Meier 2000). The practice of positioning a research question in terms of its relevance within broader conversations that are taking place within the literature is mainstream to both traditions (McNabb 2015). In the field of public management—as elsewhere—the greatest currency is given to studies that clearly identify and address a significant gap within the literature; we seek to investigate something overlooked, under-appreciated, or potentially misunderstood in our current understanding of a given phenomenon. The extent to which a study accomplishes such a contribution is a shared element of quality for both deductive and inductive traditions.

Rigor: The Point of Departure

In the previous section, we have argued that inductive and deductive approaches in public management share a common foundation in conceptualizing the quality of inquiry. Specifically, we suggest quality can be conceptualized as inquiry that addresses a significant gap in the literature in a manner that advances our general understanding of a broader phenomenon through the use of a method appropriate to the nature of the research question. Rigor, then, can be conceptualized as the appropriate execution of that method. Put simply, if quality is the what, rigor for our purposes becomes the how. It is here that inductive and deductive traditions diverge in a significant way.

It is useful to start with the negative case. Two criteria appropriate for deductive research but NOT appropriate for inductive inquiry include:

1) Is there evidence that the causal factors, processes, nature, meaning, and/or significance of the phenomenon generalize to the broader population?
2) Are the findings able to be replicated in the sense that two researchers asking the same question would come to the same interpretation of the data?

These two criteria, held sacred as cornerstones of rigor in deductive inquiry, seem to cause the greatest amount of heartburn within the field of public management and its relationship to inductive qualitative inquiry. If it is not generalizable and it does not replicate, how is that possibly science? This results in on-going frustration among qualitative scholars as they attempt to respond to criticisms of their design by reviewers, colleagues, and advisors in terms of the lack of representative sampling and/or inter-rater reliability measures. This is rooted in some fundamental misunderstandings about what inductive inquiry is and what it seeks to accomplish.

Generalizability

In deductive methods, when there are more cases that conform to an *a priori* hypothesis than do not, relative to the standard error and controlling for all other factors in the model, we reject the null hypothesis that this pattern could have been observed merely by random chance. However, in every deductive sample, there can be numerous observations which do not conform to our models. These we vaguely disregard as "error." When cases deviate substantially, we call them "outliers" and may remove them from consideration entirely. This is reasonable because the aim of deductive inquiry is to test the presence of an *a priori* relationship in the population based on a limited, representative sample (Neuman and Robson 2014). Associations deal with probabilities and likelihoods; not all cases must conform to a pattern to conclude that an association exists as long as the sample is reasonably representative and sufficient to detect differences (Wasserman 2013).

Inductive research is attempting to do something quite different. The sample of an inductive study is *never purely* random nor convenient. Instead, each case or participant should be purposively selected because they represent a *theoretically interesting* exemplar of, or key informant about, a phenomenon of interest (Patton 2014). In other words, by nature of being selected for inclusion in an inductive study, the scholar is making the argument that we should care about understanding the experience of this person(s) or the events of this case. Whether a pattern discerned in an inductive study is common in the general population is not the question an inductive scholar is seeking to

answer. In fact, the case may have been selected specifically because it represents something rare or unusual. Rather, they are seeking to use a systematic method to interpret and represent, in rich detail, *what is true* for a particular set of individual(s) and/or cases, identifying themes and patterns across cases that add insight into the phenomenon of interest. Cases with divergent patterns or informants with contradictory experiences are not ignored or discounted as measurement error or outliers. Rather, the inductive scholar seeks to understand the factors and mechanisms that explain these points of divergence (Eisenhardt et al. 2016).

Although the inductive scholar does not empirically test the extent to which an association or experience is common in the general population, this does not mean that inductive findings are not intended to have relevance for advancing general theory and practice. If done well, an inductive study should provide a detailed, contextualized, and empirically grounded interpretation of what was true in one or more cases of interest. Just as one experience in one setting should never be assumed to dictate what one might experience in another setting, it would likewise be absurd to assume prior experience is totally irrelevant if a similar set of conditions are present. In this way, qualitative inductive scholarship seeks to systematically describe and interpret what is occurring in a finite set of cases in sufficient detail as to lend insight into what might be going on in cases *like these*. Discerning the quantitative prevalence of these key patterns or experiences within populations is where deductive methods can pick up where inductive leave off. However, it is only through also gaining a grounded and detailed understanding of phenomenon of theoretical interest do we gain new insights and have hope of developing understanding and theory that has relevance to field of practice.

Replication

As mentioned previously, inductive methods are seeking to develop a deep understanding of causal factors, processes, nature, meaning, and/or significance of a particular phenomenon (Creswell and Poth 2018; Denzin and Lincoln 2012; Patton 2014). This understanding generally comes from asking a lot of questions, observing settings and behavior, and collecting stories, images, and other artifacts that aid the scholar in also gaining insight into their phenomenon of interest. Different approaches have been created to narrow in on specific types of phenomenon. For example, phenomenology looks at how individuals experience and ascribe meaning to a given phenomenon (Giorgi 1997; Moran 2002; Waugh and Waugh 2003). Grounded theory seeks to identify the causal relationships that give rise to, and result from, a given phenomenon (Glaser and Strauss 2017; Morse et al. 2016). Ethnography

seeks to uncover the cultural elements within human systems (Agar 1996; Hammersley 1983; Preissle and Le Compte 1984).

Each tradition has its own systematic process of data collection and analysis. However, regardless of the tradition, *it is always the analyst who must draw inference and interpretation from the vast array of qualitative information in front of them*. Just as there are some doctors who can observe the same patient information to diagnose root causes while others focus on first order symptomology, multiple analysts working independently on the same data sources may also come to different interpretations of what is going on (Langley 1999). One doctor is not necessarily right and the others wrong; rather the same thing can be many things at once (e.g., structural, psychological, cultural). Therefore, the appropriate criteria of rigor is not whether the same interpretation would be independently arrived upon by different analysts. Rather, in inductive analysis, the criteria is: based on the evidence provided, is a given interpretation *credible* (Patton 1999)? In other words, if an independent analyst were informed of another analyst's interpretation and then given all the same source information, would the interpretation stand up to scrutiny as being a justified, empirically grounded, exposition of the phenomenon?

Elements of Rigor

If we cannot assess inductive studies in terms of generalizability and replication, what are valid criteria upon which they might be evaluated? In very global terms, rigorous inductive research in public management can be judged on two core criteria:

1) Does the research design and its execution generate new insight into the causal factors, processes, nature, meaning, and/or significance of a phenomenon of interest to the field? (reviewed in Table 1) and

2) Is the account of these causal factors, processes, nature, meaning, and/or significance *within these cases* trustworthy? (reviewed in Table 3)

The trustworthiness and depth of insight of an inductive study is manifest in its research design, execution, reporting.

Research Design

Sampling

Because the contribution of inductive qualitative research fundamentally hinges on the theoretical relevance of the units (e.g., individuals, cases, texts) selected for study, sampling is of paramount importance. Different approaches of qualitative analysis have specific guidance on sampling consistent with that approach. For example, grounded theory uses a protocol of proposition-driven sampling in which the investigator strategically chooses cases iteratively in conjunction with data analysis in an effort to examine variation in patterns observed in the previous cases (for discussion, see Corbin and Strauss 1990; Glaser 2002). However, regardless of which analysis tradition an inductive scholar is using, the inductive qualitative sample must *always* be justified in terms of why the informants, texts, and/or cases selected should be considered of theoretical interest to the field. This description should be situated in terms of who these informants are in the broader population of possible informants relevant to the research question. Inductive scholarship should include a clear explication of why these individuals were chosen specifically and what they represent. What qualifies them as key informants of this phenomenon? Why would we expect them to have insight into this question that is particularly information rich and/or relevant to the field? How might their position likely influence the perspective they offer about the phenomenon (for discussion, see Marshall 1996; for exemplar, see Saz-Carranza and Ospina 2010 and Romzek et al. 2012 justification of both case and informant selection)?

As outlined in most introductory texts in qualitative analysis (e.g., Denzin and Lincoln 2012; Miles, Huberman, and Saldana 2013; Patton 2014), there are numerous sampling strategies that may guide participant or case selection in an inductive study. Common approaches include efforts to capture the "typical case," the "extreme case," the disconfirming case, or the "unusual case." Sampling is also often purposefully stratified to represent informants from theoretically important sub-populations. In studies of individual level phenomenon, this may include stratifying samples to include men and women, young/middle age/old, more or less experience, or different ethnicities/racial groups. In studies of higher order phenomenon such as at the organizational, coalition, group, or network level, the scholar may choose to stratify cases across geographic region or based on some developmental phase (e.g., new versus old organizations). Although there are numerous potential sampling strategies for an inductive study, they all share in common the criteria that whatever or whomever is chosen for inclusion or exclusion of an inductive study, sampling decisions must be inquiry driven, theoretically justified, and information rich.

How Many is Enough?

The question of sample size in inductive qualitative research is less straight forward than it is in deductive research. In the deductive world, the sample size criteria turns primarily on the power to detect differences given the model applied to the data (Wasserman 2013).

Table 1. Relevant and Inappropriate Criteria of Rigor for Inductive Research

	Relevant Criteria of Quality and Rigor	Inappropriate Criteria of Quality and Rigor
Significance of the research question	• Is the phenomenon of interest important to advancing the field? • Does the research question address a significant gap in the literature?	• Is the research question theoretically grounded in such a way as to inform hypothesis development?
Justification of the research approach	• Is the research question inductive in nature? • Does the research question aim to understand qualities, patterns, and mechanisms?	• Does the research question aim to understand likelihood or probability within defined populations?
Sampling	• Is there a clear and compelling justification that the cases/units/informants selected for inclusion in the study are information rich in relation to the phenomenon of interest? • Is the sampling approach justified within the analysis tradition utilized (e.g., grounded theory versus ethnography)? • Is the number of units sampled theoretically robust in both qualitative depth as well as representing important aspects of potential variability?	• Is the sample representative of the population?
Data collection	• Is the data collection protocol clearly described? • Do the elements of the protocol map clearly back to study objectives and analysis tradition? • Were all members of the research team trained on the study objectives and its relationship to the data collection protocol in order to ensure a high degree of fidelity in implementation? • Did research team members debrief regularly to review data quality in relation to study objectives and make mid-course adjustments to protocols as needed?	• Have the measures been quantitatively validated? • Were the protocols executed in a highly standardized manner across all participants to minimize method bias?
Data analysis	• Was the analysis approach clearly articulated and consistent with the analytic tradition employed? • Were practices to establish the trustworthiness of the analysis employed? (see Table 3)	• Was there adequate inter-rater reliability across coders?
Reporting	• Are interpretations presented in a credible manner that offers "thick description" (see Table 3) of the key themes or patterns identified including inclusion of relevant examples, representative quotations, and attention to non-conforming cases? • Are sources of potential bias and alternative explanations considered and reported? • Are the interpretations insightful? Do they address the research question in such a way as to advance our understanding of a phenomenon? • Are interpretations sufficiently contextualized so as to allow the reader to consider the types of situations in which such patterns might replicate?	• Are the findings replicatable? • Do the findings generalize to the population?

In inductive research, the sample size question focuses on the sources of variability of the phenomenon of interest that are of theoretical importance to the field given the research question. However, inductive studies complicate the sample size question because numerous and varied sources of data can be, and often are, integrated. For example, in several qualitative approaches, triangulation of findings among multiple data sources is one of elements of the rigor (e.g., for review see Jonsen and Jehn 2009).

Just as with deductive research, no one inductive study can address every dimension or perspective that might be relevant to understanding a phenomenon of interest. Therefore, in addition to clearly articulating the criteria upon which individuals or other data sources were sampled for inclusion into the study,

there is need to explicate the boundary criteria that sets the limits for who or what is not considered within the scope of the inquiry. Following this, the authors must clearly articulate the unit or units of analysis that define the phenomenon of inquiry. Is it case-based such as an inquiry into the factors that hindered international NGO community from being effective contributors to the response phase of Hurricane Katrina (e.g., Eikenberry, Arroyave, and Cooper 2007)? Is it organizational such as a study of service providers' usage of monitoring tools based on agency theory (e.g., Lambright 2008). Is it focused on the individual, such as examining public service motivation and transformation leadership (e.g., Andersen et al. 2016)? Or is it episodic such as a study of the process through which routines can lead to a source of innovation within an organization (e.g., Feldman 2003)? Higher order phenomenon (i.e., case-level, coalition-level, organizational-level, etc.) often require multiple data sources or informants associated with that case, group, or organization to gain sufficient depth of understanding of the dynamics present. This will necessarily place limits on the number of cases that can be studied comparatively. Alternatively, a single informant may be able to reflect on multiple episodic units based on varied experiences over time.

Qualitative Saturation

Qualitative saturation is a technique commonly referenced in inductive research to demonstrate that the dataset is robust in terms of capturing the important variability that exists around the phenomenon of interest (O'Reilly and Parker 2013). However, we advise caution in the use of saturation in defending the sample characteristics of a qualitative sample. Qualitative saturation refers to a point at which the analyst has obtained a sort of information redundancy such that continued analysis has revealed no new insight not already captured by previous cases (Morse 1995). Generally, during analysis, scholars do reach a point at which no new themes or propositions emerge and analysis of new transcripts leads only to additional instances of existing themes or relationships. However, this standard is problematic as a criterion for rigor in public management for two reasons.

First, in order to be used as a condition of sampling rigor, it requires that the scholar analyze their data as it is being collected so as to recognize the point at which no additional data collection is needed. Although this design feature is integral to grounded theory, it is uncommon in other qualitative traditions which often mimic deductive models having a distinct data collection phase preceding a data analysis phase (Miles, Huberman, and Saldana 2013). Second, the

methods by which a scholar determines saturation are generally methodologically difficult to standardize or demonstrate as a criteria of rigor (Morse 1995; O'Reilly and Parker 2013). Therefore, while saturation is an important heuristic in guiding data analysis—for example, for informing the analyst when they should transition from open coding to axial coding, we do not find it is a particularly useful concept for qualitative consumers to evaluate the suitability of a dataset in terms of whether it should be considered theoretically robust.

Consequently, qualitative consumers generally must rely on qualitative as opposed to quantitative benchmarks for determining the suitability of a given dataset for addressing an inductive research question. The questions qualitative consumers need to answer are these: 1) is the dataset analyzed information rich and 2) does it have a reasonable chance of representing variability of the phenomena of interest that are of theoretical importance given the research question (Brower, Abolafia, and Carr 2000)? In efforts to orient new inductive scholars into the general ballpark of sample expectations, some scholars have cautiously made heavily caveated recommendations (for review, see Onwuegbuzie and Leech 2007). Qualitative studies of 100 or more units are unusual and generally unnecessary for most inductive analytic traditions unless some type of quantification is desired (see below discussion on hybrid designs; Gentles et al. 2015). Studies of ten or less units would require a unique justification in terms of how such a data set provides a theoretically robust perspective on the phenomenon of interest. Within that sizeable range, qualitative consumers will have to make a subjective call about the theoretical robustness of a given dataset in relation to the research question asked, the phenomenon of interest, the analytic tradition used, and the interpretive claims made. Benchmarking sampling approaches against existing literature utilizing the same analytic approach is helpful for creating consistency within the field. Additionally, qualitative consumers may find the following questions a useful rubric in determining how theoretically robust a given dataset might be considered to be:

1) Is the phenomenon rare or infrequently encountered?
2) Are the data rare or particularly difficult to obtain?
3) Is the phenomenon simple or complex?
4) Is the phenomenon new or well investigated in the literature?
5) How information rich is each unit in relation to the phenomenon of interest?
6) Is the same unit being engaged at multiple points in time?

Data Collection Protocols and Procedures

In deductive research, constructs and relationships are articulated prior to analysis, and what one can discover is therefore necessarily constrained to what one looks to find. In inductive research, constructs and relationships are articulated through analysis, and the scholar seeks to minimize constraint on what can be discovered (Lincoln and Guba 1986). However, because in most inductive studies, the data must still be collected from individuals, the actions of the investigator will inevitably constrain and shape what the data looks like. This is done *a priori* through the creation of protocols which guide the types of questions that the investigator asks informants or the elements the investigator observes and records their observations. While these protocols can, and often should evolve over the course of the study, it is the execution of these protocols that create the data used in analysis. Consequently, the quality of these protocols and their execution is an important consideration in determining the rigor of an inductive study (Miles, Huberman, and Saldana 2013).

In demonstrating the rigor of an inductive research design, the investigator should be able to clearly describe what data was considered relevant for a given research question and how this data was obtained. Data collection protocol design should be consistent with the specific methodological tradition embraced by the study (see Table 2). Vague descriptors such as "data were obtained through open ended interviews" is not sufficient description to determine rigor. Just as in deductive research the same construct can be operationalized in multiple ways, two inductive investigators may be interested in the same research question but ask very different types of interview questions of their informants. Researchers should be able to describe the types of questions the investigator asked informants related to the phenomenon of interest. These questions should have a clear conceptual linkage to the research question of concern, the analytic tradition embraced, and be a key consideration in the analysis and interpretation of the findings (for exemplar, see Rerup and Feldman's (2011), description of the interview protocol used to illicit espoused schemas of staff in a tech start up). It is also important for the qualitative consumer to recognize that data looks different depending on the different analytic tradition one uses. Table 2 outlines some of the more prevalent qualitative traditions.

Data Analysis and Interpretation

Like deductive approaches, inductive qualitative data analysis come in many forms linked to different analytic traditions and are more or less appropriate to different types of research questions. These traditions carry with them specific guidance on design, sampling, and analysis. Methodological deviations or qualitative "mixology"

(Kahlke 2014) in which design element from multiple traditions are combined or certain design elements omitted should be well-justified and evaluated carefully by the qualitative consumer to ensure the resulting design remains robust. Just as with deductive designs, robust inductive designs should have a clear logical flow from the research question, to the data collection protocol, to the description of the analysis procedure, to the explication of the findings. There should be no black curtain behind which hundreds of pages of transcripts are magically transformed into seven key findings. Rather, the scholar should be able to provide a clear and concise description of their analysis process and its relationship to the reported findings (for exemplar, see Rivera's (2017) case analysis description in her study of gender discrimination in academic hiring committees).

As discussed, the overarching criteria of rigor associated with an inductive study is not reliability or replication. Rather, rigorous analysis is based on 1) whether the interpretation is credible in light of the data, 2) whether it was the result of a robust and systematic analytical process designed to move beyond superficial findings and minimize and/or account for investigator bias, and 3) whether it is reported with sufficient attention to context so as to facilitate the potential relevance of insights to similar contexts. These features were first described by Lincoln and Guba (1986) as the criteria of qualitative trustworthiness. They developed an initial set of practices designed to achieve these elements of rigor that have since been expanded upon by various qualitative scholars. Although these elements remain under development and debate, especially in public management (for discussion see, Lowery and Evans 2004), Table 3 offers a broad overview of some of the more commonly advocated strategies and associated aims that qualitative consumers might consider when evaluating the rigor of an inductive study. However, it is important to note that these elements represent strategies. They are not a checklist and certain strategies may be more or less appropriate in certain study designs. As such, we argue rigor is best conceptualized in terms of its functionality. Was the design logically coherent in relation to the research question? Was the analysis systematically designed to move beyond superficial findings and minimize and/or account for investigator bias? Did the design result in both credible and insightful findings? Were the findings reported with sufficient attention to context so as to facilitate empirically grounded theory building?

Deductive, Mixed, and Hybrid Qualitative Methods

We have argued that the distinction between inductive versus deductive approaches is a most relevant delineation for identifying appropriate criteria of rigor. Up

Table 2. Qualitative Data Collection and Analysis Traditions

	Typical Data	Analytic markers	Aim	Useful references
Grounded theory	Transcripts of interviews collective in phases	Iterative phases of data collection and analysis; theoretical sampling, constant comparison analysis	Develop empirically-grounded propositions concerning key elements and their associated relationships associated with a phenomenon	Glaser and Strauss 2017; Morse et al. 2016
Case Study/ Comparative Case Study	Multiple sources of data; mixed quantitative and qualitative	Theoretical or inductive content analysis Descriptive quantitative analysis Data synthesis	Understand the complexity of a specific case [or set of cases] in a holistic and comprehensive manner	McNabb 2015; Yin 2017
Descriptive Qualitative Inquiry	Any text-based data: transcripts of interviews or focus groups, printed material; open responses from surveys	Inductive thematic content analysis	Describing the outward content of experiences, opinions/ attitudes, or occurrences	Caelli et al 2003; Kahlke 2014
Ethnography	Observational field notes, cultural artifacts, formal and informal interviews	Ethnographic analysis	Describing cultural elements and their significance within a specified social setting	Agar 1996; Hammersley 1983; Preissle and Le Compte 1984
Phenomenology	Transcripts of in depth interviews	First order accounts of experience, often over time; compared across informants	Describing how humans experience and make sense of lived experience	Giorgi 1997; Moran 2002; Waugh and Waugh 2003
Narrative Analysis	Observational field notes, transcripts of in depth interviews, and written accounts	Inductive creation of conceptual groupings of stories and narrative elements	Describing how humans use elements of storytelling to express sequence and consequence; Can include "what" elements are present and "how" story is being told	Feldman et al. 2004; Ospina and Dodge 2005; Riessman 1993

Table 3. Elements of Qualitative Rigor (Adapted From Creswell and Poth 2018; Denzin and Lincoln 2003; Lincoln and Guba 1986; Morse 2015)

Phase	Strategy	Function
Data collection	Prolonged engagement with the setting	Develop a deep understanding of the context and phenomenon of interest
	Ask the same question in several ways and/or from multiple sources	Increase the depth of understanding about the phenomenon and triangulate findings
	Use multiple investigators participating in data collection	Reduce investigator bias and increase depth of insight brought to bear
	Peer debriefing and documentation during data collection	Avoid myopic initial interpretations; inform mid-course adjustments in protocols to improve data quality
	Collect data in phases, iterating between analysis and data collection (required for grounded theory- but recommended generally)	Inform mid-course adjustments in protocols to improve data quality; create the opportunity to explore more deeply into emerging themes or relationships
	Member check initial interpretation with informants during interviews through interpretative paraphrasing	Provide informal testing of the accuracy of investigators interpretations of what is being said
Analysis	Consensus based coding with multiple coders	Avoid myopic interpretations; assess the confirmability of interpretations
	Look for alternative interpretations and negative cases; describe prevalence of converging and diverging patterns	Demonstrate the robustness of a given finding or interpretation in light of alternatives
	Triangulate findings across different data sources	Demonstrate the robustness of a given finding or interpretation
	Analysis mapping and external audit	Establishing a clear audit trail of interpretative steps from raw data to final interpretation; use of external reviewer to provide feedback on data analysis steps or products
	Member check findings with setting insiders	Establish the credibility of data interpretations
Reporting	Report audit trail and other design features aimed at improving rigor of design and analysis	Establish dependability and confirmability of data
	Provide "thick description" of research context and informants	Allow readers to make judgements about the types of contexts in which findings may transfer
	Description of major findings are "grounded" in examples and/or illustrative quotations/text excerpts	Allows readers to make independent confirmations concerning the credibility of interpretations based on excepts from the raw data

to this point, we have focused primarily on inductive applications of qualitative data. However, as noted previously, not all qualitative data analysis is inductive. In this final section, we give special consideration to qualitative approaches in the field of public management that that are either deductive, mixed, and hybrid methods.

Narrative Policy Framework

In policy process research, the Narrative Policy Framework (NPF) has more recently emerged as an approach for quantifying qualitative data that has been coded from policy documents and various mediums of public comment (Shanahan, Jones, and McBeth 2013). The NPF was designed to address postpositivist challenges to policy process theories by taking into account the critical role that narratives play in generating and facilitating meaning for people and how those

narratives then relate to the politics of constructing reality (Shanahan, Jones, and McBeth 2013). Within the NPF, narratives are considered to be constructed of important elements that include many traditional parts of stories like a hero, a villain, a plot, and a moral. These narrative elements are applied as codes in a more directly deductive approach and then often used for hypothesis testing at *micro*, *meso*, and *macro* levels (McBeth, Jones, and Shanahan 2014).

Despite being derived from qualitative data, much of the work on NPF embraces a deductive model of hypothesis testing (Shanahan, Jones, and McBeth 2017). In deductive applications, the standards of rigor as it relates to representative sampling, construct validity, reliability, statistical power, and generalizability apply. These methods require the development of a stable coding framework that can be applied by multiple coders with a high degree of reliability. As such,

metrics such as inter-rater reliability are appropriate tools for demonstrating that the coding framework is being applied in a consistent manner. Another design challenge with NPF is the fact that its core propositions are associated with discrete "narratives" as the unit of analysis, which can be difficult to isolate in a standardized way across different types of policy documents which may contain multiple narratives (for discussion, see Shanahan, Jones, and McBeth 2018). Further, the representative sampling of policy documents relative to a defined population can be difficult to conceptualize (Shanahan, Jones, and McBeth 2018). Despite these challenges, NPF is valuable in its ability to examine whether specific narrative patterns have a stable and generalizable influence on different outcomes of the policy process (McBeth, Jones, and Shanahan 2014); a question ill-suited to an inductive narrative analysis approach.

Mixed Methods

Another development that has gained popularity in public management and applied social sciences more generally is the mixed methods study (see Honig, this issue). A mixed methods study is often characterized as one that uses a combination of both qualitative and quantitative data (Creswell and Clark 2018; for alternative definitions see Johnson et al. 2007). It is generally assumed that mixed methods studies will also utilize a combination of inductive and deductive approaches. The ordering of the inductive/deductive mixture can vary. For example, the scholar may use an inductive qualitative phase aimed at gaining a greater insight about a poorly understood phenomenon. Constructs, dimensions, and propositions resulting in the findings from this first inductive phase of analysis can then be translated into a second confirmatory phase in the form of survey measure development, psychometrics, and hypothesis testing. In a second variation, a scholar may use existing literature and theory to deductively create measures and propose and test hypotheses. The scholar may then design an inductive phase in which the mechanisms and contextual factor underlying these hypotheses are explored in great depth through qualitative methods (for discussion of various design options, see Mele & Belardinelli, this issue; Creswell, Clark, Gutmann, and Hanson 2003).

Considerations of rigor in a mixed methods study are two pronged. First, mixed methods studies have the dual burden of adhering to all the requirements of rigorous design associated with both inductive and deductive models. For example, the sample for the inductive phase must meet the criteria of offering an information rich, inquiry-driven sample while the sample for the deductive phase must have still sufficient power to detect differences and be a reasonably representative sample of the population. This generally makes such studies relatively large and ambitious. Second, a rigorous mixed methods study should ideally reflect some degree of complementarity between the approaches, maximizing the different advantages in inductive versus deductive designs. Each design element should reflect thoughtful attention to the points at which the findings from the different phases of analysis co-inform one another (Johnson, Burke, and Onwuegbuzie 2004).

Qualitative Comparative Analysis

Qualitative Comparative Analysis (QCA; Ragin 1998; Ragin and Rihoux 2004) represents a hybrid approach, being neither fully inductive or deductive. QCA has an established presence in public management (Cristofoli and Markovic 2016; Hudson and Kuhner 2013; Malatesta and Carboni 2015; Pattyn, Molevald, and Befani 2017; Raab, Mannak, and Cabre 2015; Sanger 2013; Thomann 2015). Like NPF, QCA involves the quantification of qualitative data and the application of mathematical models. However, different from NPF, which is principally deductive in its approach, QCA can use inductive qualitative methods to identify outcomes of interest and factors of relevance to explaining that outcome. These interpretations of the data are then quantified and entered into mathematical models designed to examine pathways of necessary and sufficient conditions that are derived from a researcher creating a numeric data table, often using binary codes.

QCA, first introduced by Ragin (1987), is intended to unify aspects of qualitative, case-based research, and quantitative, variable-based, approaches (Fischer 2011). QCA is rooted in the assumption of equifinality; that different causal conditions can lead to the same outcome, and that the effect of each condition is dependent on how it is combined with other conditions (Fischer 2011; Ragin 1987). Accordingly, QCA is not hindered by the assumptions of homogeneous effects that encumber many quantitative approaches. Rather, it enables the researcher to consider multiple pathways and combinations that may lead to the same outcome. Also unique to QCA is the focus on combinatorial logic that assumes that cases should be viewed holistically within the context of all conditions combined. As such, QCA can reveal patterns across cases that might be difficult to discern through purely qualitative approaches (for discussion see Rihoux and Ragin 2008).

One of the challenges to assessing QCA from a rigor perspective stems from its inherently hybrid nature. The samples in QCA are generally small and presumably inductively selected (Hug 2013). As such, an inductive criteria of rigor could apply. However, the results of a QCA have a distinctive deductive flavor in both the style of analysis and interpretation. For example, the process by which the specific constructs are identified for inclusion is often not well explicated and may contain a mixture of a priori theory and inductively

derived theory. Some authors embrace a fully deductive hypothesis driven approaches based on theory and using predetermined codebooks (e.g., Raab, Mannak, and Cambre 2015; Thomman 2015). Cases, which do not fit into one of the identified pathways are excluded from the output due to criteria like relevancy and consistency that enable the Boolean algebra of QCA to more readily converge on causal pathways.[3] Publications of QCA findings generally focus primarily on the pathways identified with little or no attention to the cases that deviated from these patterns.

It is our belief that as QCA applications evolve, scholars will need to, metaphorically, pick a horse to ride in their utilization of this technique in order for a study to be associated with the appropriate standards of rigor. In other words, QCA is a descriptive tool that can be used either inductively or deductively. Is a study a deductive effort to examine possible configurations of pathways toward a predefined outcome using a priori factors examined within a representative sample of a population? If so, deductive criteria of rigor would apply to a QCA as it relates to construct validity, inter-rater reliability, and representative sampling. On the other hand, QCA could also be a powerful tool used within an inductive model of research with associated inductive criteria of rigor. In this model, cases would be purposely justified as theoretically important to understanding a given phenomenon. The QCA would represent a tool within a broader process of inquiry for examining complex patterning across cases that may be difficult to otherwise discern. The inductive process by which coding categories were generated and qualitative variability that exists within coding delineations would be central concerns of the analysis. The analysis would include an empirically grounded contextualization and interpretation of the cases that conform to, as well as deviate from, the identified patterns so as to inform the mechanisms by which one pattern versus another may emerge. Either application of QCA, whether deductive or inductive, holds promise as a technique but murky applications which do not fully commit to either standard of rigor seem problematic (for additional discussion, see Hug 2013).

Conclusion

We began this article with the assertion that qualitative methods are poised to make a greater contribution in shaping our understanding of public management. We view this as a good thing; having the potential to inject new insight and depth of understanding into the questions that define the field. We perceive a general openness in the discipline to advancing bodies of literature through the integration of contributions from both inductive and deductive styles of inquiry. However, much of the discipline lacks even basic training in inductive approaches to research (see Stout 2013) which serves as a barrier. Deductive models—by virtue of the more structured task they are designed to accomplish coupled with the greater duration of time this approach has had to institutionalize—are simply more straightforward in their precepts of rigor. However, advancing the contribution of qualitative methods in public management will not happen without some shared construction of rigor that is compatible with a postpositivistic stance on science. We argue that the first step in advancing this agenda is to stop conflating data type (qualitative versus quantitative) with methodological approach (inductive versus deductive).

Beyond this, this article is positioned as a conversation-starter and as a resource for breaking down barriers for meaningful interactions that have put qualitative and quantitative methods at odds. We argue here that these past misunderstandings have less to do with the analysis of text versus number-based data, and more to do with murky or altogether misunderstood differences between the requirements of quality and rigor for inductive versus deductive methods. In clearing some of the air on quality and rigor of both kinds of methods in this space, we put forth a postpositivist stance with the understanding that not all scholars will agree, but that this perspective offers a productive pathway for broadly engaging the most common public management researcher today.

References

Agar, Michael H. 1996. *The professional stranger: An informal introduction to ethnography*. 2nd ed. UK: Emerald Publishing Group.

Andersen, Lotte Bøgh, Bente Bjørnholt, Louise Ladegaard Bro, and Christina Holm-Petersen. 2016. Leadership and motivation: A qualitative study of transformational leadership and public service motivation. *International Review of Administrative Sciences*: 1–17.

Brooks, Arthur C. 2002. Can nonprofit management help answer public management's "big questions"? *Public Administration Review*, **62** (3):259–66.

Brower, Ralph S., Mitchel Y. Abolafia, and Jered B. Carr. 2000. On improving qualitative methods in public administration research. *Administration & Society* **32** (4):363–97. doi:10.1177/00953990022019470.

Carter, Stacy M., and Miles Little. 2007. Justifying knowledge, justifying method, taking action: Epistemologies, methodologies, and methods in qualitative research. *Qualitative Health Research* **17** (10):1316–28.

Caelli, Kate, Lynne Ray, and Judy Mill. 2003. 'Clear as mud': Toward greater clarity in generic qualitative research. *International Journal of Qualitative Methods* **2** (2): 1–13.

Clark, Alexander M. 1998. The qualitative-quantitative debate: Moving from positivism and confrontation to post-positivism and reconciliation. *Journal of Advanced Nursing* **27** (6):1242–9.

Corbin, Juliet M., and Anselm Strauss. 1990. Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative sociology* **13** (1):3–21.

---

3 Within the QCA lexicon, results are referred to as causal pathways, although researchers are cautioned against the use of terms like causation as QCA uses a combinatorial logic/conjunctural causation instead of main effect/parameter estimate logic.

Corbin, Juliet, and Anselm L. Strauss. 2014. *Basics of qualitative research*. Thousand Oaks, CA: Sage.

Creswell, John W., and Dana L. Miller. 2000. "Determining validity in qualitative inquiry." *Theory into Practice*. 39 (3): 124–130.

Creswell, John W., and Vicki L. Plano Clark. 2018. *Designing and conducting mixed methods research*, 3rd ed. Thousand Oaks, CA: Sage.

Creswell, John W. and Cheryl Poth. 2018. *Qualitative inquiry and research design: Choosing among five approaches*. 4th ed. Thousand Oaks, CA: Sage.

Creswell, John W., Vicki L. Plano Clark, Michelle L. Gutmann, and William E. Hanson. 2003. "Advanced mixed methods research designs." *Handbook of Mixed Methods in Social and Behavioral Research* 209: 240.

Cristofoli, Daniela and Josip Markovic. 2016. How to make public networks really work: A qualitative comparative analysis. *Public Administration* 94: 89–110. doi:10.1111/padm.12192

Dias, Janice Johnson, and Steven Maynard-Moody. 2007. For-Profit Welfare: Contracts, Conflicts, and the Performance Paradox. *Journal of Public Administration Research and Theory* 17:189–211.

Denhardt, Robert B. 2001. The big questions of public administration education. *Public Administration Review*. 61 (5):526–34.

Denzin, Norman K., and Yvonna S. Lincoln. 2003. *The landscape of qualitative research: Theories and issues*. 2nd ed. Thousand Oaks, CA: Sage.

Denzin, Norman K., and Yvonna S. Lincoln. 2012. *Strategies of qualitative inquiry*. Vol. 4. Thousand Oaks, CA: Sage.

Dodge, J., S. M. Ospina, and E. G. Foldy. 2005. Integrating rigor and relevance in public administration scholarship: The contribution of narrative inquiry. *Public Administration Review* 65 (3):286–300.

Dubnick, M. J. 1999. Demons, spirits, and elephants: Reflections on the failure of public administration theory. Paper presented at the annual meeting of the American Political Science Association, Atlanta, GA.

Eikenberry, Angela M., Verónica Arroyave, and Tracy Cooper. 2007. Administrative failure and the international NGO response to Hurricane Katrina. *Public Administration Review* 67 (1):160–70.

Eisenhardt, Kathleen M., and Melissa E. Graebner. 2007. Theory building from cases: Opportunities and challenges. *Academy of Management Journal*. 50 (1):25–32.

Eisenhardt, Kathleen M., Melissa E. Graebner, and Scott Sonenshein. 2016. "Grand challenges and inductive methods: Rigor without rigor mortis." *Academy Of Management Journal* 59 (4): 1113–1123.

Feldman, Martha S. 2000. "Organizational routines as a source of continuous change." *Organization Science* 11 (6): 611–629.

Feldman, Martha S., and B. T. Pentland. 2003. Reconceptualizing organizational routines as a source of flexibility and change. *Administrative Science Quarterly* 48 (1):94–118.

Feldman, Martha S., Kaj Sköldberg, Ruth NicoleBrown, and Debra Horner. 2004. Making sense of stories: A rhetorical approach to narrative analysis. *Journal of Public Administration Research and Theory* 14 (2):147–70.

Fischer, Manuel. 2011. Social Network Analysis and Qualitative Comparative Analysis: Their mutual benefit for the explanation of policy network structures. *Methodological Innovations Online* 6 (2):27–51.

Freeman, John Leiper. 1965. *The Political Process: Executive Bureau-Legislative, Committee Relations*. Vol. 13. New York: Random House.

Glaser, Barney G. 2002. Conceptualization: On theory and theorizing using grounded theory. *International Journal of Qualitative Methods* 1 (2):23–38.

Glaser, Barney G., and Anselm L. Strauss. 2017. *Discovery of grounded theory: Strategies for qualitative research*. London: Routledge.

Gentles, Stephen J., Cathy Charles, Jenny Ploeg, and K. Ann McKibbon. 2015. Sampling in qualitative research: Insights from an overview of the methods literature. *The Qualitative Report* 20 (11):1772.

Gill, J., and K. J. Meier. 2000. Public administration research and practice: A methodological manifesto. *Journal of Public Administration Research and Theory* 10 (1):157–99.

Giorgi, Amedeo. 1997. The theory, practice, and evaluation of the phenomenological method as a qualitative research procedure. *Journal of Phenomenological Psychology* 28 (2):235–60.

Guba, Egon G., ed. 1990. *The paradigm dialog*. Newbury Park, CA: Sage.

Hammersley, Martyn. 1983. *Ethnography*. San Francisco, CA: John Wiley & Sons.

Haverland, Markus, and Dvora Yanow. 2012. A hitchhiker ' s guide to the public administration research universe: Surviving conversations on methodologies and methods. *Public Administration Review* 72 (3):401–8.

Head, Brian William. 2010. Public management research: Towards relevance. *Public Management Review* 12 (5):571–85.

Holland, P. W. 1986. Statistics and causal inference. *Journal of the American Statistical Association*. 81 (396):945–960.

Honig, Dan. 2018. Case study design and analysis as a complementary empirical strategy to econometric analysis in the study of public agencies: deploying mutually supportive mixed methods. *Current issue*.

Hudson, John, and Stefan Kühner. 2013. Qualitative comparative analysis and applied public policy analysis: New applications of innovative methods. *Policy and Society* 32 (4):279–87.

Hug, Simon. 2013. Qualitative comparative analysis: How inductive use and measurement error lead to problematic inference. *Political Analysis* 21 (2):252–65.

Johnson, R. Burke, and Anthony J. Onwuegbuzie. 2004. Mixed methods research: A research paradigm whose time has come. *Educational Researcher* 33 (7):14–26.

Johnson, R. Burke, Anthony J. Onwuegbuzie, and Lisa A. Turner. 2007. Toward a definition of mixed methods research. *Journal of Mixed Methods Research* 1 (2):112–33.

Jonsen, K., and K. A. Jehn. 2009. Using triangulation to validate themes in qualitative studies. *Qualitative Research in Organizations and Management: An International Journal* 4 (2):123–50.

Kahlke, Renate M. 2014. Generic qualitative approaches: Pitfalls and benefits of methodological mixology. *International Journal of Qualitative Methods* 13 (1):37–52.

Kettl, Donald F. 2000. Public administration at the millennium: The state of the field. *Journal of Public Administration Research and Theory* 10 (1):7–34.

King, Cheryl Simrell, Kathryn M. Feltey, and Bridget O'Neill Susel. 1998. The Question of Participation: Toward Authentic Public Participation in Public Administration. *Public Administration Review* 58 (4):317. doi:10.2307/977561.

Kuhn, Thomas S. 1996. The nature and necessity of scientific revolutions. In *The structure of scientific revolutions*, Kihn T. S. ed. 3rd ed. Chicago: University of Chicago Press.

Lambright, Kristina T. 2008. Agency theory and beyond: Contracted providers' motivations to properly use service monitoring tools. *Journal of Public Administration Research and Theory* 19 (2):207–27.

Langley, Ann. 1999. Strategies for theorizing from process data. *Academy of Management review* 24 (4):691–710.

Lewis, D. 1973. Causation. *The Journal of Philosophy* 70 (17):556–67.

Lincoln, Yvonna S., and Egon G. Guba. 1986. But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. *New Directions for Evaluation* 30:73–84.

Lipsky, Michael. 1971. Street-level bureaucracy and the analysis of urban reform. *Urban Affairs Quarterly* 6 (4):391–409. doi:10.1177/107808747100600401

Lowery, Daniel, and Karen G. Evans. 2004. The iron cage of methodology: The vicious circle of means limiting ends limiting means... *Administration & Society* 36 (3): 306–27.

Malatesta, Deanna, and Julia L. Carboni. 2015. The public–private distinction: Insights for public administration from the state action doctrine. *Public Administration Review* 75 (1):63–74.

Marshall, M. N. 1996. Sampling for qualitative research. *Family Practice* 13 (6):522–6.

Maynard-Moody, Steven, and Marisa Kelly. 1993. Stories managers tell about elected officials: Making sense of the politics-administration dichotomy. In *Public management: The state of the art*, ed. B. Bozeman, 71–90. San Francisco: Jossey-Bass.

McNabb, David E. 2014. *Case research in public management*. London: Routledge.

McNabb, David E. 2015. *Research methods in public administration and nonprofit management*. London: Routledge.

McBeth, Mark K., Michael D. Jones, and Elizabeth A. Shanahan. 2014. "The narrative policy framework." In *Theories of the policy process*, edited by Sabatier, Paul A., and Weible, Christopher M, 225–266. Boulder: Westview Press.

Mertens, Donna M., and Amy T. Wilson. 2012. *Program evaluation theory and practice: A comprehensive guide*. New York: Guilford Press.

Milward, H. Brinton. Forthcoming. Toward a theory of organizational networks: Structure, process, and people. *Perspectives on Public Management and Governance*.

Miles, Matthew B., A. Michael Huberman, and Johnny Saldana. 2013. *Qualitative data analysis*. Thousand Oaks, CA: Sage.

Moran, Dermot. 2002. *Introduction to phenomenology*. London: Routledge.

Morse, Janice M. 1995. The significance of saturation. *Qualitative Health Research* 5 (2):147–9.

Morse, Janice M. 2015. Critical analysis of strategies for determining rigor in qualitative inquiry. *Qualitative Health Research* 25 (9):1212–22.

Morse, Janice M., Phyllis Noerager Stern, Juliet Corbin, Barbara Bowers, Kathy Charmaz, and Adele E. Clarke. 2016. *Developing grounded theory: The second generation*. London: Routledge.

Neuman, William Lawrence, and Karen Robson. 2014. *Basics of social research*. Canada: Pearson.

Onwuegbuzie, A. J., and N. L. Leech. 2007. A call for qualitative power analyses. *Quality & Quantity* 41 (1):105–21.

O'Reilly, Michelle, and Nicola Parker. 2013. 'Unsatisfactory Saturation': A critical exploration of the notion of saturated sample sizes in qualitative research. *Qualitative Research* 13 (2):190–7.

Ospina, Sonia M., and Jennifer Dodge. 2005. It's about time: Catching method up to meaning—the usefulness of narrative inquiry in public administration research. *Public Administration Review* 65 (2):143–57.

Ospina, Sonia M., Marc Esteve, and Seulki Lee. 2017. Assessing Qualitative Studies in Public Administration Research. *Public Administration Review* 78 (4):593–605. doi:10.1111/puar.12837

Patton, Michael Quinn. 1999. Enhancing the quality and credibility of qualitative analysis. *Health Services Research* 34 (5): 1189–208.

Patton, Michael Quinn. 2014. *Qualitative Research & Evaluation Methods: Integrating Theory and Practice*. Thousand Oaks, CA: Sage.

Pattyn, Valérie, Astrid Molenveld, and Barbara Befani. 2017. Qualitative comparative analysis as an evaluation tool: Lessons from an application in development cooperation. *American Journal of Evaluation*.

Perry, James L., and Lois Recascino Wise. 1990. The motivational bases of public service. *Public Administration Review* 50 (3):367. doi:10.2307/976618

Pillow, W. 2003. Confession, catharsis, or cure? Rethinking the uses of reflexivity as methodological power in qualitative research. *International Journal of Qualitative Studies in Education* 16 (2):175–96.

Prasad, Pushkala. 2015. *Crafting qualitative research: Working in the post-positivist traditions*. London: Routledge.

Preissle, Judith, and Margaret D. Le Compte. 1984. *Ethnography and qualitative design in educational research*. New York: Academic Press.

Raab, Jörg., R. S. Mannak, and B. Cambre. 2015. Combining structure, governance, and context: A configurational approach to network effectiveness. *Journal of Public Administration Research and Theory* 25 (2):479–511. doi:10.1093/jopart/mut039

Raadschelders, J. C. 2011. The future of the study of public administration: Embedding research object and methodology in epistemology and ontology. *Public Administration Review* 71 (6):916–24.

Ragin, Charles. 1987. *The comparative method: Moving beyond qualitative and quantitative methods*. Berkeley: University of California Press.

Ragin, Charles C. 1998. The logic of qualitative comparative analysis. *International Review of Social History* 43 (S6):105–24. doi:10.1017/S0020859000115111

Ragin, Charles C., and Benoit Rihoux. 2004. Qualitative Comparative Analysis (QCA): State of the Art and Prospects. *Qualitative Methods* 2 (2):3–13.

Rerup, C., and M. S. Feldman. 2011. Routines as a source of change in organizational schemata: The role of trial-and-error learning. *Academy of Management Journal* 54 (3):577–610.

Riccucci, Norma M. 2010. *Public administration: Traditions of inquiry and philosophies of knowledge*. Washington, DC: Georgetown University Press.

Riessman, Catherine Kohler. 1993. *Narrative analysis*. Vol. 30. Thousand Oaks, CA: Sage.

Rihoux, Benoît, and Charles C. Ragin. 2008. *Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques*. Vol. 51. Thousand Oaks, CA: Sage.

Rivera, Lauren A. 2017. "When two bodies are (not) a problem: Gender and relationship status discrimination in academic hiring." *American Sociological Review* 82 (6): 1111–1138.

Rolfe, Gary. 2006. Validity, trustworthiness and rigour: Quality and the idea of qualitative research. *Journal of Advanced Nursing* 53 (3):304–10.

Romzek, B., and Melvin J. Dubnick. 1987. Accountability in the Public Sector: Lessons from the Challenger Disaster. *Public Administration Review* 47 (3):227–38.

Romzek, B., Kelly LeRoux, and Jeannette M. Blackmar. 2012. A preliminary theory of informal accountability among network organizational actors. *Public Administration Review* 72 (3):442–53.

Romzek, B., K. LeRoux, J. Johnston, R. J. Kempf, and J. S. Piatak. 2014. Informal accountability in multisector service delivery collaborations. *Journal of Public Administration Research and Theory* 24 (4):813–42. doi:10.1093/jopart/mut027

Sanger, Mary Bryna. 2013. Does measuring performance lead to better performance?. *Journal of Policy Analysis and Management* 32 (1): 185–203.

Saz-Carranza, Angel, and Sonia M. Ospina. 2010. The behavioral dimension of governing interorganizational goal-directed networks—Managing the unity-diversity tension. *Journal of Public Administration Research and Theory*. 21 (2):327–65.

Schein, Edgar H. 2003. On dialogue, culture, and organizational learning. *Reflections: The SoL Journal* 4 (4):27–38. doi:10.1162/152417303322004184

Schillemans, Thomas. 2013. Moving beyond the clash of interests: On stewardship theory and the relationships between central government departments and public agencies. *Public Management Review* 15 (4):541–62.

Shanahan, Elizabeth A., Michael D. Jones, and Mark K. McBeth. 2018. How to conduct a Narrative Policy Framework study. *The Social Science Journal*. 55 (3):332–345.

Shanahan, Elizabeth A., Michael D. Jones, Mark K. McBeth, and Ross R. Lane. 2013. An angel on the wind: How heroic policy narratives shape policy realities: Narrative policy framework. *Policy Studies Journal* 41 (3):453–83. doi:10.1111/psj.12025

Schneider, Anne, and Helen Ingram. 1993. Social construction of target populations: Implications for politics and policy. *American Political Science Review* 87 (2):334–47.

Steiner, Elizabeth. 1988. *Methodology of theory building*. Sydney: Educology Research Associates.

Stout, Margaret. 2013. Preparing public administration scholars for qualitative inquiry: A status report. *Public Administration Research* 2 (1):11–28.

Thomann, Eva. 2015. Is output performance all about the resources? A fuzzy-set qualitative comparative analysis of street-level bureaucrats in Switzerland. *Public Administration* 93 (1):177–94.

Toepler, Stefan. 2005. Called to order: A board president in trouble. *Nonprofit Management and Leadership* 15 (4):469–76.

Van Slyke, David M. 2003. The mythology of privatization in contracting for social services. *Public Administration Review* 63 (3):296–315. doi:10.1111/1540–6210.00291

Wasserman, Larry. 2013. *All of statistics: A concise course in statistical inference*. New York: Springer Science & Business Media.

Waugh, William L. Jr, and Wesley W. Waugh. 2003. Phenomenology and public administration. *International Journal of Organization Theory & Behavior*. 7 (3):405–31.

Weber, E. P., and A. M. Khademian. 2008. Wicked problems, knowledge challenges, and collaborative capacity builders in network settings. *Public administration review* 68 (2):334–49.

Weick, Karl E. 1993. The collapse of sensemaking in organizations: The mann gulch disaster. *Administrative Science Quarterly* 38 (4):628. doi:10.2307/2393339

Williamson, O. E. 1981. The economics of organization: The transaction cost approach. *American Journal of Sociology* 87 (3):548–77.

Yin, Robert K. 2017. *Case study research and applications: Design and methods*. Thousand Oaks, CA: Sage.