The Case for Gauging as a Research Purpose in Public Administration: Connecting Purpose, Frameworks, Design, and Norms in Applied Research



*Published Version:*
Shields, P.M., Casula, M., Rangarajan, N. (2025). The Case for Gauging as a Research Purpose in Public Administration: Connecting Purpose, Frameworks, Design, and Norms in Applied Research. ADMINISTRATION & SOCIETY, 57(2), 1-281 [10.1177/00953997241303932].

*Availability:*
This version is available at: https://hdl.handle.net/11585/999432 since: 2025-03-28

*Published:*
DOI: http://doi.org/10.1177/00953997241303932

**The Case for Gauging as a Research Purpose in Public Administration: Connecting Purpose, Frameworks, Design and Norms in Applied Research**

**Abstract:** This article proposes "gauging" as an unrecognized and useful research purpose for public administration. Unlike the explanatory research purpose, which uses an experimental logic and hypotheses, gauging incorporates a logic of rating and frameworks like standards. It is used in normative assessment like accreditation, process evaluation and regulatory policy. We argue that it should be part of public administration's research tradition because it is widespread in public administration, links to applied research and practice, provides summative and formative assessment, connects evidence to the logic of law and draws on pragmatism as a link to theory and mixed methods.

**Keywords**: Research Methods; Applied Research; Assessment; Pragmatism, Academic-Practitioner, Pracademic.

## Introduction

Questions about the quality of research methods used in public administration (PA) have been around for a long time.[i] Recently, Zhu et al (2018) called for a more concentrated focus on methods noting that unlike other fields there is no journal in PA devoted to research methods. They are concerned with the relevance of the field to practitioners and see issues around research methods as central to the identity crisis in PA. In addition, they find too much focus on quantitative studies and causal inference and call for a pluralism in research methodology and "rigorous designs motivated by the purpose of the research" (p. 7).

This paper responds to their call. While Zhu et al (2018) were most likely focusing their attention at the individual study level, we believe it shines a light on the importance and connection between design and research purpose. Stepping back and taking a macro view, we draw on Earl Babbie's (2007) and others[ii] taxonomy of research purposes (explanation, description, and exploration). We will show an underlying connection between these purposes and research design. But more importantly, we argue that Babbie's taxonomy is incomplete for public administration. In this paper, we make the case that a fourth purpose, we call, gauging, should be added to the taxonomy. Hence, we present gauging as an essential but unrecognized research purpose for public administration. We discuss its widespread use in the practice of public administration, its potential for framework development, and the kind of researchers that should benefit from it.

Gauging is a normative research purpose that employs a logic of rating and is often used in assessment. It has the potential to provide pathways to improve the processes and activities of public administration as well as promote meaningful connections within and across the PA landscape. Explanatory research generally employs quantitative methods and hypotheses to discover causal connections with external validity. Gauging by contrast often uses standards and case study methods to assess a process or program. Explanatory research dominates the academic journals of public administration. As a new methodological instrument, gauging enables better understanding of administrative processes and their final impact on society. Gauging research is more often connected to the world of practice and applied research, for example, Master of Public Administration (MPA) Program accreditation.

When an MPA program undergoes accreditation, it begins with a self-study. This report, which is reviewed by the accrediting body, uses a set of standards to structure the program review. It is a key document essential to determining whether a program should be accredited. This self-study is a kind of case study (the MPA Program) that incorporates data sources such as program records, surveys of students, alumni, employers, and program documents. The goal is to use this evidence to prove to or convince the accrediting body that the program adheres to the standards. The accrediting body is tasked with *gauging* whether the program meets the standards and should be accredited. The activities used to construct documents like the MPA Program self-study are ingrained within higher education management[iii], and are an under-recognized form of applied research.

Gauging research extends well beyond higher education and reaches across the globe. It is manifest in workplace health and safety regulation, auditing agencies, and best practices research. Note that we are not claiming gauging should replace or is superior to explanatory research, rather that it should be included in the research purpose typology for public administration. To achieve this goal, we begin by taking a closer look at how research purposes enhance coherence and connections across the stages of the research process. We place the gauging purpose within this larger categorical frame.

**Value of Purposes in PA Research**

We begin with the three widely recognized research purposes – explanation, description and exploration. Research purposes focus on aims or goals of a study and are related to research questions. For example, explanatory research answers the "why" question (Babbie, 2007, pp. 89-90), and looks "for causes and reasons" (Adler & Clark 2008, pp. 14). Deductive explanatory research uses hypotheses to structure their design. Hypotheses connect the research purpose to other elements of the research process. They "serve as crucial bridges between theory and data, making explicit how the variables and relationships that follow from a logical argument will be operationalized" (Sutton and Staw, 1995, p. 376). Ideally, hypotheses are grounded in the literature and research question and link to operationalization of the independent, dependent and control variables. We conclude an explanatory study by testing the hypotheses using statistics. The results should answer the question and achieve the purpose.

"What" questions are answered using the descriptive research purpose and these "what" questions are *not* concerned with causes (Strydom, 2013). According to Grinnell (2001, p. 48) description lies at the "midpoint of the knowledge continuum" between explanation and exploration. Science uses basic categorization systems (periodic table of chemistry or plant taxonomies of botany). "These baseline classification systems are a type of theorizing." (author) Classification systems can be used to answer questions like "what kind" of plants grow in the San Marcos River?

Descriptive data is usually displayed using frequency distributions and figures like pie charts or bar graphs (Ritchie and Lewis, 2004, Strydom, 2013). The descriptive purpose just presented uses deductive reasoning because the framework is constructed before the data is collected. For example, if a scholar asked, what kind of policies do coastal towns in Texas use to mitigate hurricane damage? A system to classify mitigation strategies could be developed. A survey could be constructed using the types of mitigation strategies and sent to city managers. The survey results could be displayed as frequency distributions or as charts. Descriptive studies can use inductive reasoning, which draws "inferences from specific observable phenomena to general rules or knowledge expansion" (Worster, 2013, p. 448). Inductive descriptive studies would employ qualitative methods such as direct observation and open-ended interview questions. The data could provide a thick description of the hurricane mitigation policies.

Exploratory research is preliminary in nature and unlike description and explanation addresses a variety of questions (Stebbins, 2001). It is also generally qualitative and inductive[iv]. Recently scholars are resurrecting deductive qualitative research methods and, in the process, have suggested frameworks like working hypotheses and pillar questions to help organize evidence collection in case studies (Gilgun, 2015a & 2015b, Blinded 2021).

Table 1 summarizes the horizontal connections for deductive, exploratory, descriptive, and explanatory studies. The remainder of this paper fills in the empty cells for deductive gauging research. Gauging has many possible applied frameworks. The accreditation case illustrates this. The accreditation standards are the *micro-conceptual framework*. The *methodology* is found in the instructions provided by the accrediting body NASPAA and the *data analysis* is in the reports provided by programs to the accrediting body.

[Insert Table 1 here]

The value of the research purpose taxonomy is that it demonstrates the links across the empirical research process[v]. The key link is between the purpose and what we call micro-conceptual frameworks[vi] like the formal hypotheses of explanation or categories of description.

The hypothesis and explanation are so tied together that many definitions of hypotheses include the term explanation[vii].

The underlying experimental logic of hypotheses is so compelling because it can be observed in everyday life. If people encounter a big or small problem and try an intervention to mitigate the problem, they are experimenting And, if that does not work, try something else. Parents of crying babies are constantly testing hypotheses as they seek the solution to soothing the child. This is an example of experimental logic. Our point here is that hypotheses are embedded in the way we solve problems. We seek causes as we look for ways to resolve problematic situations.[viii] When we formulate and test hypotheses, we use deductive reasoning. The same goes for any form of empirical inquiry that uses a framework to guide data collection. Almost by definition inductive explanation, description or exploration would not have a framework. This paper examines the gauging purpose using deductive reasoning.[ix]

**A closer look at the gauging purpose**

The normative and comparative practice of gauging is also embedded in everyday life and activity. Children want to know who runs the fastest, we rate the products we buy from Amazon, we receive grades for quizzes, we check out the ratings of new restaurants. We frequently and naturally apply a logic-of-rating in everyday life. Organizations are filled with opportunities to gauge. Take the hiring decision. How do we decide, which is the best candidate? A decision-making matrix with job qualifications is often employed. This matrix is the standard or framework used in candidate assessment. Like the experimental nature of hypothesis testing, gauging is found throughout life situations we encounter. Social science has a formal research purpose that captures hypothesizing. We are arguing that this fundamental logic of rating should also be captured in a research purpose and that it has particular relevance for a practitioner-oriented field like public administration.

We believe that a defining feature of gauging research is its normative nature. While the descriptive purpose asks the "what" question, gauging research often asks the "what should" question. The term "should" evokes norms. For example, Master of Public Administration (MPA) programs *should* adhere to accreditation standards. Ideally, these standards are associated with well-functioning programs. If programs do not adhere to these standards, they will not attain accreditation. Gauging research often uses standards and performance measures[x].

Gauging is inherently comparative and uses a logic-of-rating for comparison purposes. Note that even the term "rate" stems from "ratio" which uses a simple formula to compare numerator and denominator. This fraction compares the part with the whole. We also use fractions to gauge policy and government entities. The unemployment rate can be used to gauge the health of the economy. We rate states using indicators like the percent of the population with health insurance. Academic journals are rated by comparing numbers of recent citations over number of articles published (Impact Factor). Higher numbers indicate better journals. Keehley et al (1993, p. xii) describe benchmarking as a methodology which compares "the performance of one's own organization with the best practices of others." The key take-away here is that all gauging activities incorporate comparison whether it is through ratios, a standard, or across organizations (how do Texas and California compare?).

Standards embed norms into the framework. Like the hypothesis, standards provide frameworks that help researchers evaluate evidence. The process of judging whether a standard is met may employ many kinds of evidence. Unlike a statistical test, evidence may be assessed using the messier notions of verification and proof.

Standards as a framework are very different from a formal relational hypothesis yet serve a similar function. They help the researcher determine what kind of evidence to collect and how to evaluate the evidence. Hypotheses lay bare the relationships to be studied. Their concepts are usually operationalized into variables with numeric values. Statistical tests provide orthodox ways to evaluate the evidence. The process of judging whether a standard is met employs a similar deductive logic but a different way to obtain and evaluate evidence.

It is possible to do something like operationalization for the frameworks of gauging. This process is underdeveloped, invisible or ignored. Accreditation standards often come with guidelines that specify the kinds of evidence or data an accrediting body is interested in for each standard. The guidelines specify forms of evidence such as program records, survey of alumni and students, the program website, and the curriculum. These instructions help operationalize the standard's framework. They connect the assessment standards to data from the program under review.

We have made the larger case for the use of research purposes as a method to frame deductive empirical research and introduced the gauging research purpose. We now shift gears and offer five arguments for why the gauging purpose is particularly relevant to the applied field of public administration.

First, we argue the gauging function is ubiquitous within PA. Second, we show its applicability to applied research and relevance to practice. Third, gauging research has the capacity to identify poorly performing programs and can provide formative evidence useful for program improvement. Fourth, like PA itself, gauging research has a strong connection to the law. Finally, we draw on the philosophy of pragmatism to show how gauging connects to theory.

**Ubiquitous use of gauging activities in Public Administration**

For "gauging" to cross the threshold as a research purpose candidate its widespread applicability within PA should be established. This section provides extensive examples that demonstrate how members of the PA community collect and use data to fulfill the gauging purpose. We are perhaps most familiar with the use of the logic of rating at the individual level. For example, we routinely are asked to rate the quality of restaurants, products and stays at hotels. Universities rate student performance and have issued grades since the 18th century (Durm, 1993, p. 2).

Table 2 illustrates the ubiquity of gauging activities in PA practice. We show how the Bureau of Indian Standards establishes detailed labeling and marking requirements on manufactured toys. The Workplace Gender Equality Agency in Australia enforces the employer responsibility to have policies or strategies supporting gender equality. The Associação Brasileira de Normas Técnicas of Brazil mandate that safety colors be adopted in workplaces to warn about existing risks. The Canadian Food Inspection agency regulates the nutritional composition of sterilized milk. These examples are drawn from every continent and cross a variety of policy areas such as transportation, education, and the environment (columns 1 & 2). The third column underscores the applicability of gauging at the micro, meso, and macro levels of analysis. Programs and processes that have an impact on individuals, organizations, and societal level variables are gauged to assess their conformance and compliance with standards. Column 4 provides specific examples of individual standards that are part of a larger set of standards that make up a framework. The normative nature of the standards is illustrated through terms like "must, shall and should" found in bold. Evidence pertaining to these standards collected by the respective enforcement authorities will help determine the level of conformance with those standards by the entity being assessed.  The agency responsible for developing and enforcing the standard is found in column

5. Most of the standards highlighted in this table connect to the regulatory function of government. All these examples beg the question "where do the standards come from?" In public administration they emerge from a formal response to public problems such as safety concerns or establishing quality or ensuring appropriate procedures and processes. Such formal responses to public problems, which are the sources of standards, can take many forms. They can emerge from agency discourse or rulemaking processes that engage public participation.

[Insert Table 2 here]

Table 2 focuses on standards that are used in assessment. Figure 1 shows an example of a rating system. Not surprisingly, education agencies use grades (A, B, C) to rate the performance of students, schools, and school districts. Drawing from published Texas Education Agency reports we find that Cunningham elementary has received a grade of C in the "student achievement" and "school progress" domains, and a D in "closing the gaps" domain. Texas schools are graded in figure 2. Here we learn that 19.4 percent of all Texas public school campuses received a grade of "C" and that most campuses received a grade of "B". This system of rating used an ordinal scale. It is common to use something like a pass/fail rating system too.

[Insert Figure 1 here]

[Insert Figure 2 here]

In addition, auditing, and scrutinizing government agencies around the world gauge activities at the meso level. These organizations are responsible for assessing the performance of the departments across the different functions of government. We are not saying that audit agencies only engage in gauging activities or that their reports are flawless, rather that they often use standards to evaluate programs and processes. Macro gauging activities can be observed when states (in the US) or countries are compared (e.g., per capita income, average education level, percent of population that completes high school).

All these examples showcase aspects of applied research performed by practicing public administrators around the world. The collection of the evidence for reports like those at the Texas Education Agency include thousands of teachers and school administrators in the data collection process. When teachers administer a high stakes test or administrators develop school specific reports that summarize performance data, they are devoting some part of their time to larger gauging efforts. This aspect of applied research in PA is seldom recognized.

## Connection to Applied Research and Practitioners

We now move on to the second argument for the application of the gauging purpose to public administration – its connection to practitioners and applied research. The examples in the previous section link the gauging research purpose to policies and activities of practicing public administrators. It is ironic that Zhu et al (2008 p. 2) point to the increasing irrelevance to practitioners of research studies found in PA journals and dissertations. We could immediately make a profound link to practitioners by incorporating the "gauging" purpose in our research methods texts.

We articulated and saw the value of gauging research purpose as we supervised masters level capstone applied research. Many MPA Programs are filled with mid-career students who are currently working in the public and nonprofit sectors. In the mid 1980's our MPA Program was faced with an accreditation final report that identified our final projects as disorganized with weak literature reviews. The report recommended our student papers incorporate conceptual frameworks. We began grappling with this recommendation and the quizzical responses of students who asked, "what is a conceptual framework?" We gradually began to realize that the most common conceptual framework discussed in research methods books – hypotheses- worked for less than half of their research questions. What besides hypotheses could be used as larger theoretical organizing devices? We began to see the connection between frameworks and purposes (hypotheses for explanation and categories for description). We also had an uncomfortable feeling that Babbie's (2007) three social science purposes did not apply to many problems and research questions common in PA such as process evaluation. It was then that the "gauging" research purpose emerged as did a framework we called the practical ideal type. We also discovered these frameworks could be linked and justified through the literature. This transformed disorganized literature reviews connecting them to the research purpose and to theory.[xi]

Over the years, we learned that students/practitioners embraced gauging research. It resonated with their experience and way of thinking. Further, it could be applied to their work. For example, one student developed a framework to evaluate succession planning processes in her division. Somehow, the auditing arm of her agency found out and used her model (perhaps somewhat modified) in their larger work. Empirical research with a gauging purpose is seldom found in academic journals but is widely applicable to the practical day-to-day world of public administration. [xii]

### *Characteristics of Applied Research*

Research projects with a gauging purpose are generally a type of applied research. [xiii] Applied research focuses on a "specific problem" and finding the solution to that problem (Hedrick et al 1993; Provost and Kohavi, 1998 Verhoeven, 2007). It uses scientific methodology and operates in an often-messy environment with pressure for "quick and conclusive answers, sometimes in very political contexts" (Hedrick et al 1993, p.2).

The focus on a specific problem, results in research products relatively weak in external validity or generalizability of the findings. Perhaps because external validity is a characteristic highly valued in academic journals, applied research takes a back seat to studies with wide applicability. But for a practitioner, who is interested in a particular problem, external validity is of little consequence.

Hedrick et al (1993, pp. 4-9) clarify the nature of applied research by distinguishing it from basic research, which develops universal knowledge, focuses on a single question, seeks to discover statistically significant relationships or effects, focuses on causes, generally uses a single method, and employs direct observation. Applied research, in contrast, addresses problems, examines multiple questions, focuses on effects, and uses multiple methods including indirect observations. They assert basic research is more likely to use experimental design, while applied research uses quasi-experimental design. Clearly, applied research would use all three of Babbie's research purposes, although for Hedrick et al (1993) explanation appears to be privileged.

Applied research also tackles normative questions. These kinds of studies "require comparing" the characteristics of a program "to a standard". They ask, 'what is the difference between what is (or what was) and what should have been?' The standard may be based on one of a number of sources – legislative objectives, professional standards, program goals or other a priori criteria." They also note that a question like "How *well* run is the education program?" requires the development of criteria for a "well run" program. "Methods to develop standards include reviewing prior research, reviewing statements of legislative intent, and discussions with recognized experts" (p 27). All these activities (literature review, analysis of public documents and interviews) are relevant to public administration and the PA methods classroom. They go on to say, "development of standards may even become an activity of the research itself" (Hedrick et al. 1993, p. 27). We argue that these normative questions and processes which, Hedrick et al (1993) devote a few sentences, need greater attention.

The art and science of building frameworks of this sort has advanced significantly since Hedrick et al 's 1993 book. Take for example, van Popering-Verkek et al's (2022) "A Framework for Governance Capacity: A Broad Perspective on Steering Efforts in Society". They use a sophisticated mixed methods approach to build a governance capacity model for governance networks. The study concludes with a comprehensive preliminary model that informs both the "network of organizations" and organizations within the network how to analyze (assess?) their actions. The framework and table "translate theoretical concepts into items that make sense in the eyes of participants in day-to-day work situations" (van Popering-Verkeck et al, 2022, pp 1782-1783). Clearly, a possible next step could be for participants to use the framework to gauge and improve their own networks which, would be applied research. Interestingly, the authors hint at this application of the model. Their conclusion focuses on the theoretical yet can be easily operationalized for practitioners. In addition to its applicability to public administration (see for example, Farmer & Costello, 2019), other uses of this framework can be seen in fields such as arboriculture and urban forestry (see O'Herrin et al., 2023), cooperative development (see Muthuma, 2012), city ordinance effectiveness (Lerer, 2012), strategic planning (Muiruri et al, 2017) and intellectual property (see Ruddick et al., 2011).

Note that for the specific problem orientation of applied research there is limited concern with external validity or the generalizability of the findings. On the other hand, in the case of gauging research, the framework itself could have analytical generalizability (Yin, 2017). Many of them do. Take accreditation standards. They are used by the relevant programs and institutions that seek accreditation. The framework is a kind of "practical-theoretical framework" that guides data collection, analysis, and interpretation of findings.

***Possible to incorporate the "Coach" and "Cop" Perspectives***

Our fourth argument for gauging as a PA research purpose revolves around its potential to identify and stop problematic street level or administrative practices and encourage productive organizational change. We call this the "coach and cop" roles. The results of gauging research can be used by public administrators in their roles as inspectors or auditors to determine adherence with standards and whether penalties should be issued. Most organizations operate within compliance standards. There is a vast middle ground between minimal conformance and exceptional work. The results of a gauging study can reveal strengths and weaknesses as well as provide recommendations for program enhancement. Hence, many "Scrutinizing" agencies also cultivate that middle ground as a way to encourage constructive reform or improve practices[xiv]. Thus, gauging research can be used by reviewing authorities as they adopt the role of coach.[xv]

These tasks also align well with gauging as applied research that focuses on a particular problem by a given agency.

In the United Kingdom these roles are referred to as the guard dog and guide dog functions of “Scrutinizing” organizations. The watchdog is “concerned with procedural compliance” and uses explicit assessment criteria (Nutley et al 2012 p. 872). A guide dog, on the other hand, helps the client improve. This helpful canine supplement the formal criteria with inductive methods and consensus building principles (Nutley et al 2012, p. 875). Hence, the gauging purpose can be used for compliance and constructive change.

Saunders and Holahan (1997) beautifully illustrate this “coach and cop” capacity in a study about judging high school instrumental performance as a way to select the all-state band. They highlight the adjudicator (cop) role as judges make decisions about the level of student performance. The authors emphasized the value of criterion-based rating scales because the process provided “diagnostic feedback” to the students. The rating scale helps students who did not make the state band through concrete feedback on what/how to improve (coach). For example, a student who was rated midrange on “tone” would learn that they exhibited “some flaws in production (i.e., a slightly thin or unfocused sound, somewhat forced, breadth, not always used efficiently, etc.)” (p. 264). The grading rubric (or framework) was developed to serve both the coach and cop roles.

Asselineau et al (2022, p. 977) criticizes the use of numbers in public administration that rate and rank. We believe that gauging research can encourage the wise use of numbers. The seemingly opposing roles of coach and cop provide this opening. It also fits the experience of a field tasked with ensuring compliance with laws, rules, and regulations and that is a champion of organizational effectiveness, efficiency, and public service. It makes sense to incorporate the gauging research purpose as part of our research methods tool kit.

### *Gauging Incorporates a Legal Perspective*

Fifth, the gauging research purpose has strong ties to public administration through its link to law. Public administrators implement laws. Many of the standards enforced by public administrators emanate from regulatory laws. This function, in the U.S. can be tied to the establishment of the administrative state and railroad regulation. According to McSwite (1997, p.105) governments did not engage in developing regulations or standards that were monitored by administrative agencies until the widespread use of railroad transportation in the 19th century. In addition, industrialization during this period created the need for workplace safety regulations (Shields 2023). This timeframe also corresponds with the awakening of public administration as a self-aware field or discipline in the United States (Stivers, 2000).

Regulatory compliance incorporates legal notions of evidence including *proof* and *verification*. Verification establishes the truth, accuracy, or reality of a claim (Webster's dictionary) Proof establishes the cogency of evidence that compels acceptance by the mind of a truth or fact (Webster's dictionary).

Laws are explicitly normative and one of the sources of the norms that frame many gauging research activities (Hedrick et al., 1993). Research that uses standards based in law to reach conclusions and judgment, implicitly or explicitly incorporates legal notions of evidence into its processes. Standard compliance investigation requires *evidence* of adherence or violation of a standard. In both research and law evidence is a core concept (Rycroft-Malone et al., 2003, p. 82). The concept of “evidence” is rooted in the notion of experience, relating to what is manifest and obvious. In law it helps to refute or corroborate the issue at hand (Rycroft-Malone et al., 2003, p. 82). Legislators *expect* laws to be upheld. In their simplest form, hypotheses are *statements of expectations*. A set of standards can serve as a framework of gauging. They are also *statements of expectations* about what should be. Prosecutors bring a *case* to trial when they believe there is enough *evidence to refute the expectation of innocence* (innocent till proven guilty). There is a shared reasoning and vocabulary between the process of adjudicating the law and gauging research.

Clearly "evidence" is a pivotal concept. We will take a slight detour to examine the nature of evidence in practice. "Evidence has meaning only insofar as it is used in some context with regard to some claim. Hence, evidence can be regarded as playing a role in the context of an argument. Evidence is proffered in arguments as probative, this having the status of truth or compelling assent, or belief in the state of affairs described" (Morse et al 2001, p. 15).

Evidence from scientific studies is commonly used to inform practice. In the health field, for example, a medication must be subjected to rigorous experimental trials before it can be widely prescribed. The results of these science-based studies provide evidence that the medication affects patients' health (provides a cure, mitigates a symptom). Evidence from these quantitative studies point to an underlying causal connection (e.g., medication will cause a reduction in pain). Rycroft-Malone et al (2003 p. 82) note that "evidence derived from quantitative research tends to be more highly valued than evidence obtained from other sources."

PA has incorporated these notions of evidence also. The *Journal of Behavioral Public Administration*, which "focuses on behavioral and experimental research in public administration." (https://journal-bpa.org/index.php/jbpa/index ) illustrates this trend. We are not questioning scientific evidence. Rather, the evidence used to verify adherence to standards must meet something like the legal test of "proof." [xvi] Types of evidence that could establish proof include interviews, documents, agency records, surveys, direct observation as well as findings from scientific studies. The approach to evidence collection and analysis is similar to the concept of triangulation used by Yin (2017) and mixed methods (Johnson & Onwuegbunzie, 2004). The cop function of gauging research often requires proof that a standard is met.

Whether evidence is derived from a quantitative scientific study or through a qualitative process that reveals proof, it is always provisional. Another study could provide findings that do not support the hypothesis, or another witness could appear that contradicts earlier testimony. Hypothesis testing and inferential statistics provide clear procedures to evaluate the evidence of scientific studies. There is nothing like statistical tests, which have a kind of universal applicability, to evaluate evidence credibility and demonstrate a standard is met. Somehow, the gauging researcher must still amass evidence relating to what is "manifest and obvious".

It should be noted that legal standards of proof depend on the severity of the case. The more serious the consequences the higher the standard of proof. In addition, establishing proof often relies on the judgment of a group of individuals (jury) or an individual (judge). In the world of practice, gauging research like MPA program accreditation reports involve a panel of peers who evaluate the evidence. A faculty search committee is another example of a group that assesses evidence using a standard (hiring rubric). It is not our intent to delve deeply into the meaning and nuances of legal proof. Rather to make the case that legal concepts of evidence and proof apply to the gauging research purpose. This has unique applicability to applied public administration research and perhaps not to social science research.

Note that the gauging research purpose does not limit itself to judging evidence by legal standards, rather it is open to both scientifically and legal ways of conceptualizing evidence. It would also keep a critical evidentiary norm common to both science and law – evidence should be subject to critical scrutiny or *independently observed and verified* (Rycroft-Malone et al, (2003 p. 83). Gauging often uses a standard generated by government officials as the frame of reference. Nutley et al (2012, p. 871) note that the studies using standards often fail to show a clear connection between *standards and evidence*. Explicitly incorporating the gauging research purpose into PA research methods discourse could create a path to clarifying the connection between standards and evidence.

### *How Gauging Connects to Theory – Building tools for practice*

In the introduction we referred to Zhu et al's (2018, p. 7) call for rigorous designs motivated by the purpose of the research. They also called for methodological pluralism and a stronger focus on the connection between methods and theory (p. 1). At first glance, gauging would provide opportunity for methodological pluralism and be relatively weak where theory is concerned. In this section we will look at the role of theory in gauging research and how it connects to methodological pluralism, classical pragmatism, and mixed methods research.

### *Philosophical Underpinnings*

Before venturing into theory, we briefly detour to the methodology literature and explore the philosophical underpinnings of quantitative, qualitative, and mixed methods approaches to empirical research. Because a wide variety of evidence can be used to establish proof, gauging research aligns with mixed methods research [xvii].

Quantitative methods are associated with positivism/postpositivism, which assumes an objective reality and that actual causes are possible to decipher with validity and reliability (Levers, 2013, Twining et al., 2017, Johnson & Onwuegbunzie, 2004). In addition, positivism relies on the correspondence theory, which asserts that "to each concept there corresponds a set of operations involved in its scientific use" (Kaplan 1964, p. 40).

The interpretivist philosophical orientation of qualitative research takes a very different approach. It assumes reality is relative and knowledge is subjective (Levers, 2013). Thus, context free generalities of positivism are impossible. "Causality, explanation, prediction, experimentation depend on assumptions about the correspondence between concepts and reality, which in the absence of an objective reality is impossible" (Blinded, 2021). For interpretivist scholars, empirical inquiry is "contextualized emergent understanding rather than the creation of testable theoretical structures" (O'Connor et al. 2008, p. 30). Clearly, in their pure sense these philosophies are incompatible.[xviii] Public Administration scholars Nowell and Albrecht's (2018) highly cited "A Reviewer's Guide to Qualitative Rigor" sidesteps the debate by focusing on "practical" ways to use qualitative research. They also note that qualitative research has a "rich history" of integration within postpositivism (p. 349).

Mixed methods research bridges this incompatibility gap by drawing on a philosophy known for dissolving dualisms. The classical pragmatism of William James and John Dewey[xix] uses a "transformative, experimental notion of inquiry" nested in pluralism that focuses on constructing conceptual and practical tools to resolve "problematic situations" (Blinded 2017). According to Johnson & Onwuegbunzie (2004, p. 17) pragmatism provides "an immediate and useful middle position philosophically and methodologically; it offers a practical and outcome-oriented method of inquiry that is based on action." Mixed methods and its partner pragmatism have a pluralistic approach that incorporates all kinds of evidence, which is helpful when gauging whether a standard is met.

The mixed methods literature, however, has not identified a framework that, like the hypothesis, ties the research purpose, methods, and evidence together. On the contrary, frameworks are often missing or unrecognizable (Heyvaert et al. 2013, Blinded 2021)[xx] While the mixed methods literature embraces pragmatism as a philosophical home, it has not incorporated Dewey and James's notion of theory as a way to organize and bring greater horizontal coherence across design components.

**Pragmatists' Notions of Theory**

Pragmatism offers a practice-oriented way to describe theory. Larry Hickman (1990, p. 102) compares Dewey's notion of theory to the Greek philosopher's fixed notion of "forms". In contrast, for a pragmatist, theories are *instruments* or tools for solving problems. Pragmatist philosopher William James (1907) describes a synergy between pragmatism and theory. This synergy is positioned in the problematic situation. He employs a memorable hotel corridor metaphor to make the point. "Pragmatism lies in the midst of theories like a hotel corridor. All rooms open out to it and all rooms can be entered. Pragmatism owns the corridor and the right to move freely from room to room" (p. 54). James is articulating the pluralism inherent in classical pragmatism.

For Dewey theory acts like a map which facilitates navigation. Like maps, theories or frameworks should be assessed by their usefulness. "There is no expectation that a map is a true representation of reality. Rather, it is a representation that allows a traveler to reach a destination (achieve a purpose). Hence, theories should be judged by how well they help resolve the problem or *achieve a purpose*" (Blinded 2013, p. 23). Contemporary scholars such as Horthersall (2018, p. 5) capture the pragmatic notion of theory. "Any theoretical model is for the pragmatist nothing more than a framework through which problems are perceived and subsequently organized". The micro-conceptual frameworks[xxi] linked to research purposes introduced in Table 1 are a type of theory and are defined as the *"way the ideas are organized to achieve the project's purpose*" (Blinded 2013 p. 24). Pragmatism is about purposeful and problem focused inquiry. Theory is not equated with fixed truth, rather it is a helpful tool that organizes inquiry connecting data and problem. Its problem orientation makes it well suited for applied research.

The frameworks of gauging research are a kind of micro level theory that enhance alignment. These frameworks can be found in rating systems like those used by the Texas Education Agency to rate individual schools or by the state music association to find students to fill the state band (Saunders & Holahan, 1997). Gauging frameworks based on standards or regulations create minimum performance expectations. Other gauging frameworks and processes like best practices and the practical idea type (Blinded, 2013) have an aspirational quality. All of them can be viewed as a tool that guides and shapes encounters with the world of practice.

Nowell and Albredht's (2018) call for a practical, pluralistic approach also emphasizes that qualitative methods are uniquely suited to advancing theory. Van Polering-Verkerk et al (2022) built their conceptual framework (theory) for governance capacity to provide a "comprehensive understanding" and to "facilitate analysis" of governance capacity in practice (pp 1780-1782). Inductive qualitative methods were an important component of the framework's design. Their conceptual framework is easy to recognize as an example of Dewey's theory as tool. Further, their broadly applicable conceptual framework incorporated the academic literature using an impressive two-part design. We believe that the gauging purpose, with its aspirational frameworks and standards, can connect theory and practice in concrete and valuable ways.

We see the standards used in MPA Program accreditation, an elementary school rating system used by school districts, a hiring matrix, and a grading rubric as examples of theory-in-use. These tools (theory) are used often by practitioners and help them to keep their projects organized as they work to achieve a purpose. It is also within the rating framework that the details of the "what should" component of gauging is laid bare.

**Conclusion and debate**

The aim of this article was to highlight a research purpose, gauging, which has been hiding in plain sight. We show how Babbie's three research purposes of exploration, description and explanation fail to incorporate normative research problems in public administration and policy. We began with a table (see Table 1) that showed the value of research purposes to enhance coherence across the research process using micro conceptual frameworks. We now return to that table and incorporate the gauging purpose (See table 3 below). We make the case for gauging as a useful research purpose for public administration and policy through five arguments: ubiquity of gauging within PA, its applicability to applied research and practice, its capacity to provide formative evidence useful for program improvement, its strong connection to the law and its connection to theory as informed by philosophical pragmatism.

[Insert Table 3 here]

From a methodological and epistemological point of view, as synthesized in Table 3, with this article we demonstrated the links across the empirical research process, and how the proposed fourth purpose called gauging differs from exploration, description, and explanation in terms of its micro-conceptual framework, methodology adopted, and data analysis procedure used. Throughout the paper we have demonstrated how standards are used as a framework for gauging research in policy and administration. The “standards” framework provides a minimum bar that must be met for compliance. In addition to the framework of standards, we paired the gauging research purpose with a conceptual framework called the practical ideal type which is similar to best practices. The value of the practical ideal type is a topic for another article. Methodologically, we show that for the gauging purpose, qualitative, quantitative, and mixed methods are all potentially applicable to a particular applied research project. What really sets the gauging research method apart is that the data analysis involves a legal/normative logic to the evaluation of evidence. Implementation of law is at the core of public administration and policy. None of the other three purposes (exploration, description, and explanation) have this explicit connection to law as does gauging. This explicit connection to law is a strong link to the world of practice.

We recognize that suggesting gauging as a new research focus for public administration studies is a challenging prospect for our academic community and for stakeholders and policy actors interested in analyzing complex applied research projects. In conclusion, we need to consider a set of points for discussing the potential future development of this line of inquiry.

To establish gauging as a legitimate research focus for public administration, we have begun by validating the significance of gauging as a purpose. We recognize the complexity and importance of developing standards and frameworks for gauging, and we believe that this is a crucial step in advancing gauging as a research focus. Essentially, we are now seeking suggestions for the processes involved in developing a valid framework as the next phase of this research.

It is important to consider at least three limitations when using gauging as a research purpose. These limitations must be addressed by carefully curating assessment criteria, favoring formative assessments over punitive approaches, and regularly reviewing standards to keep them relevant. Firstly, the effectiveness of the gauging process relies on the robustness of the framework used to assess the criteria and procedures. Therefore, implementing a stronger comprehensive set of standards should lead to a more thorough and reliable gauging process. Another limitation is the potential for punitive application, making it counterproductive. Certain administrative roles have the authority to create regulations, often based on gauging methods, which can be detrimental to those they affect. Lastly, it is important to recognize the contextual dependency of gauging standards. We must acknowledge that these standards may not accommodate changing circumstances. Standards that are relevant in one decade or for a specific context or case study may become outdated in other administrative settings unless they are carefully reviewed and updated.

Despite acknowledging these potential limitations, this article presents a new, useful research purpose for our community. The gauging purpose takes into account an assessment practice widespread in public administration The example of NASPAA's accreditation/reaccreditation processes demonstrates that gauging is widely used in public administration education. This process prompts critical self-reflection and establishes program quality, which is valuable for various stakeholders, including faculty, staff, students and employers.

We argue that government officials conducting inspections or audits using established standards are employing gauging activities. Similarly, students who use ask normative evaluation questions in a class project can develop criteria for processes like succession planning which could lead to program and process improvements.. Our experience working with student practitioners shows that gauging research bridges the gap between academia and practice and has the potential to bring the two communities closer together[xxii].

## References


Adams, G. B., & White, J. D. (1994). Dissertation research in public administration and cognate fields: An assessment of methods and quality. *Public Administration Review*, 565-576.

Adler, E. & Clark, R. (2008). *How It's Done: An Invitation to Social Research, 3rd ed*. Thompson-Wadsworth.

Asselineau, A., Grolleau, G., & Mzoughi, N. (2022). A good servant but a poor master: The side effects of numbers and metrics. *Administration & Society*, *54*(5), 971-991.

Babbie, E. (2007), *The Practice of Social Research* (11th edition), Belmont CA: Thompson, Wadsworth.

Babbie, E. & Mouton, J. (2010). *The practice of social research*. Oxford University Press.

Briesch, A. M., Chafouleas, S. M., Neugebauer, S. R., & Riley-Tillman, T. C. (2013). Assessing influences on intervention implementation: Revision of the Usage Rating Profile-Intervention. *Journal of school psychology*, *51*(1), 81-96.

Brown, M., & Hale, K. (2014). Applied research methods in public and nonprofit organizations. John Wiley & Sons.

Cleary RE (1992) Revisiting the doctoral dissertation in public administration: An examination of the dissertations of 1990. *Public Administration Review* 52(1): 55–61.

Cleary RE (2000) The public administration doctoral dissertation reexamined: An evaluation of the dissertations of 1998. *Public Administration Review* 60(5): 446–455.

Dewey, J. 1938. *Logic: The Theory of Inquiry.* New York: Holt Rinehart and Winston.

Durm, M. W. (1993). An A is not an A is not an A: A history of grading. In *The educational forum* (Vol. 57, No. 3, pp. 294-297). Taylor & Francis Group.

Farmer, J. L., & Costello Jr, W. A. (2019). Social Media as an Innovative Policy Tool: Lessons and Recommendations from the City of Austin. Journal of Public Management & Social Policy, 26(2), Social-Media.

Gilgun, J. F. (2015a). *Deductive qualitative analysis as middle ground: theory-guided qualitative research.* Amazon Digital Services LLC, Seattle, WA, USA.

Gilgun, J. F. (2015b). Beyond description to interpretation and theory in qualitative social work research. *Qualitative Social Work*, *14*(6), 741-752.

Grinnell, R.M. (2001). *Social work research and evaluation: quantitative and qualitative approaches*. New York: F.E. Peacock Publishers.

Guba, E.G.: What have we learned about naturalistic evaluation?, *Evaluation Practice*, **8**(1), 23–43 (1987).

Hedrick, T. E., Bickman, L., & Rog, D. J. (1993). *Applied research design: A practical guide*. Sage Publications.

Heyvaert, M., Maes, B., Onghena, P.: Mixed methods research synthesis: definition, framework, and potential. Quality & Quantity, **47**(2), 659–676 (2013).

Hickman, L. 1990. *Dewey's Pragmatic Technology*. Indiana University Press.

Hothersall, S. J. (2019). Epistemology and social work: enhancing the integration of theory, practice, and research through philosophical pragmatism. *European Journal of Social Work*, *22*(5), 860-870.

Howe, K.R.: Against the quantitative-qualitative incompatibility thesis or dogmas die hard, *Education Research*, 17(8), 10–16 (1988).

Hyde, K. F. (2000). Recognising deductive processes in qualitative research. *Qualitative market research: An international journal*.

James, O., & Van Ryzin, G. G. (2019). Rates and the judgment of government performance. *Journal of behavioral public administration*, *2*(2), 1-10.

James, W. 1907. *Pragmatism: A New Name for Some Old Ways of Thinking*. The Riverside Press.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational researcher*, 33(7), 14-26.

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of mixed methods research*, 1(2), 112-133.

Kaplan, A. 1964. *The Conduct of Inquiry: Methodology for Behavioral Science*. Chandler Publishing Co.

Lerer, D. (2012). Big Things in Small Packages: Evaluating the City of Berkeley's Nanotechnology Ordinance Effectiveness as a Model of Targeted Transparency. Pace Environmental Law Review 30, i.

Levers, M. J. D. (2013). Philosophical paradigms, grounded theory, and perspectives on emergence. *Sage Open*, *3*(4), 2158244013517243.

McDonald III, B. D., Hall, J. L., O' Flynn, J., & van Thiel, S. (2022). The future of public administration research: An editor's perspective. *Public Administration*, *100*(1), 59-71.

McKevitt, D. (2015). Debate: Value for money—in search of a definition. *Public Money & Management*, *35*(2), 99-100.

McKevitt, D. (2017). Harvesting public audit knowledge: implications for theory and practice. *International Journal of Auditing*, *21*(2), 190-197.

McSwite, O.C. (1997). *Legitimacy in Public Administration: A Discourse Analysis*. SAGE.

Menand, L. (2002). *The metaphysical club: A story of ideas in America*. Macmillan.

Morse, J. M., Swanson, J., & Kuzel, A. J. (Eds.). (2001). *The nature of qualitative evidence*. Sage.

Muiruri, K. E. F., Bwisa, H., Muturi, W., & Kihoro, J. (2017). Investigation of the adoption of strategic planning by small and medium sized manufacturing firms in Kenya. *International Journal of Advanced Research in Management and Social Sciences*, 6(4), 105-126.

Nowell, B., Kate Albrecht, A Reviewer's Guide to Qualitative Rigor, *Journal of Public Administration Research and Theory*, Volume 29, Issue 2, April 2019, Pages 348–363,

Nutley, S., Levitt, R., Solesbury, W., & Martin, S. (2012). Scrutinizing performance: how assessors reach judgements about public services. *Public Administration*, *90*(4), 869-885.

O'Connor, M. K., Netting, F. E., & Thomas, M. L. (2008). Grounded theory: Managing the challenge for those facing institutional review board oversight. *Qualitative Inquiry*, 14(1), 28-45.

Perry, J. L., & Kraemer, K. L. (1986). Research Methodology in the "Public Administration Review," 1975-1984. *Public Administration Review*, *46*(3), 215–226. https://doi.org/10.2307/3110436

Provost, F., & Kohavi, R. (1998). Guest editors' introduction: On applied research in machine learning. *Machine learning*, *30*(2-3), 127-132.

Ritchie, J., Lewis, J. (2003). *Qualitative Research Practice: A Guide for Social Science Students and Researchers*. Sage, London (2003)

Saunders, T. C., & Holahan, J. M. (1997). Criteria-specific rating scales in the evaluation of high school instrumental performance. *Journal of Research in Music Education*, *45*(2), 259-272.

Smith, J.K.: Quantitative versus qualitative research: an attempt to clarify the issue. *Education Research*. 12(3), 6–13 (1983a).

Smith, J.K.: Quantitative versus interpretive: the problem of conducting social inquiry. In: House, E. (ed.) Philosophy of Evaluation, pp. 27–52. Jossey-Bass, San Francisco (1983b).

Smith, J.K., Heshusius, L.: Closing down the conversation: the end of the quantitative-qualitative debate among educational inquirers. *Education Research*, 15(1), 4–12 (1986).

Springer, J. F., Haas, P. J., & Porowski, A. (2017). Applied Policy Research. Garland Publishing, Inc., London.

Stallings RA and Ferris JM (1988) Public administration research: Work in PAR, 1940–1984. *Public Administration Review* 48(1): 580–587.

Stebbins, R.A.: Exploratory Research in the Social Sciences. Sage, Thousand Oaks (2001)

Strydom, H. (2013) An evaluation of the purposes of research in social work. *Social Work/Maatskaplike Werk* **49**(2), 149–164.

Sutton, R. I., Staw, B.M. (1995). What theory is not. *Administrative Science Quarterly*. 371–384.

Twining, P., Heller, R. S., Nussbaum, M., & Tsai, C. C. (2017). Some guidance on conducting and reporting qualitative studies. *Computers and Education,* Vol. 107 pp A1 – A9.

van Popering-Verkerk, J., Molenveld, A., Duijn, M., van Leeuwen, C., & van Buuren, A. (2022). A Framework for Governance Capacity: A Broad Perspective on Steering Efforts in Society. *Administration & Society*, *54*(9), 1767-1794.

Verhoeven, N. (2007). *Doing research: The Hows and Whys of Applied Research*. Lyceum Books.

White, J. D. (1986). Dissertations and publications in public administration. *Public Administration Review*, 227-234.

Worster, W.T. (2013) The inductive and deductive methods in customary international law analysis: traditional and modern approaches. *Georgetown Journal of International. Law* **45**, 447- 521.

Yin, R. K. (2017). *Case study research and applications: Design and methods*. Sage Publications.

Yun, J. A., Hamidullah, M. F., & McDougle, L. M. (2020). An analysis of gender differences in public administration doctoral dissertation research. *Journal of Public Affairs Education*, *26*(1), 73-95.

Zhu, L., Witko, C., & Meier, K. J. (2019). The public administration manifesto II: Matching methods to theory and substance. *Journal of Public Administration Research and Theory*, *29*(2), 287-298.

[1] See for example Perry & Kraemer, (1986), Stallings & Ferris, (1988), Addams & White, (1994), Cleary (1992, 2000), Yun et al (2020) and McDonald et al (2022).

[2] See for example Babbie (2007), Adler & Clark (2008), Strydom (2013), Babbie & Mouton (2010), Blinded, (2006), Blinded, (2017)

[iii] University accreditation by regional accreditation bodies like the Southern Association of Colleges and Schools are an example.

[iv] For an excellent discussion of the value of qualitative, inductive empirical research in public administration see Nowell & Albrecht (2018).

[v] Nowell and Albrecht (2018) stress the horizontal connections within a research project and argue that different kinds of research questions are" more or less appropriate to different designs" (p. 352). We are arguing something similar – different kinds of research purposes are appropriate for different types of frameworks. The combination of purpose and framework lead logically to different kinds of designs. Note that our focus is on deductive research. The inductive research that Nowell and Albrecht (2018) focus on would not have front end frameworks.

[vi] Micro-conceptual frameworks refer to theoretical tools like a set of hypotheses that are operationalized in a particular empirical study.

[vii] Here are a few examples. A hypothesis is "an idea or *explanation* that you then test through study and experimentation" (*Vocabulary.Com*), "an idea or *explanation* for something that is based on known facts but has not yet been proven (*Cambridge Dictionary*), "a proposed *explanation* for a phenomenon" (*Wikipedia*). Italics added.

[viii] Description and categorization are also very basic. To describe the content of a kitchen, look at the drawers and cabinets. These common features of a kitchen are tangible categories. The kitchen would not function if the plates, forks, and spices were in the same drawer. We naturally categorize to make sense of the world and to cook!

[ix] We agree with Nowell and Albrecht (2019) about the importance of distinguishing between deductive and inductive research approaches.

[x] For example, Asselineau et al (2022) in an *Administration & Society* article emphasize the importance of performance measures and develops a system to rate (gauge) the effectiveness of numerical targets using the measures aligned with the real pursued goal and their use of additional information in the assessment. Numerical targets which align with the goal and are supplemented by additional information are rated as "ideal". Whereas numeric targets unaligned with goals and un-supplemented received the "worst situation" rating.

[xi] Note that applied research does not need to demonstrate that it fills a gap in the literature. Rather, it should help resolve in some way a problem such as the lack of a succession plan or the need for an in-house assessment of a fire department's incident command system. When the frameworks of applied gauging research draw on the professional literature they connect theory and practice and should reveal the relevance of PA scholarly literature for practice.

[xii] This endnote will provide examples of gauging student papers. The sources are omitted to ensure a blind review. See Reyes, O'Neill, Vaden, Pogue, McLemore.

[xiii] We were surprised at the relative paucity of references to applied research. Aside from Hedrick et al (1993) we found just a few more such as Brown & Hale (2014) and Springer, Haas & Porowski (2017).

[xiv] Note that this article is an outgrowth of recommendations of a scrutinizing effort by NASPAA's Commission on Peer Review and Accreditation.

[xv] Thanks to Stu Bretschneider for using this terminology as one of the authors sat on NASPAA's accrediting body engaged in decisions about which programs received accreditation or reaccreditation. He guided members of the accrediting body stressing that they were acting in the capacity as both coach and cop.

[xvi] Note, the results of scientific studies can be used as evidence and can be used to provide proof in a legal context.

[xvii] Note that any one gauging study could use purely quantitative or qualitative approach.

[xviii] For a discussion on incompatibility theory see: Guba (1987); Howe (1988); Smith (1983a; 1993b); Smith et al. (1986).

[xix] To this list we also add Charles Sanders Peirce, Oliver Wendel Holmes, Jr., and Jane Addams (Menand, 2002).

[xx] In an article that compared nine mixed methods studies Heyvaert et al, (2013, p. 663)_ could find an explicit framework in only two studies. And, these frameworks were vague (transformative and pragmatic) (p. 663).

[xxi] The term conceptual framework is applicable in a broad context (See Ravitch & Riggan, 2012). The micro-conceptual framework narrows to the specific study and informs data collection.

[xxii] Examples of published students research will be included. They are not currently included to ensure author anonymity.