Deconstructing Public Administration Empiricism

A paper written for

Anti-Essentialism in Public Administration
a conference in
Fort Lauderdale, Florida
March 2 and 3, 2007

Larry S. Luton
Graduate Program in Public Administration
Eastern Washington University
668 N. Riverpoint Blvd., Suite A
Spokane, WA 99202-1660
509/358-2247 voice
509/358-2267 fax
lluton@mail.ewu.edu
Deconstructing Public Administration Empiricism:
Short Stories about Empirical Attempts to Study Significant Concepts and Measure Reality

Larry S. Luton
Eastern Washington University

Like much social science empiricism, public administration empiricism at its best recognizes that it presents probabilistic truth claims rather than universalistic ones. In doing so, public administration empiricists are (to a degree) recognizing the tenuousness of their claim that their approach to research is better connected to reality than research that does not adhere to empiricist protocols. To say that one is 95% confident that an empirical model explains 25% of the variation being studied is not a very bold claim about understanding reality.

They are also masking a more fundamental claim that they know an objective reality exists. The existence of an objective reality is the essential foundation upon which empiricism relies for its claim to provide a research approach that is superior to all others and, therefore, is in a position to set the standards to which all research should be held (King, Keohane, & Verba 1994; Brady & Collier, 2004). In one of the most direct expressions of this claim, Meier has stated “There is an objective reality” (Meier 2005, p. 664). In support of that claim, he referenced 28,000 infant deaths in the U.S. in 2000. He went on to attribute about one-third of those deaths to an inadequate health care system. Recognizing that the one-third figure was conjectural, he described the suspiciously rounded 28,000 number as real. This brief anecdote points toward the direction of my argument. This paper will not challenge the existence of an objective reality; it will take an agnostic position on that question since it’s unclear to me whether we can ever be certain about such a thing. The basic argument here is that public administration empiricist claims to measure reality are quite problematic. Refusing to dispute the existence of infant deaths is not the same as conceding the accuracy of an attempt to measure them.

My paper’s approach will be based on close examination of some of the best public administration empiricist research available. It will begin by reviewing key aspects of the empiricist claim to be connected with reality. It will then examine those aspects by focusing on specific examples of public administration empirical research, research published in top rank peer review academic journals, examples that most any public administration researcher would recognize as quality work. The shortcomings found in such work can, I think, be fairly characterized as representing typical shortcomings of public administration empirical research.

In addition, since research is only valuable if it addresses important questions (cf. Lynn, Heinrich, & Hill, 1999, p. 11), this paper will focus on research that has done so. In order for empirical work to make a significant contribution to the growth of knowledge, it needs to be thoughtfully connected to important theory and concepts. The issues addressed by the research examined here include red tape, representative bureaucracy, and performance measurement. It seems unlikely that many would question whether the concepts of red tape, representative bureaucracy, and performance measurement have been important ones for public administration in the last several decades. Pandey has identified red tape research as an area where much
progress has been made in recent decades (2002). Meier has described the past 30 years’ work on representative bureaucracy as “high quality empirical work” (Meier, 2005, p. 663). Performance measurement has also specifically been recognized as an “empirical exemplar” (Miller & Fox, 2006, p. 19).

In sum, using focused attention on three narratives, this paper will suggest that because of the tenuousness of their claim to do research that is connected to reality, the knowledge claims of public administration empiricists are so weak that the field of public administration ought not be persuaded to grant them the increased space in public administration journals that they seek (cf. Gill & Meier, 1999, p. 193).

The foundational claims of public administration empiricists

Having granted the possibility that an objective reality exists, I begin with the empiricist claim that their work is connected to it. At its heart, that claim suggests that the variables used in empirical research correspond to specific parts of objective reality, that they are valid measures of what it is that they claim to be measuring. Others have taken issue in general with the credibility of that claim (e.g., Miller & Fox, 2006), but this paper’s approach will be to examine the stories that specific studies tell to determine whether that claim has been adequately supported.

A related claim that empiricists must support in order to deliver on their promise to study objective reality in a useful fashion is that the measures that they utilize are reliable. The reliability of a measure depends upon its consistency. “We can measure something accurately only when we can also measure it consistently” (Leedy & Ormrod, 2005, p. 29).

A third claim upon which empiricist research approaches rest is that the relationships their statistics measure between variables are indicative of direct influences, not spurious relationships. If a correlation is found between or among variables, it may not mean what it seems to mean if there is an undetermined third variable that ‘intervenes’ in the relationship. Public administration empiricists claim that they engage in ‘precise thinking,’ which should help them to avoid problems like spurious relationships (e.g., Lynn, Heinrich, & Hill, 2001, p.6).

One of the ways that they assure the precision of their thinking is by carefully constructing models of the relationships they study. Those models specify the relationships they expect to find among variables and are tested to discover whether we should have confidence in them. Sometimes those models include suggested causal relationships, something that Lynn, Heinrich and Hill have described as the central theoretical challenge faced by empirical research (2001, p.17). Translating the models into hypotheses, the hypotheses are then subjected to statistical tests that give results which either support the hypotheses or fail to support them. This is the current manifestation of Popper’s falsification theory (1959). According to that theory empirical claims cannot ultimately be verified, but they can be falsified. The centrality of falsifiability to empiricism is testified to by the frequency with which it is pointed to in the arguments made by scientists against intelligent design proponents (cf. Kitzmiller v. Dover Area School District, 2005).

Finally, the public administration empiricist claim that binds together all of the previous claims is that by utilizing their methodology, adhering to its standards and delivering on its promises, they are building a body of knowledge (cf., Lynn, Heinrich, & Hill, 2001, pp.11, 14). In order to build a body of knowledge they would need to link their work to the work of others in a way that cumulates and holds together.
Having reviewed the above claims that are fundamental to public administration empiricists arguments that their research approach is superior to (and sets the standards for) all other research approaches, it is now time to examine whether they adhere to their standards, deliver on their claims, and make good on their promises.

**Red Tape**

As I mentioned earlier, a key factor in selecting the empirical research that is examined in this paper is that it addresses an important concept in the field of public administration. It would be difficult to identify a concept that is more identified with the field of public administration than ‘red tape.’ Given the amount and depth of his work on studying the concept of red tape using empirical methods (DeHart-Davis & Pandey, 2005; Pandey, 1995; Pandey & Breschneider, 1997; Pandey & Kingsley, 2000; Pandey & Scott, 2002; Pandey & Garnett, 2005; Pandey & Welch, 2005; Rainey, Pandey, & Bozeman, 1995; Scott & Pandey, 2000 & 2005), Pandey is clearly one of the most accomplished public administration empiricists working on this topic. In his own opinion, in the last thirty years articles on this concept have paid “great attention to carefully defining red tape and specifying operational measures deliberately … [and] are definitely ‘giant leaps’ for red tape research” (2002, p. 566). (More on this claim about building a body of knowledge later.) It is, therefore, fair to say that Pandey’s work on red tape recognizes and is part of the best empirical work done on that concept. To the extent that his work fails to fulfill the claims made for public administration empiricism, one might begin to wonder about the ability of public administration empiricists to deliver on their claims.

In one of his most thorough articles, Pandey and Scott (2002) review the history of empirical work on red tape to address the questions regarding whether “different measures of red tape and formalization measure the same underlying realities” (p. 553). They also conduct an empirical examination of operational measures of red tape to examine “the relationship between theoretical construct and organizational reality … the link between organizational reality and red tape” (p. 555). Including detailed information on the operational definitions, survey items and computational steps taken (see Exhibit 2, p. 570), the article is an exemplary model of transparency for empirical work. As Wright, Black and Flournory have pointed out, “without an adequate understanding of the research process, one can hardly expect to understand, replicate, or build upon the research findings of others” (as cited in Lynn, Heinrich, & Hill, 1999, p. 113).

The source of data for this 2002 *Journal of Public Administration Theory and Research (JPART)* article was a data set containing 857 observations that was compiled in the work of the National Administrative Studies Project (NASP). NASP includes researchers from the University of Denver, Florida State University, the University of Georgia, Ohio State University, and Syracuse University. They have gathered data in Colorado, Florida, and New York from a variety of public, private, and nonprofit organizations, each having at least ten employees. In a quick review of the literature it appears that over a dozen peer reviewed journal articles in such highly rated journals as the *JPART, Public Administration Review (PAR), Administration & Society (A&S)*, and *Public Performance and Management Review* have been produced utilizing this data set.

This 2002 article (like the others based on this data set) utilizes several approaches to measuring red tape, involving a global measure and three more specific
measures of personnel, administrative delay-, and number of decision makers-based red tape. The study includes some items Pandey and Scott characterize as attitudinal and some they describe as objective. As will become apparent below, however, both kinds of measures involve judgments and evaluations. Judgments and evaluations are part of organizational reality, but they are better described as attitudinal than objective. To the degree that public administration empirical research relies upon or is influenced by judgments and evaluations, it is not justified in its claim that it is linked to objective reality.

The global measure is one they recognize as attitudinal. It asks their respondents to use a scale of 0-10 (with 10 representing the most red tape) in making the following evaluation: “If red tape is defined as burdensome administrative rules and procedures that have negative effects on the organization’s effectiveness, how would you assess the level of red tape in your organization?” As I interpret it, this item involves the respondents in making two judgments to reach the overall evaluation requested. In order to determine whether red tape exists in their organization, and if so how much there is, they must decide what ‘burdensome’ and ‘negative’ mean. Since it is not possible to actually observe red tape, they must infer its existence based on their feelings about whether administrative rules and procedures are burdensome and have negative effects. It is not difficult to understand why Pandey and Scott did not characterize this as an objective measure of red tape.

Pandey and Scott also recognize their measure of personnel red tape as attitudinal. It is based on summing the scores on five Likert scale items. A more complex construction, this measure involves requiring their respondents to make seventeen judgments, five causal and twelve evaluative. In the interest of saving space, I will specify the four judgments involved in only one of the five items utilized in this measure: “The rules governing promotion make it hard for a good manager to move up faster than a poor one.” The causal judgment required in this indicator is whether it is the rules that cause the difficulty. Since the respondents can be safely assumed not to have engaged in empirical research to determine the cause of the difficulty, this item is reliant upon their unscientific judgment. The evaluative judgments in this item involve interpreting the terms ‘hard,’ ‘good,’ and ‘poor.’ The item does not ask respondents what qualifies as hard. If good managers do sometimes move up faster (as could be the case), how often would this need to take place or how much faster would it need to happen for it not to be ‘hard’? What criteria are used to identify the ‘good’ managers? What criteria are used to identify the ‘poor’ ones? Whatever answer one might get to these last two questions, it is quite doubtful whether each of the respondents utilized the same criteria. This undermines any claim to consistency and reliability. These kinds of considerations also explain why Pandey and Scott recognize that objective organizational reality has not been captured by their attitudinal measures.

Pandey and Scott describe their measures of administrative delay- and number of decision makers (or approvals)-based red tape as objective, but these measures also require the making of judgments, this time by the researchers instead of the respondents. Let’s look closely at the administrative delay-based measure. In order for the researchers to categorize a response to this measure as an indicator of red tape, the researchers had to make judgments about reported lengths of time between initiating hiring, firing, and reorganizing requests and approval of those requests. What numbers indicate delays? Moreover, how can the researchers be confident that the amounts of time, say for reorganizing, were caused by ‘burdensome rules and procedures’ rather than something else—perhaps by members of the organization.
resisting the request? Moreover, the answers respondents gave to the survey items could be better indicators of the respondents’ attitudes toward their organization than measures of organizational reality. They were not asked to report the amount of time between a specific request and its approval, they were required to generalize. It is doubtful that the respondents collected data on a number of requests and reported the mean. Their estimate of the time required was most likely influenced by their attitudes.

Thus, the objectivity of the information being studied is at least questionable. It is not at all clear what they were measuring. If Pandey and Scott had been able to measure red tape objectively, they should be able to tell us what a ‘unit’ of red tape is. As has been seen they have good reasons for not claiming to be able to do that. Nonetheless, Pandey and Scott should be given much credit for the thoroughness of their research and the thoughtfulness with which they conducted it. The concept of red tape is not directly observable. As their global measure recognized, seeing something as red tape requires interpretation. Seeing organizational problems as being caused by red tape involves more interpretation. Finding links between such a concept and organizational reality requires observation and judgment. It is not strictly an empirical matter. It is not even significantly an empirical matter. It is inherently a matter of interpretation. The problem is not with the quality of the research but with the ability of empirical public administration research to deliver on its claim that it measures reality.

Moreover, there is nothing about the empirical research on red tape that meets the expectation of falsifiability. While Pandey and Scott evidence concern about whether the concepts of formalization and red tape are confounded, their work is based on an assumption that red tape exists and does not allow for that assumption to be falsified. The relationship between red tape and challenges involved in personnel administration are similarly not falsifiable. The same holds true with regard to administrative delays and numbers of decision makers.

Even if we were to accept the work based on this data set as an empirical contribution to knowledge, there remains a problem with the claim to be cumulating a body of knowledge. Pandey and Scott do a marvelous job of summarizing and analyzing the previous literature, but that in itself is not sufficient. As was pointed out earlier, much of the recent empirical research on red tape derives from analysis of the NASP data set. Mining a data set does little to build a body of knowledge. Repetition is not accumulation.

So, the story of empiricist research on red tape fails to deliver on several essential claims made by empiricists. It is not objective, reliable, falsifiable or cumulative.

**Representative Bureaucracy**

The second public administration empirical research story to be deconstructed in this paper focuses on research on representative bureaucracy. According to Meier, “What has happened to the concept of ‘representative bureaucracy’ over the past 30 years is an excellent example of conceptual development that included both high quality empirical work and precise thinking by theorists. … This type of sustained long term interplay between empirical work and conceptual thinking is desperately needed for dozens of other key public administration concepts” (Meier, 2005, p. 663). Clearly, the empirical public administration research on representative bureaucracy has been identified by one of the leading advocates for public administration empiricism as an exemplary model for the field. Examining one of the contributions
to the representative bureaucracy literature in which Meier participated (Andrews, Boyne, Meier, O'Toole, & Walker, 2005), thus, gives us an opportunity to see how well this high quality empirical work meets empiricist standards and claims.

In a 2005 *JPART* article, Andrews et al. address the argument that representative bureaucracy results in organizations that are more responsive to the public. In order to facilitate the use of empirical data to examine that argument, they reduce that claim to a performance claim: “Representative bureaucracy will be associated with higher levels of organizational performance” (p. 492). Unfortunately, the result is that instead of examining the argument that a representative bureaucracy will be more responsive, they test the relationship between representative bureaucracy and performance.

Further, prior to reporting their empirical research, they replace that reduced (but still significant) performance claim with a trivial one: “the representative bureaucracy—performance relationship is moderated by management” (p. 490) [my emphasis]. As a result, instead of exploring the relationships between significant concepts, they have set as their central contention a claim that is so obvious it hardly seems worth examination. But this should not be simply dismissed as a failure of this specific example of empirical research. The need to work with measurable data often results in reduction of significant concepts to less significant, but measurable, data. Not infrequently, that reduction goes farther than may be theoretically or conceptually sound. This is a case in point, not an aberration.

The measures of performance that Andrews et al. use are attempts to measure public perceptions of performance. Public perceptions of performance are at best, incomplete indicators of performance. They may be remarkably poor indicators of performance. As Catlaw (2005) has pointed out, the dynamics of public perceptions of bureaucracy may not be as simple as is suggested by the model used by Andrews et al. for their empirical examination. If Catlaw is correct, improvements in performance may have little or nothing to do with whether the public’s perception of a bureaucracy improves. Whether Catlaw’s theory is correct or not, I suspect that public administration empiricists would readily agree that perceptions are often not reliably linked to reality. However, Andrews et al. never address that problem.

Moreover, the data upon which Andrews et al. rely for their empirical examination are even weaker proxies for the concept of responsiveness than is the concept of performance. Their measures related to representative bureaucracy are data (originally created for England’s central government) on population deprivation and diversity and workforce diversity. Their deprivation measure is an average score derived from thirty-three indicators covering six domains: income, employment, health, education, housing, and access to services. Readers of their article are provided no direct explanation for how the collection of this data met empirical research standards. People who have been involved with data collection know that this is an issue that should not be ignored. “Empirical studies are only as good as their data. All of the benefits of empirical social sciences research … are predicated on accurate measurement” (Lynn, Heinrich, & Hill, 2001, p. 113). Those who have been involved in data collection by government agencies know that it is possible that social science standards have been compromised. Moreover, as Yanow (2003) has demonstrated, ethnic minority data collected in a census may itself be a poor reflection of the reality it purports to measure.

In addition, although the article does not report sufficient detail to be sure, it is reasonable to suspect that the deprivation measure is an arithmetical mean calculated using incommensurable and/or ordinal numbers. Social science and public
administration empiricists often manipulate ordinal numbers as if they were interval data, but statisticians are generally agreed that it is not proper to use ordinal numbers in calculating a mean. This practice is not unusual among social science empiricists; for some reason they are not bothered by acting as if ordinal numbers are more precise than they are. In addition to raising questions about the accuracy of their measurements, this lack of precision undermines their claims to objectivity and reliability.

To measure population diversity Andrews et al. take the percentages of each of the fifteen ethnic minority groups identified in the census, square them and subtract the sum of the squares from 10,000. It is not made clear why they thought this “Herfindahl index” approach was an appropriate mathematical computation. Without an explanation of their reasons for utilizing a calculation that is commonly used to assess the degree of monopolization in a market, one might be reasonably left with a concern that each mathematical manipulation weakens the measure’s ability to claim to reflect objective reality. Some might be tempted to conclude that the mathematical manipulations were used to “mask the absence of evidence” (Miller & Fox, 2006, p. 14), but sometimes empiricists seem to become so enamored with their statistical manipulations that they fail to make them sufficiently transparent. In this article, this potential weakness in the data is never addressed, but it would not be fair to dismiss this as a criticism specific to this article. That kind of omission is commonly found in public administration empiricism.

Andrews et al. use data from the same sources to measure organizational performance. One performance variable is customer satisfaction as recorded by the Audit Commission as one of its Best Value Performance Indicator measures. It is described as “percentage of citizens satisfied with the overall service provided by their authority” (p. 495). The other performance variable derives from an index called “Core Service Performance” (covering six aspects of performance) that is calculated by the Audit Commission. The Commission gives each aspect a score (1-4), based on judgments by others that they translate or convert to achieve a common metric. They also weight the scores “to reflect their relative importance and budget” (p. 496). Those weights vary based upon whether they come from a county, a metropolitan borough, unitary authority, or London Borough. “The weighted data are then converted to the 1-4 scoring system” (p. 496, note 6). Andrews et al. then “converted those results to overall performance scores” (p. 496). The measures of performance, then, included one relatively straightforward attitudinal measure and another rather complicated index where the final variable summarizes six measures that begin as judgments made by respondents then are subjected to translation, conversion, and/or weighting by the Audit Commission before they are converted by Meier and associates. The attitudinal measure Andrews et al. recognize as somewhat subjective. The complicated index which derives from respondent judgments on six areas that are converted, weighted, and converted again is characterized by Andrews et al. as the ‘objective’ measure. It is rather questionable whether this measure can be considered an objective reflection of reality, since it begins with respondents’ judgments (not observable behavior), and then is placed in a hall of mirrors. As was the case in the examples above, the computational manipulations obfuscate any connection there may be between the data and reality. In this instance—since the data clearly rely upon judgments made by the respondents, are weighted by the Commission, and further manipulated—they would be better described as ‘creative’ than ‘objective.’

To measure management (as in “management moderates performance”), Andrews et al. gather original data by asking 4,184 informants in 386 English local
authorities questions designed to facilitate categorizing their strategic styles into one of three categories: prospector, defender, or reactor. Complete data was obtained only from 846 informants in 80 authorities. The prospector style was operationalized by asking whether informants saw their service or authority as being “at the forefront of innovative approaches” (p. 498). The defender style was operationalized based on judgments regarding “[whether their approach to service delivery focused on their ‘core business’]” (p. 498), and the reactor style was operationalized based on an inquiry into the extent to which they waited for direction to come from auditors and inspectors. Clearly all three of these measures are subjective in nature, but while Andrews et al. admit that using strategy as a proxy for management means that management is “somewhat underspecified” (p. 498), they do not address the lack of objectivity in the measure. As the red tape research also demonstrated, it can be very difficult to gather objective data. It could be that because of the nature of public administration as an object of study, the challenge is too difficult for public administration empiricists to meet. It could be that objective reality is a very small portion of public administration reality.

In any case, given the distance from reality and the opportunities for distortion involved in the construction of these measures, it would be reasonable to have rather low confidence in them. To their credit Andrews et al. describe their findings as “provocative but clearly not definitive” (p. 501). Moreover, they express concern about the meaning of performance and the proper way to measure it. It would have been reasonable for them to express similar concerns about diversity and representative bureaucracy.

Any body of knowledge built with such uncertain materials, instead of accumulating will disintegrate. The internal workings are compromised by cancerous imprecision and the conceptual framework is so weakened by reductions and substitutions that it no longer provides even a promise of connecting to significant issues.

Performance Measurement

The third empirical story to be deconstructed in this paper involves studying performance measurement. As was noted earlier, even among post-modernists, performance measurement has been recognized as an “empirical exemplar” (Miller & Fox, 2006, p. 19). The work to be examined here is based on a large data set provided by the Texas K-12 educational system. That data set appears to have been used by public administration empiricists to produce at least fifteen articles that have appeared in highly ranked peer review journals, including PAR, A&S, JPART, the American Review of Public Administration and the Journal of Policy Analysis and Management.

One article that has been produced as from this educational administration data set seems particularly appropriate to examine in this paper. It was published in PAR and received the Mosher Award for the best article by an academician in 2004 (O’Toole & Meier, 2004). One of the authors, Laurence O’Toole, received the 2004 Dwight Waldo Award for outstanding contributions through an extended career. As an award winning academic article published by award winning authors in the primary journal associated with the American Society for Public Administration, it can fairly be characterized as an example of the best empirical work done in the field of public administration. Nonetheless, it, too, falls short in delivering on the empiricist claims for why and how their work is the standard bearer for how to do public administration research.
To begin with, one might question whether the field of public administration should accept the Texas school system as indicative of the general practice of performance measurement. It would also be fair, as was the case in the research on representative bureaucracy, to question whether the data set available from more than 1,000 Texas schools contains data in which one should place confidence. Most anyone who has experience with administratively generated data knows that its quality can be compromised in a number of ways, from the way it is gathered to the way it is coded or reported. More specifically, a story told by Miller and Fox (2006) gives reason to question the reliability of the data collected by Texas schools, but O’Toole and Meier provide no explanation of why they have confidence in the data set that has been made available to them. If they had generated the data themselves, they could reasonably have been expected to explain how their procedures met social science expectations. Having obtained the data from others, they are not absolved of the responsibility for assuring that proper procedures have been followed. Again this is not a problem just for the article presently being addressed. In their review of public administration literature, Wright, Black, and Flournory (1999) noted “the nearly complete neglect of issues regarding the reliability or validity of the research measures” (as cited in Lynn, Heinrich, & Hill, 1999, p. 117).

The data itself also fails to connect the dots between reality and hypotheses. Perhaps the data most tightly connected to reality are their measures for “resources allocated to the core instructional function,” but those data are better characterized as measures of money spent on instruction (percentage spent on instruction, per student spending on instruction, teachers’ salaries) than resources allocated to the core instructional function. Using the budgetary resources allocated as an indicator of the resources allocated to a program is a common administrative practice, but those who do so generally recognize that it is an exercise in ‘picking the low hanging fruit.’ The only measure reported for performance is percentage of students who pass all portions of the Texas Assessment of Academic Skills (TASS). The article contains no recognition of the controversy over whether the TASS represents a reasonable indication of student performance. Moreover, apparently they examined other performance measures, but did not report anything about those results (p. 347). For the purposes of this article, they presumably were quite satisfied with the TASS as a measure of performance. There is more reason to have concern about the quality of their measures. When O’Toole and Meier address the determinants of contracting, they acknowledge Boyne’s critical review of research on that topic and point out that he was concerned about both poor measures and reciprocal relationships, but they choose to ignore the problem of poor measures and focus on the problem of reciprocal relationships. Their ready reliance on a measure about which there has been considerable public policy disputation may be a case in point supporting Stivers’ concern about public administration’s tendency to utilize empiricism to provide cover for treating answers as final when they are, or ought to be, subject to continuing political discussion (forthcoming).

O’Toole and Meier also have a problem in meeting social science research expectations with regard to assigning causality. When it comes to the relationship between increasing bureaucracy and increasing contracting, they address the ‘correlation does not equal causation’ problem by using panel analysis—an approach based on the idea that if A happens before B it could be a causal factor. The results of their panel analysis, however, show what they characterize as “reciprocal causation” (p. 349). In years following increases in bureaucracy, contracting increased. In years following increases in contracting, bureaucracy increased. Based on finding this
pattern in a longitudinal study covering only three years, they conclude that bureaucrats promote increases in contracting and contracting promotes increases in bureaucrats. They never appear to have considered the possibility that there was an ‘undetermined third variable’ (or variables) underlying the results they found. It could be that an increasing student population created the conditions for both increases. It could be that increased monitoring and/or reporting requirements led to the increase in bureaucrats. Similarly, their interpretation of the meaning of positive correlations between contracting and revenue per pupil and percentage of funds from local sources is that in wealthy districts contracting is used to provide enhancements such as guest conductors for orchestras (p. 348). They fail to consider the possibility of an ‘undetermined third variable’ driving the correlations. It could be, for example, that the people and the school boards of wealthy districts tend to have ideologies favorable toward market models and their ideologies mean that they are inclined toward contracting. One generally finds caveats about this dynamic in introductory research textbooks (e.g., Leedy & Ormrod, 2005), so it is surprising to see it so inadequately addressed in an award winning article.

Moreover, O’Toole and Meier are inconsistent in their willingness to address problems arising from looking at correlations in order to address the issue of causality. Whereas, in the relationship between bureaucracy and contracting they recognize the possibility of reciprocal causation, they simply interpret a positive correlation between high levels of teacher turnover and contracting as indicating that the instability that resulted from teacher turnover led to an institutional lack of capacity that was filled by contracting (p. 348). Their only interpretation of that correlation posits that the turnover caused the contracting. Based on the data and its statistical analysis, however, it is also possible that increases in contracting created the conditions that gave rise to higher teacher turnover. They never address that (or any other) possibility.

This tendency to be satisfied with a single causal interpretation, instead of considering a set of reasonable interpretations is not limited to one or two instances. Just to cite one additional example, they interpret a finding that a negative correlation exists between percentage of the budget spent on instruction and percentage contracted as meaning that “contracting constricts instructional resources” (p. 346). It is just as likely that other factors, such as reduced budgets and/or greater demands on budgeted dollars, are responsible for constricting the percent of the budgets that were spent on instruction.

It would appear that even award winning public administration empiricism fails to meet reasonable expectations arising from social science standards and empiricist claims that they measure reality and engage in precise thinking to connect theory and concepts to that reality.

Conclusion

Miller & Fox have concluded that “the proposition that performance can be reduced to variables that can be measured, and that outcomes, results, effects, and even nonevents can be attributed to programs and policies, amounts to fanciful faith in what social science methodology can accomplish” (2006, p. 37), but one need not go that far to conclude that public administration empiricists have not yet delivered on their claims of having a methodology that is superior because it connects with reality. Performance measurement promises much more than it has delivered. Red tape and
representative bureaucracy have not been shown to be objective realities capable of measurement. They look more like metaphors that have been reified.

As Miller & Fox have noted, “Reality, it turns out, is exceedingly difficulty to measure or represent” (2006, p. 15). Because measuring reality is exceedingly difficult, I do not want to be as hard on the authors of these three public administration empiricist stories as might be some of their colleagues. Lynn, Heinrich and Hill, for example have taken the following position: “Studies that fail to deal forthrightly with the limitations of concepts, methods, and data deserve the disrespect of practitioners” (Lynn, Heinrich, & Hill, 2001, p. 157). There are many reasons why public administration empiricists may not always live up to the standards they espouse. As Lynn, Heinrich, & Hill (2001, p. 18) have recognized, a central problem in empirical research is obtaining data that allow them to explore a range of theoretical possibilities, not simply test one hypothesis. A more general problem that all public administration scholars face is the practical matter that journals accept articles but impose space limitations. Those space limitations make it difficult if not impossible to address all of the limitations of their research in each published article. It could also be that their rational self interest in being published disinclines authors from addressing such limitations.

I do not want to ignore or minimize the difficulties public administration empiricists face in living up to their billing. I am more interested in suggesting that they ought to lighten up in their claims about how much their research contributes to our enlightenment. The stories I have presented, for example, suggest that they ought to at least qualify the common empiricist claim that they measure objective reality while others engage subjective interpretation. The phenomena they measure are not facts they found in observing organizational reality. Most often they have been created for the research process—or possibly worse—for the bureaucratic accountability process.

My short stories about these public administration research efforts do not add up to a definitive grand narrative about public administration empiricism. Based as they are on an interpretation of the narratives provided by the specific researchers examined, they are simply suggestive of the kinds of challenges faced by public administration empiricists and the ways that those researchers have or have not adequately come to terms with those challenges. The moral of my story is that upon close reading, the stories offered by public administration empiricists decompose.

References


