

Jason L. Jensen
University of North Dakota

Robert Rodgers
University of Kentucky

Cumulating the Intellectual Gold of Case Study Research

Two criticisms of public administration research have been leveled. First, knowledge in the field is not being cumulated. Second, the research has low quality. A preference for case studies as a form of scientific inquiry is behind both criticisms. The authors propose a solution to the first problem by showing how meta-analysis can be used to cumulate knowledge using case study evidence. Viewed cumulatively, case studies comprise an intellectual goldmine awaiting discovery. The authors challenge the second criticism, proposing that quality judgments should be based on the important principle of knowledge cumulation, which acknowledges the value of all research methods.

Evaluations of public administration research have permeated *Public Administration Review* for over a decade (Perry and Kraemer 1986; Stallings 1986; Stallings and Ferris 1988). Two themes pervade the literature criticizing research in the field: Critics contend that knowledge is not being cumulated (Adams and White 1994; White et al. 1996) and that research in general (dissertation research in particular) is of poor quality (Adams and White 1994; Cleary 1992; McCurdy and Cleary 1984; White 1986a; White et al. 1996). Criticisms that public administration research lacks quality remain largely unchallenged (see Bailey 1992 and Box 1992, for two rare exceptions).

The use of case studies as a preferred research methodology in public administration lies behind both themes. Critics contend that case studies play a limited role in knowledge cumulation (Adams and White 1994) and that they fare poorly on indicators of quality (Adams and White 1994; McCurdy and Cleary 1984). Both claims can be traced to the perception that case studies are not generalizable, a perception we show to be terribly misconceived.

We contend that public administration is well suited to case studies because they satisfy the recognized need for conditional findings and in-depth understanding of cause and effect relationships that other methodologies find difficult to achieve. Case studies are a key part of the solution, not part of the problem. We will show this by first defining case studies and by outlining a typology of case studies. Next, we will show how the knowledge-cumula-

tion problem can be solved by using meta-analysis. Finally, we will explain how the criticism that case studies have poor quality is based on a problematic definition of quality and misguided criteria.

The critics' test of quality generally involves counting the number of quality criteria that are satisfied in a study. The higher the count, the better the quality of the study. That is, the criteria are additive (see McCurdy and Cleary 1984). When a whole range of quality measures are applied to any single case study, it fares badly. Similarly, when a whole range of quality measures are applied to a study using a different methodology (survey research, for example), it is also likely to fare badly. The errors of this logic lie with the application of all quality criteria to each single, isolated study pertaining to a particular point in time (or several points in time), and with the assumption that all measures of quality are appropriate for all methodologies.

For example, consider a quality indicator such as the subtlety and richness of findings embedded in a meaningful historical context. A study that is not a narrative

Jason L. Jensen is an assistant professor at the University of North Dakota in the Department of Political Science and Public Administration. His research interests include public organizations and management, as well as research methods in public policy and administration. He has used meta-analysis to study the diffusion of public organizations across states and the resulting rise of organizational populations. Email: Jason_Jensen2@und.nodak.edu.

Robert Rodgers is a professor of public administration at the University of Kentucky. He champions the use of case study research for dissertations. Email: pub708@ukcc.uky.edu.

case study would fare poorly on such a criterion. Measures of quality should make sense for the methodology being used, with an understanding that good-quality studies, using a variety of methodologies, will add up to produce knowledge over time and that knowledge cumulation is actually dependent on the use of multiple methodologies. Exclusion of any methodology undermines knowledge cumulation.

The charge that case studies are of poor quality stems, in part, from the perception that they are not generalizable, and thus cannot be cumulated. Both of these criticisms are untrue. Case studies are often done in clusters. Even when not so designed, they can be considered retrospectively as a group for a broader and richer analysis. When case studies are considered cumulatively, a wide variety of special conditions can be recognized to ascertain whether the findings are generalizable.

Meta-analysis aims to solve the related issues of cumulation and generalizability; the cumulation of case studies using meta-analysis allows specific tests of generalizability. Meta-analysis is a method of combining findings across research studies that has become increasingly popular in the social sciences (Hedges and Olkin 1985; Hunter et al. 1983). With 982 meta-analyses published in the social sciences between 1980 and 1993 (Bausell et al. 1995), a revolution in knowledge cumulation has occurred in many fields.

For many public administration topics, only a meta-analysis that includes case studies can provide a solution to the knowledge-cumulation problem. If we abandon case study research, the cumulation of knowledge about some public administration topics—whose study is most conducive to case study methodologies—will not occur. Thus, the continued use of case study methodologies is essential for knowledge to advance.

We will compare our solution—continued use of case studies and meta-analysis to address the knowledge-cumulation problem—with the implicit solution of public administration research critics. The critics' solution entails conforming all research to the notion of quality as they have defined it—a path that discourages the use of case studies. Our evaluation of this solution leads us to conclude that they are incorrect about the quality of research in the field. Case studies contain golden nuggets that await discovery.

What Is a Case Study?

Case study research in public administration has an esteemed history. In 1951, pioneers of case study research formed the Inter-University Case Program (ICP), which included members from almost 50 institutions. Members of the ICP met to develop ideas for cases and to refine the

case method. The group published an acclaimed treatise on the case method (Bock 1962) and published cases on many aspects of public administration (see Stein 1952a).¹

ICP case writers focused on the processes and critical events that lead to administrative decisions. They crafted narratives about a variety of administrative processes, regardless of whether the entity under study was a single decision maker, a single agency, or a complex governing system. ICP case authors also showed an overriding concern for rigor and generalization, which Stein (1952b, xxi–xxiv) termed a “constant” process. They urged that generalizations be tentative due to their complexity.

From its inception, ICP members expressed concern over the scientific utility of case study research. Fesler (1962) outlined several contributions that case studies can make in the research process, including sorting out the plausibility of hypotheses, aiding the formation of hypotheses, and providing data for hypothesis testing. Waldo (1962, 63) recommended making case studies more scientific and generalizable, but he saw “no responsible alternative” to using the case method as a research tool.

The ICP group had a sound notion of the meaning of *case study*, an involved definition to which Stein devoted a large part of his 1952 essay on public administration cases. He described how case studies differ across fields, but “common ground exists in the inherent intellectual approach: examination of particulars prior to or as a part of generalization” (1952b, xx). In general, an ICP case study in public administration is a narrative of decision processes, a definition that centers on the method of analysis as opposed to the research setting or the particular entity being studied.

Over time, the term “case study” has been applied to an ever-increasing group of studies, and the definition has evolved to center on the entity under study. For instance, Adams and White (1994) proposed a broad definition of case studies as “research conducted within a single agency or political jurisdiction, or research that spanned a handful or less of single agencies or political jurisdictions for comparative purposes” (573). The Adams and White definition blankets a large number of studies under the general heading of case study, including most research on policy implementation and bureaucratic effectiveness, which generally focuses on a single agency or a few agencies. For example, Moe (1985) considered the National Labor Relations Board, Scholz and Wei (1986) considered OSHA, and Wood (1988) considered the Environmental Protection Agency.

Although definitions of the term case study have become broader, an overriding concern for generalizability remains at the root of all definitions. The generalizability issue has led most critics to identify case study research as contributing to poor-quality research in public administra-

tion. For instance, Adams and White (1994) concluded that—compared to five other fields—the quality of dissertation research in public administration is poor in three main areas: (1) practice research (most often case studies); (2) foreign focus research; and (3) case study research.

The Entity under Study

Individual case studies, by their nature, are difficult to generalize. We define a case study as any study where generalization to a larger population is called into question² because the study focuses on a single entity. That is, the underlying sample size of a case study is always equal to one. The entity in question might be an organization, a decision, a department, an agency, a person, an institution, a policy, a city, a state, or a county. A concern that is shared by both advocates and critics of case studies is the lack of generalization in case study research. By our definition, the specific entity being studied takes center stage because generalization requires that there be comparison across entities.³

By “entity” we refer to the research setting as opposed to the research question or research method. To the extent that case study writers focus on process, they often treat the research entity as having secondary importance. But, if generalization is to occur, we must ask how well a process generalizes to similar entities. By our definition, a case study can take many different forms and use many different methodologies. The particular methodology used in a case study (whether it is qualitative or quantitative) makes no difference to our definition.

A case study may address several research questions, examine several processes, or survey a large sample of individuals or subunits. For example, a case study of NASA might include information on external institutional relationships, such as interaction with Congress, the president, or outside contractors. Additionally, such a case study might refer to internal processes, such as launch decisions or the hierarchical structure of the organization. The generalizability of findings regarding any of these processes can be questioned because results apply only to NASA, and the underlying sample size is equal to one. Even so, such a case study possesses intellectual gold that is absent from a large-sample study of many agencies, because detailed information is reported about specific conditions that are present and critical events that occur. The public administration study with the richest information is thus a case study that reports detailed information about the conditions, critical events, and processes of a single entity.

Case studies do, to some extent, trade detail for generalizability. Critics of case studies believe this kind of detailed information has little worth, because it is never apparent whether findings are generalizable to other entities. Critics of large-sample studies, on the other hand,

believe that rich details specific to each entity in a sample are overlooked or simply assumed away. However, when the findings of case studies are analyzed cumulatively, the foundation for both criticisms vanishes because the generalizability of each study is addressed while the richness of detail is preserved.

A case study, by our definition, can be qualitative or quantitative. The methods used to test for effects in case studies with quantitative data are analogous to the “within,” “between,” and “mixed- subject” designs used in experimental settings. In public administration, the unit of analysis is generally not an individual person, but a larger entity (organization, institution, policy, etc.). More specifically, a researcher could test hypotheses by identifying variance between entities within a single unit (comparing different programs or departments in an agency), across an entity over time (a longitudinal analysis of a program or programs in an agency), across entities in a larger population (comparing program effects in many agencies), or across several entities in a larger population over time. Variance may occur between entities (programs, agencies, etc.) in a single time period (between-cases design), across time in a single entity (within-cases design), or as a combination of the two (mixed design). Clearly, even if a case study utilizes hypothesis testing and sophisticated quantitative methods, generalizability is still questionable. A case study may be qualitative and descriptive without quantitative hypothesis testing and still contain useful information related to one or more of the design categories above. The manner in which the entity is studied thus assumes many different forms.

In the interest of disentangling the complexities of different designs and approaches, we introduce a typology of case studies in the next section. This typology lays the foundation for categorizing case studies and shows how case study evidence can be cumulated using meta-analysis.

A Typology of Case Studies

Snapshot Case Studies

Case studies reflect a wide variety of designs. Critics claim that the least meaningful case study is a “snapshot” description of a single organization (or other entity) at a single point in time.⁴ Such studies (for example, Mintzberg 1971; Taylor 1916) would certainly not have high quality according to the critics’ standards, but nonetheless have a distinguished history in public administration. Snapshot studies utilize various qualitative and quantitative approaches, or some combination of the two. In the case of a snapshot study of an agency, hypotheses are often formally tested with a between-cases design, utilizing a cross-section of programs, departments, job types, or even persons.

The qualitative snapshot study in public administration was refined by the ICP as a narrative of decisions and processes; what the ICP deemed the “case method” has consequently become a popular mode of analysis throughout the field’s history. A narrative approach to the snapshot study is generally detail-rich. The type of study conducted by ICP authors was not autobiographical, but rather was written from the perspective of a detached observer. It excluded the abnormal, made few value judgments, and covered a specific time period (Stein 1952b). Qualitative case methods include many different approaches and have enjoyed a resurgence and refinement in technique, particularly in the past decade (see Orosz 1997, for a review of resources for public administration researchers).

Longitudinal Case Studies

Case studies can be longitudinal, providing a different type of information than the snapshot study. This type of study is a time-ordered analysis of events that occur during a period of the entity’s history (DiIulio 1994; McCurdy 1991; Ross and Staw 1986). Longitudinal case studies in public administration may focus on political entities or institutions (such as presidencies), on a particular agency (see Wood 1988), or on policies, programs, or decisions. Longitudinal case studies may be quantitative or qualitative in character and may involve a formal report and analysis of critical events or processes. In the case of an agency, the focus could be on changes that occur organization-wide or in sub-units of the organization, or on responses to external factors that influence outcomes (Wood 1988). Time is the organizing device and the dynamics of change are the primary focus. In some longitudinal case studies, the full experience of an entity is reported from its birth to its demise.

Pre-Post Case Studies

The “pre-post” study goes a step beyond the longitudinal design to provide evidence on the outcomes of implementing a particular program, policy, or decision. It is longitudinal in design, but it also includes an assessment before implementation of the program, policy, or decision and follow-up assessment after implementation. Pre-post case studies are more sophisticated in the sense that causal inferences can be drawn. Once results are systematically aggregated, cause-and-effect assertions become possible.

One problem that has plagued public administration research is the mishmash of studies that assume many different forms (snapshot, longitudinal, pre-post) and focus on a single entity. Consider any large government agency, policy area, or governmental unit (federal, state, or city), and chances are very good that there are many case studies in the literature of the same entity (a within-unit vari-

ance situation). Similarly, consider any policy, program, or process, and chances are good that there are several case studies that span organizations, institutions, or other entities (a between-unit variance situation). Each case study may have a different purpose, use a different approach to evaluating the research question, and cover a different time period. How can knowledge from case studies with different purposes, using different methods, over different time periods, be cumulated? Researchers have attempted to address this question by engaging in studies which integrate the findings from several case studies. This reflects a first step toward cumulating the findings of research that bear on a particular research question. Such integrative studies assume two forms—the patchwork case study and the comparative study.

Patchwork Case Studies

Patchwork studies integrate several case studies that have evaluated a particular entity at different points in time as snapshot, longitudinal, or pre-post designs. Roethlisberger’s (1941) patchwork summary of the Hawthorne studies is a classic example of a patchwork case study. Results of each case study are used to construct a patchwork case, which becomes longitudinal since it has the advantage of drawing on within-case variance provided by analyses of the same entity at different points in time.

The strengths of the different forms of case studies—snapshot, longitudinal and pre-post—all contribute to the richness of the patchwork case study. Any single case study provides only a glimpse of an organization, policy, or decision, which is limited in scope, time, and perspective (that is, not generalizable). Patchwork studies create a more complete and holistic picture of the dynamics associated with the entity. The patchwork case study is limited, however, in the sense that it focuses on only a single entity.

Comparative Studies of Cases

A second type of cumulative study, similar to the patchwork study, also integrates the findings of several case studies. The objective of the comparative study, however, is cross-unit comparison as opposed to within-unit synthesis. Comparisons are made in an attempt to tease out generalizations about an underlying commonality reflecting a policy, process, program, or decision. Comparative studies broaden the perspective to include many different entities, and thus offer insights that can be unearthed from comparative analysis.

The comparative study has the potential to yield important scientific discoveries (for example, Allison 1971; Kaufman 1960; Tipple and Wellman 1991). Baumgartner and Jones (1993) did a comparative case analysis of policy agenda-setting (as a process) across several distinct policy

areas. Both qualitative and quantitative comparisons were made in an effort to generalize. The comparative study, however, falls short in systematically integrating results from other studies that examined the same research question.

The most meaningful analysis will assemble the findings of all case study evidence on a particular topic, thereby integrating all of the scientific evidence into a single analysis. No evidence is initially discarded. The method that accomplishes this purpose is meta-analysis, which solves the knowledge-cumulation problem because the findings of studies using different formats, designs, and methodologies are analyzed together.

The first step in solving the problem of knowledge cumulation is to raise the level of analysis so that all of the evidence from case studies is synthesized and considered cumulatively, providing a golden opportunity to evaluate change within units as well as between units over time. To accomplish this feat, meta-analysis is used to cumulate the findings of single studies in a formal and systematic fashion.

Our Solution: Cumulate Case Study Evidence Using Meta-Analysis

Although it is relatively new to public administration, meta-analysis has rapidly become a popular way of cumulating knowledge. At least six meta-analyses with public-sector applications have been published in the research literature. Yin and Heald (1975) studied the effects of decentralization in agencies by cumulating the findings of 215 case studies. Miller et al. (1991) analyzed the relationship between technology and organizational structure by cumulating the findings of 25 case studies (which included 33 samples). Rodgers and Hunter (1992) evaluated the impact of public-sector management-by-objectives programs by cumulating the findings of 30 case studies. Wolf (1993) used meta-analysis to compare several theories of bureaucratic effectiveness as reported in 44 case studies (which included 85 observations). Wolf (1997) also studied entrepreneurial administrative reforms by analyzing 170 case studies of 104 U.S. federal agencies. Finally, Robertson and Seneviratne (1995) studied planned organizational change by integrating the findings of 47 independent samples, all of which were taken from case studies. These meta-analyses used slightly different techniques, but all analyzed evidence from case studies.

Critics devalue case study evidence because they think only in terms of each case study as an independent finding. They fail to appreciate each case study's value to the whole network of research and to the final synthesis. Data from single studies (each study may have more than one result), with the appropriate adjustments,⁵ are the pieces from which a complete database of study evidence, orga-

nized around a specific unit of analysis, can be constructed. Data from case studies are then entered into a data set of findings and analyzed using the tools and methods of meta-analysis. A cumulative meta-analysis of findings from many case studies is a powerful test of generalizability.

We have described how case studies represent a range of foci and methods, all of which are centered on a particular research entity. A view of quality based on knowledge cumulation respects all types of case studies, since each provides something that others miss. Application of the quality criteria advocated by critics of public administration research would lead us to conclude that some, if not all, types of case studies are of low quality. We will show how the generalizability of individual case studies—the foremost concern underlying any form of case study research—can be tested by cumulating evidence from individual case studies into a meta-analysis.

Testing the Generalizability of Case Study Results

Drawing conclusions about an important theoretical framework, or comparing alternative theories pertinent to the character of a field, is arguably the most worthwhile mission in science. Indeed, “the foundation of science is the cumulation of knowledge from the results of many studies” (Hunter et al. 1983, 10). Meta-analysis systematically tests the generalizability of case study results by using evidence that examines the same phenomena in different settings, at different times, and utilizing different methodologies. The results from a summary analysis of case studies reflect the cumulative evidence from many studies of the same setting, policy, process, or program. Thus, the generalizability of any single study is a function of the degree to which it conforms to the results of other cases that have been entered into a meta-analytical database.

The Research Question

To cumulate knowledge from case study research successfully and to address generalizability, a meta-analyst must begin with a research question. By specifically defining a research question, the researcher will also define the universe of case study evidence bearing on the question. For example, suppose the research question is the effect of process reengineering on organizational productivity. The unit of analysis is the organization. Any case study that considers the consequences of reengineering on productivity may be relevant. However, because the meta-analyst can test generalizability only by cumulating across organizational entities, the context of each single case study becomes important. Conceivably, 100 studies might consider the effects of process reengineering, but only 30 of them report evidence relevant to productivity outcomes.

Thus, based on the research question, only 30 of the studies can be entered into a database for meta-analysis.

If the research question of the meta-analysis is direct and simple, the search for case study evidence that addresses the question directly or indirectly will be straightforward. Conversely, if the research question is complex, entails several secondary questions, or reflects a theory with multiple conditions, the search for study evidence becomes more difficult. By delineating a research question at the outset, the meta-analyst narrows the range of applicable evidence that must be gathered and defines the limits of generalizability.

Case Study Design

The research question of each individual case study relates to the study design that is used to address the question; thus, the meta-analyst must also give attention to issues of study design. Regardless of whether cases are qualitative or quantitative, or whether they test for within- or between-unit effects, meta-analysis affords the opportunity for a systematic, quantitative analysis of findings. Meta-analysis can use evidence that has been extracted from snapshot (for cross-sectional analyses), longitudinal (for longitudinal analyses) or pre-post case studies (for mixed analyses and causal assertions). Evidence from patchwork and comparative studies may also be entered into the meta-analytic database. Since the population of study evidence is the existing body of all published and unpublished case study evidence, it should represent the unit of analysis under study. Findings always apply to the existing body of research evidence, so meta-analysis is comparable to a comprehensive literature review that is quantitative and systematic.

When critics subject the findings of just one case study to the evaluation criteria, generalizability and validity are called into question. For instance, when a single case study assesses a program or policy before and after a particular intervention, results may show a change. Since only one entity was studied, there are issues of internal validity (extraneous factors may have caused the change) and external validity (the result may not generalize to other organizational contexts). The opportunity to draw a definitive conclusion with regard to causality emerges, however, at the point when the pre-post study evidence is evaluated cumulatively.

Consider, for example, the cumulative analysis of pre-post case study evidence reported by Rodgers and Hunter (1991, 1992) and by Rodgers, Hunter, and Rogers (1993). They analyzed all case studies that evaluated the impact of management by objectives (MBO) on organizational productivity or job satisfaction (the research question). Because they were considering the impact of MBO on specific outcomes, they selected pre-post case studies which

evaluated MBO programs during different years and under different temporal influences. All studies showed performance gains after MBO implementation. Their performances improved during periods of downsizing and rightsizing, during periods of Democratic and Republican administrations, etc. Thus, regardless of the external events present at the time of the study, meta-analysis showed that all public organizations experienced an improvement in performance. This is both a causal assertion and a statistically powerful demonstration of generalizability.

Threats to Validity

Confounding factors that may threaten generalizability can also be tested with meta-analysis. There may be a question about whether studies with or without a certain feature should be included in the analysis of study evidence, or if the presence or absence of a feature somehow influences (or confounds) the cumulative result. Any hypothesized threat to validity that can be coded can be tested as a potential moderator variable (or confounding factor) in a meta-analysis through quantification and classification. For instance, variables that could affect the relationship under study include the author's involvement in program implementation, whether the study was published, the use of control groups, the use of quantitative or qualitative methods, or the use of retrospective assessments. Studies can be grouped according to the presence or absence of the feature in question (see Rodgers and Hunter 1994). If the hypothesized confounding factor is actually moderating the relationship under study, the cumulative result will be different for studies in which the feature is present.

Sample Size

If the research question of the meta-analysis involves microphenomena (such as job satisfaction), then sample sizes and reliability estimates may be used to make statistical adjustments to the effect sizes in individual case studies, even though the data apply to a single entity. Weighting effect sizes according to sample size acknowledges the sampling error that is associated with each study. Correcting for attenuation recognizes the degree of unreliability that is associated with the single-study estimate. Conversely, if the research question focuses on a macroentity (that is, an organization, institution, or policy), each case study of that entity counts as a single data point ($N=1$) in the cumulative analysis and no sample-size adjustment is indicated. If a study reports findings for multiple entities, it is assigned more weight. For example, a comparative study of three agencies will have three times the weight of a pre-post study of one agency. Thus, weighting the findings of case studies on the basis of sample size and correcting for statistical artifacts increases the generalizability and validity of the findings.

Standardization of Single-Study Outcomes

The variety of approaches used to report the outcomes of case studies are an obstacle to generalization. Meta-analysis solves this problem by converting case study outcomes to a common metric through standardization. By converting each case study outcome to a common metric, it is possible to derive a cumulative result for all studies. Further, confounding factors can be tested to determine whether they moderate the cumulative result under examination. For example, does the cumulative result change as a function of the type of agency, location of the agency, or level of government? Standardization of single-study outcomes depends on whether the data from the single studies are qualitative or quantitative.

Standardization of qualitative case studies entails using various content-analytic procedures which result in quantitative data used in the meta-analysis. Essentially, the case author's narrative statements register on a predetermined rating scale (for instance, a decision was made or a decision was delayed; the outcome was positive, negative, or there was no change). Preferably, several meta-analytic raters are used and inter-rater reliability is estimated for each scale.

Quantitative single-study results are standardized by choosing a common metric and using available formulae to convert the results of each case study to this common metric (for example, an r or a d statistic). Effect sizes for each single study can be calculated from any of several univariate statistics reported in the studies (most commonly, F -test, t -test, chi-square value, r -value, and means and standard deviations of a variable before and after an event). Since an "effect size" is calculated for each study, it is possible to derive an estimate of the cumulative effect size (or mean effect) for all studies.⁶

Unfortunately, there is a formidable challenge in converting the findings of most case studies to a common metric because of the lack of standardization in reporting formats. Three steps in particular will alleviate the challenge of locating study evidence and converting study findings to a common metric, thereby facilitating the use of meta-analysis as an accepted means of knowledge cumulation in public administration.

First, standards should be adopted by public administration journals to include—for quantitative articles—a correlation matrix and means and standard deviations of all study variables so that single-study effect sizes are readily available. Such standards already exist in several other fields.

Second, standards should be adopted for the publication of qualitative case studies. Our suggestions regarding such standards are listed in table 1. Standards include stating the research question(s) to ease the meta-analyst's search for relevant case studies.⁷ We have explained why a

single research entity underlies all case study research and why generalization of case study outcomes can only occur with regard to like entities. Thus, clear definition and detail of the entity under study is also a crucial reporting requirement of case researchers.

Table 1 Template for Reporting Case-Study Findings

Research question: What is the primary research question of the case study? What secondary questions are related to the primary question?	
Related case studies: What other case studies report evidence bearing on the same research question(s)?	
Research entity: What is the primary entity under study and the specific setting in which the research took place (institution, organization, department, city, state, county, country, etc.)? Identify sub-entities.	
Information breakdown and size of the sub-entities: Is evidence on outcomes reported for subunits (for instance, evidence on departments within an agency that is the primary entity)? If so, what are the sub-entities studied and what is the size of samples within each entity (departments, subjects, etc.)?	
Design: Is the study a snapshot, longitudinal, pre-post, patchwork, or comparative design?	
Measurement: How were outcomes assessed (qualitative and/or quantitative)? What outcomes were observed (productivity, performance, satisfaction, turnover, goal accomplishment, decision, etc.)?	
Threats to validity: What factors may have influenced the outcomes?	
Time frame: Describe the intervention (if any) and all critical events within the context of a time frame. For example:	
Results: How did outcomes change over time (improve, deteriorate, or show no change)? What were the outcomes for sub-entities?	
Additional information: Who should be contacted for further information?	

The third, and perhaps most important, recommendation is that the Inter-University Case Program be reinstated. The ICP would set reporting standards conducive to knowledge cumulation, refine case methodologies for use in public administration research, and develop the application of meta-analysis for the cumulation of both quantitative and qualitative public administration research.

Testing the Generalizability of Public Administration Theory

Each case study contains a rich body of information that can be used to evaluate the conditions for which a theory is held to be true. The advantage to the meta-analyst is that detailed descriptions of potential confounding factors are almost always available in the case study report for extraction and entry into the database.

Even if the needed study evidence is missing, a researcher can usually obtain the missing information by contacting the author directly or by extracting it from archival data. The unit of analysis and author are almost always revealed in a case study, so a researcher is in a position to collect the additional data and subject any theory to a rigorous empirical test. Pre-post and longitudinal case studies provide information about change across time (the “within condition”), and all types of case studies provide the evidence needed for cumulative analysis of change across units (the “between condition”). Virtually any theoretical condition can thus be coded and treated as a moderator variable in the meta-analysis (Hunter and Rodgers 1995).

To illustrate how the generalizability of theory can be explicitly evaluated with meta-analysis, consider again Rodgers and Hunter’s (1992) meta-analysis of pre-post case studies evaluating the impact of MBO in different agencies. The MBO program was found to be more effective with strong commitment and support from top management. Theory tells us that if top management commitment is present, productivity should be greater, regardless of the type of organization in which the program is implemented. The theoretical importance of top management commitment was thus found to be generalizable across a variety of organizational forms.

Meta-analysis can be used to study questions and theories that are central to the field. Such pursuits are cutting edge because the studies have not yet been cumulated and the researcher is providing interpretation of a body of literature in a way that has not been done before (see Miller et al. 1991). Definitive conclusions to theoretical stances that are important to the field can be drawn, and novel theoretical propositions can be advanced using evidence from case studies.

An indicator of importance may be the number of studies written about a particular theory or topic. Meta-analysis requires a critical mass of studies that can be cumulated. Thus, any meta-analysis that analyzes the findings of public administration case study evidence is also a study that, by definition, addresses a topic that is central to the field. Analysis of study findings is organized around the key question posed in the meta-analysis, and all study evidence that bears on the question is identified and included in the meta-analysis. Since the meta-analysis is topic-driven, case studies appearing in print across all disciplines are included and analyzed. Thus, the findings of a meta-analysis will be a rich test of generalizability because they are based on a large number of cases reflecting various methodologies and contexts.

Meta-analysis of case studies satisfies all concerns about generalizability that are germane at the single-study level and, at the same time, solves the knowledge-cumulation problem. Thus, meta-analysis is a logical solution to the two alleged problems in public administration research. Our solution, however, is contrary to the implied solution of critics, which is to abandon the use of case studies as a form of scientific inquiry in favor of studies that satisfy all quality criteria simultaneously.

The Critics’ Solution: Conform Single Studies to Quality Criteria

The implied solution of the critics allegedly improves the overall quality of research in the field by applying a rigid set of quality criteria for each study to satisfy. For their solution to succeed, several conditions must be met. First, there must be agreement about what constitutes a high-quality study. Second, application of the same criteria should replicate across evaluators. Third, all criteria should be systematically related to one another since all purportedly represent an underlying dimension of study quality. Finally, the classics in the field should satisfy the criteria of a high-quality study. We will show in the following section that none of the above conditions are satisfied.

Lack of Agreement about the Criteria

By advocating certain criteria to measure the quality of a study, critics of public administration research implicitly advocate a particular type of study. For instance, McCurdy and Cleary (1984) were the first to evaluate the quality of public administration dissertations by using the following criteria: (1) purpose; (2) validity; (3) impact 1 (the testing of theory); (4) impact 2 (casual relationships); (5) importance 1 (central topic); and (6) importance 2 (cutting edge).

If the quality of a single study is a quantifiable construct open to reliable evaluation, there must be agreement on how to assess it. Table 2 lists four studies (in-

Table 2 Specific Criteria Used to Judge Dissertation Quality

McCurdy and Cleary (1984)	White (1986a)	Cleary (1992)	Adams and White (1994)
Research has a purpose	Purpose	Research has a purpose	Framework guiding the research
Validity	Validity	Validity	Flaws in the research
Impact #1: testing of theory	Theory testing	Impact #1: testing of theory	Relevance of the findings
Impact #2: causal relationships	Hypothesis testing	Impact #2: causal relationships	Importance of the topic
Importance #1: central topic	Causality	Importance #1: central topic	Overall quality
Importance #2: cutting edge		Importance #2: cutting edge	

cluding McCurdy and Cleary 1984) that evaluate the quality of dissertations. Taken as a whole, the four studies apply 11 different criteria to evaluate quality. Note that not one criterion is used by all of the studies to evaluate the quality of dissertation research. Thus, there is lack of agreement about which quality criteria should be applied. It is also not clear whether studies that appear to use the same criteria are actually rating the same constructs. That is, if a criterion is used in more than one study, there is likely divergence across studies regarding its definition. To address the lack of consistency in how the criteria are defined, we now turn to the level of replicability between evaluators.

Lack of Replicability

To address replicability, we will consider the findings of studies that have evaluated the quality of the same dissertations using some of the same quality criteria. White (1986a) reanalyzed the dissertations used in the McCurdy and Cleary (1984) study. He found that 37 percent met his validity criteria, compared with 21 percent that met the validity criteria in McCurdy and Cleary’s study. But when White rated the studies according to the central importance criteria of McCurdy and Cleary (“importance 1: central topic”), he found that 83 percent met his central importance criterion, while only 45 percent met the criteria of central importance when rated by McCurdy and Cleary. Thus, even when the same criteria and dissertations were used, different results were obtained. Based on the inconsistent use of criteria and the divergent findings when the same criteria were used, we conclude that the quality of any single study is not objectively definable as measured by quality indices.

Lack of Internal Consistency

When considered independently, the criteria in table 2 are appealing as quality indicators because each represents a virtue of positivist social science research. But considered additively, are the indicators an appropriate way to measure the overall quality of any single study considered in isolation? In evaluations of study quality, the authors attempt to measure quality based on the number of criteria that are met by a study. For example, if a study meets all the criteria used to evaluate its quality, it earns the highest-quality ranking, whereas if a study meets none of the criteria, it ranks as the lowest quality. In treating the quality indicators as a single additive index, the critics thus assume that quality can be measured by a single underlying dimension. Hence, the indicators should all be positively correlated with one another if they are indeed measuring the same qualities. That is, it must be possible for all indicators to be present in a single study.

Using McCurdy and Cleary’s (1984) six quality indicators, table 3 indicates that, of the 15 possible pairs of indicators, it is questionable whether four of the pairs can even be present in the same study. First, in public administration it is very difficult to study either a cutting-edge topic or a topic of central importance (as contained in public administration textbooks), while at the same time meeting the validity criterion (experimental or quasi-experimental designs). Note the question mark in table 3 between “cutting edge” and “validity” and between “central topic” and “validity.” Second, it is virtually impossible to test an existing theory and still be cutting edge. If a researcher is interested in testing a theory that is an integral part of public administration, the topic would likely not be considered cutting edge. Note the question mark for the interface of “theory testing” and “cutting edge.” Third, if a topic is important to the field, as evidenced by its inclusion in leading public administration textbooks, then it is extremely unlikely that the topic would be considered cutting edge. Again, note the question mark at the interface of “central topic” and “cutting edge.” These inconsistencies convolute the meaning of quality, showing that it is impossible for any single study to satisfy all of the criteria simultaneously. Thus, there are serious problems with the internal consistency of the quality indexes.

Table 3 Likelihood that Quality Criteria Could Occur Together in a Public Administration Study

Criteria	1 Purpose	2 Validity	3 Theory testing	4 Causal relationships	5 Central topic
1. Purpose					
2. Validity	yes				
3. Impact #1: theory testing	yes	yes			
4. Impact #2: causal relationships	yes	yes	yes		
5. Importance #1: central topic	yes	?	yes	yes	
6. Importance #2: cutting edge	yes	?	?	yes	?

Criteria in the table are from McCurdy and Cleary (1984).
Notes: “Yes” indicates that both criteria could likely be met in the same study. A question mark indicates that meeting both criteria in a single study is questionable.

Lack of External Validity

Thus far, we have been unable to ascertain the type of study that would satisfy the critics’ image of a high-quality study. Indeed, the evidence suggests that quality may not be open to reliable evaluation. Still, if quality does possess certain characteristics, then it should also be the case that research that has been accepted as having high quality should possess all (or most) of the characteristics of study quality espoused by critics. This proposition can be tested by evaluating whether studies generally recog-

nized to have high quality also have the characteristics stipulated by the critics.

Two dissertations that have become classics in public administration are Herbert Simon's *Administrative Behavior* (1945) and Dwight Waldo's *The Administrative State* (1948). Although we would not define either as a case study, it is clear that neither study would be judged as high quality using any of the 11 criteria used by critics. It may even be the case that neither study could satisfy any of the criteria, short of having a purpose. Simon (1991, 88) clearly recognized the importance of his dissertation as a precursor to his life's work: "I would not object to having my whole scientific output described as largely a gloss—a rather elaborate gloss, to be sure—on the pages of *Administrative Behavior* where these ideas were set forth (especially pages 39–41, 204–12, and 240–4)."

It is ironic that a dissertation that ultimately led to a Nobel prize in economics for its author would be considered low-quality according to the standards set forth by current critics of public administration dissertation research.

It is also ironic that studies based entirely in logical-positivism cannot possibly meet the positivist criteria applied by critics. Bailey (1992) specifically cites many public administration classics and concludes, "Despite their long-term value to the field, these studies would fail the scientific tests established by positivist social science" (52).⁸ Indeed, not one of the 51 studies contained in Shafritz and Hyde's (1992) *Classics of Public Administration* would come close to being judged as high-quality using the indices the critics have propagated.

It has been clear all along that a case study fails the test for a high-quality study. The evidence presented so far has led to the conclusion that even if quality is open to objective assessment (which is itself questionable), any single study would be hard-pressed to meet all of the quality criteria simultaneously.

Summary

Discontinuing the use of case studies in an attempt to conform all research to quality criteria is ill advised. We have pointed out that the criteria used to assess quality fall short of a workable solution to the knowledge-cumulation problem because (1) there is no clear agreement on the "right" criteria of study quality; (2) use of identical criteria has yielded inconsistent findings; (3) the criteria themselves are not internally consistent; and (4) even classics in the field do not satisfy the criteria for study quality. Even if we were able to describe a study that would satisfy all criteria simultaneously, relying exclusively on that study type (or any study type for that matter) would undermine knowledge cumulation because the diversity of methodologies so critical to knowledge cumulation would be too narrowly constrained.

Conclusion

Why do public administration researchers rely on case study methodologies to such an extent? Case studies are a tradition of public administration research because the focus is often an aggregate unit of analysis, such as the organization, institution, governmental unit, policy, program, or process. Thus, if we discourage researchers from conducting case studies, we will exhaust the supply of intellectual gold and, therefore, hinder knowledge cumulation. Many people think that case studies are the scourge of the field. It is ironic, then, that the knowledge-cumulation issue cannot be resolved *without* case studies.

The critics are entirely correct in one important regard: Knowledge is not being adequately cumulated in the field. We have proposed a solution to the knowledge-cumulation problem that embraces case studies and recognizes public administration as a field that speaks to the interests and needs of both practitioners and academicians.

The quality of research should not be judged according to quality indices. The answer to a particular research question must embrace many forms of study, using many different methodological lenses. Only through multiple methodologies can a cumulative picture emerge; in this sense, no methodology is inherently any better than another. Moreover, certain methodologies—for a variety of reasons—are more conducive to addressing certain questions. None should be dismissed. Ensuring rigor within any particular methodology is accomplished through dissertation committees and peer reviews of journal articles.

We began this endeavor with a quiet note of concern over the arguments that continue to surface in *Public Administration Review* which imply that case studies lack quality and prevent knowledge cumulation. We have shown how these concerns about generalizability and knowledge cumulation can be addressed, so long as the evidence from diverse methodologies, including case studies, is available for meta-analysis. Case studies are an integral part of knowledge cumulation. The intellectual gold of our field needs to be mined, not undermined.

Acknowledgments

The authors wish to thank Irene Rubin, *Public Administration Review* reviewers, Herbert Kaufman, Julie Olberding, and three anonymous reviewers of the Academy of Management (Public and Nonprofit Division) for helpful comments on earlier drafts of this paper. An early draft of this paper was presented at the Public and Nonprofit Sector Division of the 1997 Academy of Management meeting in Boston.

Notes

1. ICP history is from Harold Stein (1952), from one of our reviewers, and from our communication with Herbert Kaufman. The roots of ICP can be traced to Harvard's Graduate School of Public Administration, which led to the Committee on Public Administration Cases, established in 1948.
2. Any study has generalizability (external validity) issues, but these issues increase proportionately as the number of entities studied approaches one.
3. The degree to which studying a single entity creates generalization issues is always related to the referent population. For instance, a cross-sectional study of political processes in all 50 states may not have generalization issues because the entire population is under study; however, a study of political processes in industrialized countries may have generalization issues if the United States is the only country studied because it is only one out of a population of countries.
4. Our snapshot category would likely include the "practice research" described by Adams and White (1994), who stated, "On every indicator, the practice research dissertations in public administration were of low quality" (572).
5. This may include sampling error, attenuation (unreliability), and range restriction adjustments (Hunter and Schmidt 1990).
6. For elaboration on data conversion, see Hunter and Schmidt (1990) or Hedges and Olkin (1985).
7. We thank two manuscript reviewers for stressing that case study authors need to clearly delineate research questions (what is my case a case of?) so that generalizability can be tested and primary research studies can be more easily identified for meta-analysis.
8. White (1986b) also pointed out that Waldo's *The Administrative State* would not meet the criteria of causality, testability, and validity advocated by McCurdy and Cleary (1984).

References

- Adams, Guy B., and Jay D. White. 1994. Dissertation Research in Public Administration and Cognate Fields: An Assessment of Methods and Quality. *Public Administration Review* 54(6): 565–76.
- Allison, Graham T. 1971. *Essences of Decision: Explaining the Cuban Missile Crisis*. Boston: Little, Brown, and Co.
- Bailey, Mary Timney. 1992. Do Physicists Use Case Studies? Thoughts on Public Administration Research. *Public Administration Review* 52(1): 47–54.
- Baumgartner, Frank R., and Bryan D. Jones (1993). *Agendas and Instability in American Politics*. Chicago: University of Chicago Press.
- Bausell, R. Barker, Yu-Fang Li, Meei-Ling Gau, and Karen L. Soeken. 1995. The Growth of Meta-Analytic Literature from 1980 to 1993. *Evaluation and the Health Professions* 18(3): 238–51.
- Bock, Edwin A. 1962. Introduction. In *Essays on the Case Method*, edited by Edwin A. Bock, v–viii. New York: Inter-University Case Program.
- Box, Richard C. 1992. An Examination of the Debate over Research in Public Administration. *Public Administration Review* 52(1): 62–9.
- Cleary, Robert E. 1992. Revisiting the Doctoral Dissertation in Public Administration: An Examination of the Dissertations of 1990. *Public Administration Review* 52(1): 55–61.
- DiIulio, John D. 1994. Principled Agents: The Cultural Basis of Behavior in a Federal Government Bureaucracy. *Journal of Public Administration Research and Theory* 4(3): 277–318.
- Fesler, James W. 1962. The Case Method in Political Science. In *Essays on the Case Method*, edited by Edwin A. Bock, 65–88. New York: Inter-University Case Program.
- Hedges, Larry V., and Ingram Olkin. 1985. *Statistical Methods for Meta-Analysis*. San Diego, CA: Academic Press.
- Hunter, John E., and Robert Rodgers. 1995. Flawed Theories about Study Flaws. In *Research in Organizational Change and Development*, edited by R. W. Woodman and W. A. Pasmore, 297–335. Greenwich, CT: JAI Press.
- Hunter, John E., and Frank L. Schmidt. 1990. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. Thousand Oaks, CA: Sage Publications.
- Hunter, John E., Frank L. Schmidt, and Gregg B. Jackson. 1983. *Meta-Analysis: Cumulating Research Findings Across Studies*. Thousand Oaks, CA: Sage Publications.
- Kaufman, Herbert. 1960. *The Forest Ranger: A Study in Administrative Behavior*. Baltimore, MD: Johns Hopkins Press.
- McCurdy, Howard E. 1991. Organization Decline: NASA and the Life Cycle of Bureaus. *Public Administration Review* 51(4): 308–15.
- McCurdy, Howard E., and Robert E. Cleary. 1984. Why Can't We Resolve the Research Issue in Public Administration? *Public Administration Review* 44(1): 49–55.
- Miller, Chet C., William H. Glick, Yau-De Wang, and George P. Huber. 1991. Understanding Technology-Structure Relationships: Theory Development and Meta-Analytic Theory Testing. *Academy of Management Journal* 34(2): 370–99.
- Mintzberg, Henry. 1971. Managerial Work: Analysis From Observation. *Management Science* 18(2): B97–B110.
- Moe, Terry M. 1985. Control and Feedback in Economic Regulation: The Case of the NLRB. *American Political Science Review* 79(4): 1094–116.
- Orosz, Janet. 1997. Resources for Qualitative Research: Advancing the Applications of Alternative Methodologies in Public Administration. *Public Administration Review* 57(6): 543–9.
- Perry, James L., and Kenneth L. Kraemer. 1986. Research Methodology in *Public Administration Review*, 1975–1984. *Public Administration Review* 46(3): 215–26.

- Robertson, Peter J., and Sonal J. Seneviratne. 1995. Outcomes of Planned Organizational Change in the Public Sector: A Meta-Analytic Comparison to the Private Sector. *Public Administration Review* 55(6): 547–58.
- Rodgers, Robert, and John E. Hunter. 1991. Impact of Management by Objectives on Organizational Productivity. *Journal of Applied Psychology* 76(2): 322–36.
- . 1992. A Foundation of Good Management Practice in Government: Management by Objectives. *Public Administration Review* 52(1): 27–39.
- . 1994. The Discard of Study Evidence by Literature Reviewers. *Journal of Applied Behavioral Science* 30(3): 329–45.
- Rodgers, Robert, John E. Hunter, and Deborah Rogers. 1993. Influence of Top Management Commitment on Management Program Success. *Journal of Applied Psychology* 78(1): 151–5.
- Roethlisberger, Fritz J. 1941. *Management and Morale*. Cambridge, MA: Harvard University Press.
- Ross, Jerry, and Barry M. Staw. 1986. Expo 86: An Escalation Prototype. *Administrative Science Quarterly* 31(2): 274–97.
- Scholz, John, and Feng Heng Wei. 1986. Regulatory Enforcement in a Federal System. *American Political Science Review* 80(4): 1249–70.
- Shafritz, Jay M., and Albert C. Hyde, eds. 1992. *Classics of Public Administration*. Pacific Grove, CA: Brooks/Cole Publishing Co.
- Simon, Herbert A. 1945. *Administrative Behavior*. New York: The Free Press.
- . 1991. *Models of My Life*. New York: Basic Books.
- Stallings, Robert A. 1986. Doctoral Programs in Public Administration: An Outsider's Perspective. *Public Administration Review* 46(3): 235–40.
- Stallings, Robert A., and James M. Ferris. 1988. Public Administration Research: Work in PAR, 1940–1984. *Public Administration Review* 48(1): 580–6.
- Stein, Harold, ed. 1952a. *Public Administration and Policy Development: A Case Book*. New York: Harcourt, Brace and Co.
- . 1952b. Case Method and the Analysis of Public Administration. In *Public Administration and Policy Development: A Case Book*, edited by Harold Stein, xx–xlv. New York: Harcourt, Brace and Co.
- Taylor, Frederick W. 1916. *The Principles of Scientific Management*. New York: Harper and Brothers.
- Tipple, Terence J., and J. Douglas Wellman. 1991. Herbert Kaufman's Forest Ranger Thirty Years Later: From Simplicity and Homogeneity to Complexity and Diversity. *Public Administration Review* 51(5): 421–8.
- Waldo, Dwight. 1948. *The Administrative State*. New York: Ronald Press.
- . 1962. Five Perspectives on the Cases of the Inter-University Case Program. In *Essays on the Case Method*, edited by Edwin A. Bock, 39–63. New York: Inter-University Case Program.
- White, Jay D. 1986a. Dissertations and Publications in Public Administration. *Public Administration Review* 46(1): 227–34.
- . 1986b. On the Growth of Knowledge in Public Administration. *Public Administration Review* 46(3): 15–24.
- White, Jay D., Guy B. Adams, and John P. Forrester. 1996. Knowledge and Theory Development in Public Administration: The Role of Doctoral Education and Research. *Public Administration Review* 56(5): 441–52.
- Wolf, Patrick J. 1993. A Case Survey of Bureaucratic Effectiveness in U.S. Cabinet Agencies: Preliminary Results. *Journal of Public Administration Research and Theory* 3(2): 161–81.
- . 1997. Why Must We Reinvent the Federal Government? Putting Historical Development Claims to the Test. *Journal of Public Administration Research and Theory* 7(3): 353–88.
- Wood, B. Dan. 1988. Principals, Bureaucrats, and Responsiveness in Clean Air Enforcement. *American Political Science Review* 82(1): 214–34.
- Yin, Robert K., and Karen A. Heald. 1975. Using the Case Survey Method to Analyze Policy Studies. *Administrative Science Quarterly* 20(3): 371–81.