Mixing Qualitative and Quantitative Methods in Political Science: A Primer

Ann Chih Lin

Kenyatha Loftis

Gerald R. Ford School of Public Policy and Department of Political Science
University of Michigan
Ann Arbor, MI 48109

Please address all comments to annlin@umich.edu and loftisk@umich.edu

Prepared for presentation at the 2005 Annual Meeting of the American Political Science Association, Washington, D.C., September 1-4, 2005. All comments very welcome; please address to annlin@umich.edu and loftisk@umich.edu. We gratefully acknowledge funding from the Horace H. Rackham School of Graduate Studies at the University of Michigan .

With the upswing in critical attention to qualitative methods in political science, multiple or "mixed method" research has gained more attention as well. Mixed method work demonstrates the researcher's facility with different methodological tools and allows the researcher to reach a wider audience. The attractions of synergy have proved especially tempting for graduate students, who are getting more varied methods training and see mixed methods research as a way to prove that they are on the leading edge of a wave. There is some irony in this, since mixed-method work has a long pedigree in political science: for example, the 1960 classic, *The American Voter*, used cross-tabulations to assess the answers to closed-ended questions from a representative national survey, content analysis on free-response answers from the same survey, and extended quotations in the discussion of almost every finding. (Campbell, Converse, Miller and Stokes 1960) Indeed, Maxwell and Loomis (2003) cite evidence suggesting that across the social sciences, "a case could be made that mixed methods research was more common in earlier times, when methods were less specialized and compartmentalized and the paradigm wars were less heated." (242) It also may be, however, that as increasing sophistication in methods creates more knowledge about the limitations of each specific method, scholars are abandoning the possibility that developments in technique alone will lead to greater validity and insight. Mixed method work presents the alluring possibility that different methods could be chosen to complement, rather than supercede, each other.

The increase in mixed method work has been accompanied by an increase in literature on how to design these projects. Some (Tarrow 2004) of this literature is primarily intended to explain the contributions of qualitative methods to quantitative scholars; but much of it begins with the assumption that their readers, who have a certain level of sophistication with research

design, are looking to expand their repertoire of research options (Bennett and Braumoeller 2004; Greene and Caracelli 1997; Jick 1979; Tashakkori and Teddlie 2003). This is an important audience, but it is not the only one. Beginning researchers who have methodological tools but little experience in designing projects, and readers who may never consider doing mixed method projects themselves, also have an increasing need to learn how to think about and evaluate mixed method projects.

This article provides a primer for that audience. Instead of offering a comprehensive set of mixed method research designs, our goal is to offer some principles for deciding when mixed method projects are most appropriate, and to distill these into suggestions for practice and evaluation. We start by defining mixed method research as the combination of qualitative and quantitative methods of data analysis within a single project. While we believe that this is the most commonly accepted definition, it is not the only one, and we explain the consequences of defining it in this way. We then present a basic decision tree for use in planning empirical research and we explain where, along its branches, mixed method research is most useful. Our decision tree introduces three purposes for mixed method research designs – supplementary validation, convergent validation or triangulation, and theory generation – and explains the kind of claims associated with each. This tool can help the researcher produce more clarity about her intentions and in doing so, make it easier for others to evaluate the success of her project. Because researchers are often not explicit about their intentions, however, we conclude this paper with a set of guidelines for evaluating mixed-method research. Since the purposes for using mixed methods differ, we argue, the criteria for the successful use of mixed methods must differ as well.

Defining Mixed-Method Research: Data, Analysis, and Theory

Few empirical studies in political science can exist without both words and numbers. A book that uses econometric techniques to examine budgetary data may refer to historical events that place those data in context; a book that examines the development of Rousseau's thought may compare sales figures for his various books. Mixed-method empirical research, therefore, must mean more than research that includes both words and numbers. We define it here as *research distinguished by both qualitative and quantitative methods of data analysis:* by variety in analytical technique, and not simply variety of data.[1] This definition excludes research that uses qualitative/quantitative data simply for the sake of context. Reporting government statistics to set the stage for a discussion of how participants in focus groups talk about tax cuts is not mixed method research; nor is a study of political business cycles that starts with a vignette about some Federal Reserve Board chairman. At the same time, this definition also excludes research where only one set of tools is used for data analysis, even when other sets of tools are used to derive theoretical propositions for testing. Thus, neither a book that uses survey data to examine popular adherence to different philosophical models of justice, nor one that uses game theoretic models to derive predictions for historical cases, would qualify as a mixed-method study.[2] On the other hand, a game theoretic model would qualify if it were tested first through in-depth archival study of a historical event and then with a large-n data set of cases, as would a Foucauldian study

---

[1] Other writers have defined mixed method research differently, in particular Bennett and Braumoeller (2004). Instead of starting with the distinction between qualitative and quantitative methods, they list three types of methods – game theory, econometrics, and case studies – and discuss the benefits of combining them. More generally in the literature, however, the primary distinction is made between qualitative and quantitative methods, with the understanding that both categories contain multiple specific methodologies.

[2] To us, the great strength of game theory for social science is precisely its symbolic, non-empirical system of reasoning, a strength it shares with philosophy. Thus we do not consider game theory an empirical method for data analysis.

of prisons using both participant observation and the quantitative analysis of prison characteristics over time.

Note that the definition of analytical technique here is quite broad. The general term "quantitative" covers techniques as common as multiple regression; as sophisticated as the time-series analysis of autocorrelated economic data; or as simple as percentages in cross-tabulation. Similarly, "qualitative" work includes the compilation and selection of archival material; the creation of ethnographic narratives based on months or years of fieldwork; or the application of semiotic techniques to the selection and interpretation of excerpts from interviews. Defining these terms so broadly will no doubt trouble some scholars, who would argue that as methods training, qualitative and quantitative, becomes more widespread, researchers should be required to do more than rely on basic techniques. But a more restrictive definition runs the risk of defining research as "mixed-method" only if it is "sophisticated" research.[3] Given the high variance in methods training across the discipline, and the active debate over the legitimacy and appropriateness of various methodological techniques, more restrictive definitions would quickly become ideological. Given the time and training it takes to become an expert in more advanced methodological techniques, and/or to conduct extensive fieldwork, it is also unrealistic to expect that all practitioners of multiple methods develop multiple specialties (Bennett and Braumoeller 2004).

Note, also, that the focus on analysis sidesteps questions about research paradigm and epistemology. There is wide agreement, from multiple traditions, that "methods" do not imply "paradigms": that is, that the tools of data gathering and analysis can be used in the service of

[3] This is not to suggest that the sophistication of the methodologies in mixed-method research is not an important question, but only to urge that this be considered a question separate from, and logically posterior to, the definition of mixed-method research.

multiple epistemologies. (Geertz 1973; Harding 1987; Greene and Caracelli 1997; Lin 1998; Yanow 2003; Dessler 2003) It is, of course, possible that some epistemologies cannot be served by some methods: that survey research can not be used in the support of interpretivist projects (Stoker 2003), or that a regression analysis of the conditions under which winks and twitches occur adds little or nothing to an ethnographic understanding of their meaning (See Geertz 1973:6-7, 12; King, Keohane, and Verba1994:38-40 for a rejoinder). But the disagreement over what various research paradigms might be, and how to define them if they exist, creates the need to examine method and paradigms separately. In explaining how and when methods might be combined, we leave open the possibility that their pairing also leads to the combination of research paradigms, without prejudging whether this will occur.

Deciding to Use Mixed Methods: A Decision Tree

The use of multiple methods has the potential to increase the validity of the theoretical propositions we offer and to extend their depth and richness. Yet it is but one strategy to achieve these ends. The strengths of one method can compensate for the weaknesses of the other, but multiplying studies is no guarantee of targeting the original method's shortcomings. Instead, it may be more effective to compensate for the weaknesses of a method by using the same method on new observations, or new measurements of the original observations. (King, Keohane, and Verba 1994: 217-218) Similarly, depth of knowledge can often be better achieved by more intense, careful, and thorough use of one method than by the superficial application of many. Research design, like most other activities, is a benefit-cost exercise: the advantages of versatility can be offset by the time and research effort needed to mix methods well.

Single-method research, in other words, may sometimes be more appropriate than mixed method research. Collier, Brady, and Seawright, for instance, summarize several schools of thought that characterize qualitative studies as providing more detailed knowledge, lower levels of measurement (nominal vs. ordinal vs. interval), and fewer simplifying assumptions. Quantitative studies, by contrast, "provide explicit, carefully formulated criteria" that can lead to the systemization of knowledge (2004:244-249). By implication, therefore, research areas about which little is known, or where there is a great deal of data but little agreement as to its organization, qualitative and quantitative methods, respectively, are often more appropriate.

Single-method research may also be more appropriate for reasons of feasibility. For questions on which there is very little previous research and thus little established theory, the in-depth use of any one method – qualitative or quantitative -- will usually generate enough description and inductive theory to make a substantial contribution to the literature. The cost of gathering more than one type of data is also an important consideration. Collecting new data, or exploring the possibilities of data which were collected for some other purpose, is a full-time task, whether the collection method is a new national survey or extensive ethnographic fieldwork. The effort to create new theory from new data is monumental in itself, and is seldom improved by repeating the practice with additional data analysis techniques.

This is not to say that multiple methods are not useful for generating new theory; as we argue below, there are certain theoretical tasks for which mixed methods are especially appropriate. But those tasks are best undertaken when there is an extant body of theory and literature to allow for the targeted use of mixed methodologies. This principle suggests that it may also be appropriate to discontinue one part of a mixed method project when the use of the first method is generating surprising new insights, or when the data being explored are richer

than first assumed. The measure of a good research project should always be the findings produced by the project, not the complexity of the design.

As a single method approach reaches fruition, however, it is often appropriate to consider whether there are "observable implications" (King, Keohane, and Verba 1994: 223-228) of one's theory that could be best explored through additional data. We call this limited use of mixed methods *supplementary validation:* gathering evidence from different sources that is consistent with the central argument. Arguments crafted from even exemplary uses of a single method can often be strengthened by bringing in confirmatory data: specific illustrations of a general case, trends over time that are consistent with the archival record, public opinion polling suggesting the generalizability of a particular worldview. In contrast to convergent validation, which we describe below, one generates the data for supplementary validation through deduction from the results of a different method: the findings of one method suggest that some other statement should also be true, and the researcher then goes looking for additional data to confirm or deny that possibility. Thus the use of such data usually does not rise to the level of an independent test of the theory; it merely adds to or detracts from the validity of the primary method.

A more extensive use of mixed methodologies is most suited for research questions in which the state of knowledge is intermediate: when some theoretical expectations have been established, and/or when the limitations of the data that have been used to explore a question have become more obvious. In the first case, when some theoretical expectations exist but the conceptual framework or the appropriate propositions need to be developed, mixed methodologies can help to extend our understanding of those theories. For instance, based on previous literature, it might be possible to run a large-n study and then use case studies to discover the mechanism behind its results; it might also be possible to take an established but

underspecified theory, and break it down into sets of narrower propositions that can then be explored through either qualitative or quantitative data. In the second case, when data limitations have stymied research, an unconventional data analysis method can generate new ideas for measurement or uncover new indicators. For instance, known contradictions between theoretical predictions, or theories that make predictions at different levels of aggregation, can lead to a mixed-method research design that resolves controversies by testing each aspect of a theory with appropriate information.

These different research design choices are summarized in Figure 1.

INSERT FIGURE 1 HERE

The most important consideration in distinguishing between the use of single- and multiple-method designs is the state of knowledge and theory about the question at hand. But as the decision tree shows, after deciding that mixed methods are appropriate, there are still different purposes to which mixed methodologies can be directed. Mixed methods can be used for additional theory generation -- by further increasing knowledge of the important concepts embedded in a theory, by developing specific propositions to flesh out a theory, or both. Mixed methods can also be used for convergent validation, also known as triangulation, where the validity of a theory is established through multiple independent tests. We describe each of these potential uses below.

Supplementary Validation: Multiple Sources for Data and Analysis

Expanding studies that are primarily qualitative or quantitative to include other types of methods and data is an important use of mixed methods, one that has largely gone unexamined by writers in this area. This neglect is understandable given the fact that this type of expansion seldom includes both the collection of qualitative and quantitative primary sources, and the original use of qualitative and quantitative analyses. More commonly, secondary sources will be used to provide examples of one method while primary sources are used for another.[4] For instance, the author may perform an original analysis of roll call votes in the U.S. Congress but rely on other people's studies to "tell a story" of the coalitions on one bill. In other studies, primary data of both sorts may be collected – for instance, government data on crime rates and journalistic accounts of crime victims – but one type of data will be analyzed while the other will simply be reported. For purists, this is not a mixed-method approach; it merely reflects the collection of many different types of data and/or the borrowing of other researchers' analyses.

But the lack of complexity is also a virtue. In contrast to convergent studies, which must be designed into a project at its conception, data collection and analysis for supplementary validation are flexible. It can follow the discovery of initial findings; help the researcher to choose between alternative readings of her data; and enrich both the researcher and the reader's understanding of the situation being studied. As long as researchers take care to contextualize and critically assess the additional data or studies they present, supplementary validation through the use of mixed methods can be a valuable asset to single-method research.

---

[4] "Primary source" and "secondary source" are familiar terms to most social scientists, but a definition may be in order. Primary sources are original documentation of events; contemporary records, documents, or objects; and data collected by the researcher. Secondary sources, by contrast, are collected, published, or compiled by other researchers. A political scientist using an eyewitness account of a coup is using a primary source; a political scientist using a historian's account of that coup is using a secondary source.

Margaret E. Keck's study, *The Workers' Party and Democratization in Brazil* (1992), provides an example of how quantitative work can supplement a primarily qualitative study. Keck traces the evolution of the Workers' Party in Brazil, focusing on its relationship with the labor movement in São Paulo. She argues that the Party was a novel development in Brazil because it set out to express the interests of workers and the poor, because it sought to be democratic internally, and because it wanted to represent and be accountable to its members. She also argues that the Party developed in response to Brazil's political transition – that understanding Brazil's political transition illuminates the actions and evolution of the Party. Her study uses primary and secondary archival materials, including party documents and published journals. But to bolster her arguments about the Workers' Party's constituents and success over time, she also brings in elections results and statistics by area, demographic information, and industry statistics,

John Mueller's book, *Policy and Opinion in the Gulf War* (1994) presents the opposite case: qualitative work supplementing a quantitative analysis. Mueller examines the long-term effect of the Gulf War on public opinion and argues that it had little salience after hostilities ended. The argument depends primarily on the presentation and analysis of polling data from multiple sources. But in order to link the polling data to events and to political attempts to capitalize on them, Mueller also includes both news stories and some secondary sources on policymaking during this time period. While the trends in the polls themselves establish Mueller's argument, the accounts of policy development help the reader match the trends to their contemporaneous contexts.

It is important to understand the difference between these uses of mixed methods and convergent validation. The classic definition of convergent validation requires that one get the

same findings from separate studies, preferably through methods that correct for each other's weaknesses. For instance, had Keck been able to survey Workers' Party members as to whether they felt the party was internally accountable, she could have corrected for the possibility that party documents portrayed only a partial and biased account of its commitment to accountability. But if these uses of mixed methods do not quite rise to the level of convergent validation, they do provide "supplementary validation": evidence from different sources that is consistent with the central argument. The great advantage of supplementary validation is that it is an easy yet powerful approach to enriching research. Social phenomena exist in a world of both numbers and words, yet the blinders of one's chosen data set can prevent researchers from taking advantage of this fact. Looking beyond one's primary method and data sources need not require new data collection or additional research design; the researcher can rely on accessible public sources or on the work of previous scholars. The relatively small effort expended in locating data or studies that would support one's work can pay off in a substantial gain of perspective on and confidence in one's findings.

The limitation of supplementary validation, however, is that it is sometimes merely illustrative. Journalists often adopt this approach: in order to make vivid the difference between the 40% of people who support the president's foreign policy and the 42% who do not, they will quote two seemingly average citizens, one a supporter and one a critic. Numbers can be illustrative as well: statistics about a city's unemployment or a neighborhood's poverty rates can add color to an oral history interview with a grandmother who has struggled to raise children and grandchildren in impoverished circumstances. The problem is not so much that the quotations or statistics might have been deliberately selected to make a point, although critics are often quick to levy this charge. Rather, the problem is that many scholars who mix methods for the purpose

of supplementary validation neglect to give the reader a context for the information they provide. A quotation from a bureaucrat that confirms agency data about an increase in red tape over time is more convincing when the reader is told about the bureaucrat's position, portfolio, reputation, and time in office. Similarly, a set of crime statistics from a poor neighborhood sketches a more convincing picture of abnormal levels of crime when it is presented in the context of crime in the rest of the city.

To be effective, therefore, a project that uses supplementary validation should adhere to certain guidelines. First, the additional data provided for supplementary validation should be grounded in a discussion of its context, and the context should be integrated into the explanation or "proof" that it provides. Second, the researcher should indicate whether the supplementary data are controversial in and of themselves. If they admit of multiple explanations, it may be that supplementary validation is not an effective strategy. Instead, triangulation – in which alternative explanations for both kinds of data can be thoroughly explored – may be a better approach.

Multiplying Data and Offsetting Weakness: Triangulation

The use of both qualitative and quantitative methodologies in research projects and publications has long been understood as a way of compensating for the shortcomings of any one method. Thus, Donald Campbell and D.W. Fiske argued that "multiple operationalism" could "ensure that the variance reflected that of the trait and not of the method." (Jick 1979: 602) A famous 1979 article by Todd Jick elaborated on this concept, which he called "triangulation": "The effectiveness of triangulation rests on the premise that the weaknesses in each single method will be compensated by the counter-balancing strengths of another ." (604) Sidney

Tarrow points out that triangulation is also useful when available data have shortcomings: "Triangulation is particularly appropriate in cases in which quantitative data are partial and qualitative investigation is obstructed by political conditions." (Tarrow 2004: 178) The convergence of findings across different techniques raises confidence in their validity.

Two characteristics distinguish a project that uses multiple methods for the purposes of triangulation: (a) the emphasis on testing the same hypothesis multiple times, using different methods in each iteration; and (b) the focus on aggregating knowledge, rather than on discovering new relationships. That is, each component of a triangulation project independently illustrates the central argument of the research or independently supports the hypothesis. (Caracelli and Greene 1997:22) The relationship between the component studies is one of *joint reinforcement*; each component can stand alone, although they make a stronger argument in combination.

As this explanation suggests, one need not combine qualitative and quantitative methods in order to triangulate. Kathleen Blee's *Women of the Klan* (1991) argues that women who joined the Ku Klux Klan in the first half of the twentieth century did so not out of alienation from the institutions of their day, but as part of a more expansive profile of community involvement: joining the Klan, in other words, was a mainstream, even progressive activity, no different than joining one's local church or women's club. Blee's data are entirely qualitative, but she practices triangulation by looking closely at the careers of leaders of the Women of the Ku Klux Klan (WKKK), by examining affiliations of regular members by looking at newspaper obituaries, and by interviewing former Klanswomen, their contemporaries, and their victims (Blee 1991:4-5). She asked the same questions of each type of data: what were the affiliations of WKKK members, how did they compare to non-WKKK members, and how did those

multiple affiliations affect each other? Collecting different types of data can correct for the biases implicit in each source. For instance, data from Klan sources, both individual and organizational, might over-report mainstream involvement in an attempt to hide extremism under a veneer of normality; data from non-Klan sources, even sources biased against the Klan, can help to put information from Klan sources into context. Combining different types of qualitative methods can also target weaknesses in inference inherent to each method. For instance, interviews of former Klanswomen (especially when those women are in their 70s and 80s) necessarily suffer from failures of memory. Examination of contemporaneous archival records can correct for this problem.

Moving from triangulation within the family of qualitative or quantitative methods to triangulation *across* methods, however, can increase the value of this practice. Bennett and Braumoeller argue that "the outliers in any particular piece of research could be . . . the product of stochastic processes, the result of measurement error, the consequence of a variable that affects the outcome of the outlier case and also those of other cases in the sample, or the consequence of a variable that is unique . . . .Combining different methods provides more powerful means than any one method used alone for discerning which of these holds true." (Bennett and Braumoeller 2004:7-8). The large-N research that is usually identified with quantitative methods (but see Collier, Brady, and Seawright 2004: 245-6) can provide leverage for deciding whether Klanswomen who proudly wear their racism when their robes come off are outliers, or whether it is Blee's Indiana communities that were atypical. Similarly, the small-N traditions of qualitative work may be able to find similarities in what at first might seem like examples of communities that handled the Klan quite differently.

Mixed-method research that pursues triangulation should, therefore, follow a set of common guidelines. First, this is research that starts from a clear set of expectations: a hypothesis or set of hypotheses that are grounded in theory and/or past empirical work, that are clear enough to be respecified in several different ways. Research that develops knowledge and theoretical expectations is very valuable and can be pursued with mixed methods; as we explain in the next section, however, it is *not* triangulation. Second, this research must ensure that the results derived from different tests of the hypotheses are independent. Because qualitative and quantitative methods usually use different data sources, this guideline can seem to be trivial; but since quantitative data are often produced by abstracting from qualitative data, independence can not be taken for granted. Third, researchers must determine whether their choice of methods is primarily driven by the data, or also driven by the possibility of offsetting the weaknesses of one method with another. In the latter case, the researcher should be transparent about the weaknesses she is correcting for, and the nature of the correction. It is not enough to say, for instance, that surveys are representative but focus groups are not: if a set of focus groups and a survey are both conducted in Texas, neither is likely to be representative of the rest of the United States. By contrast, an explanation of the weakness of a survey (low coverage of groups that are hard to reach by and reluctant to speak on the telephone) and the ways in which the selection strategy of the focus groups correct for that weakness (by finding under-represented groups and increasing their comfort level through a small group format) makes clear the distinctive contribution that triangulation brings to research.

Creating Concepts and Exploring Propositions: Theory Generation

The task of triangulation is clearly theory-testing, which by definition requires the existence of theory. But as many scholars have forcefully argued, the work of social science is more often the iterated refinement of concepts and propositions, and not the evaluation of relationships whose meaning is clear to all (Adcock and Collier 2001; Gerring 2001; Glaser and Strauss 1967; McKeown 2004; Ragin 2004). Thus if triangulation is the contribution most commonly claimed for mixed methods, theory generation is the contribution that can bear the most fruit. As Jick argues in his discussion of triangulation, "divergence can often turn out to be an opportunity for enriching the explanation . . . In seeking explanations for divergent results, the researcher may uncover unexpected results or unseen contextual factors." (607-8)

As can be surmised from this statement, many projects that are phrased in the language of triangulation actually resemble theory generation instead. In these projects, the findings produced by one type of research method set the agenda for use of a different type of method. Thus, in comparison to triangulation, the studies produced by different methods are *integrated* and not independent (Caracelli and Greene 1997: 23). The classic example is what Tarrow calls "quantitative data as point of departure for qualitative research: a quantitative data set serves as the starting point for framing a study that is primarily qualitative." ; he also suggests other types of studies in this vein, such as "process tracing" and the "focus on tipping points." (Tarrow 2004:174). Bennett and Braumoeller (2004), building on Lijphart (1971) and Eckstein (1975), distinguish between "disciplined-configurative" studies, in which the major propositions produced by one method structures an examination by the other, and "heuristic/theory-generating" studies, in which one method identifies outliers that the other method investigates; both types of studies, they argue, serve a theory-building purpose and benefit from iteration (18-20).

As with triangulation, the methods that are mixed in the service of theory generation need not be both qualitative and quantitative. The use of factor analysis to clarify families of concepts, followed by regression analysis to discover relationships between them, is an example of the mixing of different quantitative methods in a project. Most studies of this type, however, rely on the proposition that qualitative small-N or single-case studies and large-N quantitative work have distinctive, complementary strengths. This perspective holds that small-N studies are better for establishing causal mechanisms or for process tracing: that is, for exploring the many interactions that cause an event to occur or lead to particular outcomes. The generalizability of these mechanisms, however, is best established through large-n research, which can establish whether patterns hold across space and/or time. The ordering of the small- and large-N projects is not crucial: at times the qualitative project is presented as discovering mechanisms that can be tested with statistical work, while at other times the quantitative work establishes patterns that must then be explained by an in-depth study. In some studies there may even be iterative movement back and forth, as in research where there are multiple factors to be explored in paired small- and large-N investigations. What distinguishes this type of work is the assumption that each type of research is, in and of itself, *partial*:[5] it is the combination of the various projects that proves the argument.

William M. Downs' book, *Coalition Government Sub-national Style: Multiparty Politics in Europe's Regional Parliaments* (1998), provides an example. Downs seeks to explain and predict how politicians choose to form state, provincial, and regional coalition governments, with a particular focus on Belgium, France, and Germany. His hypotheses, formulated through an

[5] Here, "partial" does not mean "incomplete": chapters from books like these are often first published as articles. Rather, it means that the full argument is understood to be dependent upon both types of research.

extensive discussion of both theoretical and empirical literature, are first tested through a survey of politicians at the subnational level and then fleshed out and made more detailed through interviews and archival data on each country's election experience. Throughout this study, Downs introduces new methodologies to converge on a theory that explains those patterns revealed in the first phase of his analysis. Each method serves to refine and clarify those empirical findings he has previously established.

Downs uses historical analysis to establish the frequency and intensity with which governing coalitions are formed post-election in Germany, Belgium, and France. Compiling election statistics from government sources and medial coverage of the coalition formation process, Downs finds evidence that coalition precedents stem from the sub-national level to the national level. That is, at initial glance, it seems that national leaders take their cues from sub-national politics when forming governing coalitions. Yet is it unclear if this national reflection of sub-national coalitions is conscious or coincidental.

Downs uses an extensive closed-ended mail-in questionnaire to understand the nature of these interactions and to refine the hypotheses of the study. Analyzing survey responses from elected officials at both the national and sub-national levels, he is able to gauge the attitudes, behaviors, and motivations of these actors as well as the circumstances under which their attitudes and behaviors are discrepant from their motivations. In analyzing the survey data, Downs finds that there are many prevalent barriers that could prevent bottom-up coalition influence including entrenched party identities and the lack of effective organizational communication across the sub-national platform. Yet, he finds that national party intervention on regional issues facilitates the bottom-up flow of coalition patterns and that the weaker is party discipline, the more influential is the regional government in coalition formation.

In the final leg of this study, Downs interviews elected representatives and party leaders at both levels of government. These interviews provide him with the "contextual material essential for full interpretation … and ameliorate some of the inherent restrictiveness of closed-ended questionnaires (p. 84)." They reveal that regional officials are aware that their actions may have national-level influence and consider the potential impact in their decision-making processes. They consciously try to influence national level coalitions with their actions at the regional level.

Downs' study, of course, is but one of many examples of theory generation, and the capaciousness of this category makes general guidelines at once necessary but difficult to formulate. Mixed method projects of this kind, however, do share some requirements. First, the task of theory generation cannot begin without a statement of the problem: the concepts that need clarification, the propositions that lack precision, the relationship that puzzles. The necessity to generate theory in a particular area means, almost by definition, that existing hypotheses are obviously inadequate. Thus starting a project with the existing hypotheses may not be the best approach, but the reverse flaw - the wholesale reinvention of concepts and propositions – is also to be avoided. Second, explicit consideration of how one's data are biased (as all data inevitably are) is especially crucial in theory generation, since the answer greatly affects both the scope and the assumptions of any theory one generates. The danger of creating theory to fit one's data is especially acute, since the iterative process encourages repeated use of the same data set. This is not wrong – as Gerardo Munck points out, "it would be an important constraint on the accumulation of knowledge if analysts did not routinely revise their explanations of a set of cases and then test the new explanation – if need be, with the same set of data." (Munck 2004:119) but it does mean that any limits imposed by the data need to be

factored into one's theory-building. Third, since the iterative nature of this type of study is its greatest strength, the connections between the various parts of the project need to be fine-tuned and made explicit. A concept developed from ethnographical observations can be proxied so that econometric analysis can be used, but any findings from the analysis need to be read back into the ethnography as a check on the concept.

Applying Research Design to Evaluation

The simple classification system – supplementary validation, theory generation, and triangulation -- that we have presented above is one focused on the research purpose rather than the research design. Researchers who already know which of these purposes they plan to pursue, or whose primary interest is in a set of findings – for instance, in evaluating whether a program has achieved its intended results, in generating evidence to support or disprove a clear hypothesis, in describing different aspects of a puzzling phenomenon -- will get more ideas about appropriate research design by studying one or more of the many elaborate typologies that researchers in this area have compiled. (Teddlie and Tashakkori 2003) Our discussion, however, is especially well-suited to researchers who begin consideration of a project from a puzzle that they are want to solve and a place in the literature from which that puzzle arises. In this we follow Maxwell and Loomis (2003), who argue that as useful as mixed method typologies can be, understanding the "logic-in-use" of a study requires understanding "the different components of a research study (including purposes, conceptual framework, research questions, and validity strategies, in addition to "methods" in a strict sense) and the ways in which these components are integrated with, and mutually influence, each other." (242-243) Starting from the question of whether one's puzzle reflects the need for more conceptual

development, clearer propositions, or different measures for different kinds of data, our classification system directs the researcher to the tasks she needs to accomplish and the criteria for understanding whether she has accomplished them well.

This description is also true of another group of people concerned with mixed method projects – readers, whose main job is not to determine the appropriateness of the research design used but rather the contribution made by a particular mixed-method project. Readers and evaluators of mixed-method research also start from the puzzle, as presented by the author and bolstered by their own knowledge of the literature. From outside the project, they make judgments about the state of theory and knowledge on a particular issue; from outside the project, they also need to decide whether the researcher has an adequate, and hopefully interesting answer.

Yet this task is often more complex than it might at first seem. The evaluation of mixed methods research has two components: the validity of each of the component analyses, and the effectiveness of the combination of methods. On the first, the standard rules of validity for the qualitative and the quantitative components of a mixed methods analysis apply, making the evaluation of internal validity relatively straightforward. Even here, however, the standards of proof for supplementary validation and convergent validation should differ; in supplementary validation, as we have discussed, the secondary method provides additional data but not a full-fledged study, and thus analyses of validity should focus on the context for that data rather than on whether the procedures producing that data would be sufficient for an independent analysis. Even more daunting is to evaluate how different aspects of the study work together. Frequently, researchers are not explicit about the sequencing of their questions in theory generation, nor even about the common hypothesis tested in triangulation. Evaluators must then work inductively:

making judgments about the research task that the author is trying to accomplish, classifying the nature of the findings produced by each method, and determining how those findings relate to the original research task. An essential element in this process is to judge mixed-method projects by what they intend to accomplish: triangulation should be judged by different standards than theory generation, because the function of the methods and their integration will both be different.

Figure 2 explains this process.

INSERT FIGURE 2 HERE

Two considerations determine how and how well the methods in a mixed method project are combined: the nature of the argument, and the relationship between the methods. The evaluator's first task is to identify the argument of the book, teasing out the researcher's thesis statement, guiding questions, and propositions. The most important aspect of this task is to evaluate how the researcher characterizes the state of research on her question and explains the status of her findings. Does she believe that she is resolving a controversy or puzzle in the literature, or between the literature and empirical reality? Is she adding theoretical structure to a set of topics which have been amply described but not well-understood? Or does she argue that her findings extend the understanding of a phenomenon that heretofore went unnoticed or unstudied? The first characterization of her task would lead the evaluator to expect convergent or supplementary validation; the second and third, theory generation.[6]

[6] Of course, the evaluator may also have a substantive critique, if her understanding of the relevant literature does not match the researcher's.

In order to answer questions like these, the evaluator must next determine the separate function of each method in relation to the argument. A method may be used to (a) dispute a hypothesis or previous finding; (b) support a hypothesis; (c) provide an explanation; (d) describe processes; or (e) establish patterns. Once these functions are established, the evaluator can determine the relationship between the methods and match them to the forms of mixed methods analysis previously discussed: supplemental validation, triangulation, or theory generation. This match is determined by the extent to which each method influences the analysis and the degree to which the methods are integrated in the analysis.

Frequently, both quantitative and qualitative methods are used to support a single hypothesis. These methodological dynamics are indicative of triangulation and supplementary validation. To determine which framework should be used in the evaluation of the process, consider the extent to which the second method influences the analysis. If one method is dominant, then the project should be evaluated as a work of supplementary validation. If each method offers an independent test, or one method is used to support the hypothesis while the second disputes alternatives, the project uses convergent validation.

By contrast, theory generation is indicated by one of three other combinations of functions. The first of these is when one method is used to describe a process while another is used to provide explanations for it. This type of work usually results in the formulation of propositions. In the second combination, one method is used to show the stable/predictable conjunction of two or more phenomena, while the second method explains why these relationships are related. The third combination is pattern establishment coupled with process tracing or description. In this approach, the relationship in question usually occurs in the form of a sequence of events, rather than in the conjunction of phenomena. In these two latter types of

research, the researcher essentially adds greater sophistication to the concepts and relationships under investigation; the result is a more detailed, better specified theory.

Distinguishing between these different uses of mixed methods helps the evaluator to assess whether the combination of methods is appropriate for the question(s) posed. A researcher who claims that she is resolving a controversy should test the same hypothesis with both methods of data analysis. If the hypotheses are different – perhaps because it is not possible or desirable to measure variables in the same way with qualitative and quantitative data – she should make the different measures explicit, and defend them. To the extent possible, the data used to test the hypotheses should also be independent; the case studies that are used to test one hypothesis may also be represented in a large-n data set, but the sources for the first should be demonstrably different from the sources used to compile the second.

Neither of these conditions, however, needs hold true for theory generation. The iterative feedback of quantitative and qualitative analysis is instead the important trademark: the evaluator should consider whether the explanations, processes, or patterns discovered by the research are integrated. The results of one study should lead directly into the design of the second, as when a pattern of wars under certain conditions leads to a case study of one of those conditions, an exploration of a counterfactual, and an argument about its impact. An alternative possibility is when the results of one method lead to several possibilities, each of which is then separately explored by other methods: for instance, when the process tracing of a particular war leads to a set of propositions, some of which can be explored through a large-n study and others which are developed through different case studies. The data for theory generation also need not be independent: a common practice, for instance, is to code for systematic features in a set of interviews, and then to examine the transcripts of those interviews more closely to explain why

those coding choices were made and what they signify. By contrast, the methods used for supplementary and convergent validation need not be integrated. In this case, one might find a pattern of wars and their attendant conditions in large-n data, and then compare the initiation of two wars, one in which those conditions obtained and one in which they did not. While both studies together clearly add explanatory value, the conclusion would stand even if one of the analysis techniques and its findings (the secondary technique, in the case of supplementary validation) were dropped from the research.

Conclusion

An important strand of the debate about the uses of qualitative and quantitative research lies in whether the two are governed by one research logic, or instead represent different research paradigms. King, Keohane, and Verba, for instance, argue that there is one "inherent logic underlying all social scientific research" (King, Keohane and Verba 1994: 230). In an essay comparing Karl Marx's *Eighteenth Brumaire of Louis Napoleon* and Emile Durkheim's *Suicide,* David Laitin agrees: "All social scientists are interpretivists . . . a reading of these two works demonstrates the futility of carving out a separate methodological realm for qualitativists called interpretivism." (Laitin 2003:6-7 By contrast, others argue that identifying different research epistemologies broadens the scope of what social science contributes. McKeown (2004) argues that the Popperian notion of "testing" in the statistical mode is usefully supplemented by a "folk Bayesian", "interactive processing" approach which moves "back and forth between theory and data, rather than taking a single pass through the data." (159). Lin (1998) identifies the study of causal mechanisms as particularly suited to "interpretive" research and the study of causal relationships as particularly suited to "positivist" research , and advocates mixing the two to

allow for the study of both.[7] Brady, Collier, and Seawright, while agreeing that "quantitative and qualitative methods are founded on essentially similar epistemologies" (Brady, Collier, and Seawright 2004: 7), argue that "data-set observations" and "causal-process observations" together increase causal leverage. (Collier, Brady, and Seawright 2004:259).

The classification of mixed methods research that we have presented here maps, in some ways, onto these distinctions. Theory generation, in particular, can be understood in terms of combining the strengths of different epistemologies to create richer theories; those who accept only one logic of inference can also accept one version of this research task: the large-n testing of hypotheses developed from small-n work. Convergent validation, by contrast, fits less comfortably. Positivists might argue that triangulation – testing one hypothesis in multiple ways – is exactly what social science should do. On the other hand, they might also be concerned about the validity of a small-n test. By contrast, the very notion of validation is a hotly debated concept among interpretivists. Some hold that the research process can be valid or invalid, but that it is impossible to judge the truth-claim of a finding. Others would argue that the truth-claims of interpretive work are established not by a probabilistic account but rather by the recognition of those who are the subject of that work. (Kirk and Miller 1986; Altheide and Johnson 1994; Adcock and Collier 2001) Thus an interpretivist would be unlikely to accept the premise that interpretive work could be validated through positivist work, or vice versa.

This being said, people who believe that mixing methods entails mixing paradigms may be able to see a purpose for triangulation under a broader understanding of convergence. One may see "convergence" as proving not whether different studies are valid, but instead that the phenomena they study is consistent and common. For instance, Paul Brass' *Theft of an Idol*

[7] Importantly, both McKeown and Lin reject the idea that quantitative and qualitative work come

(1997) analyzes statistical data to establish that in three highly publicized cases, actual patterns of caste-related violence in India differ dramatically from the way that they are presented by local, state, and national political actors. Comparing the statistical work to media accounts and politicians' speeches, he finds that politicians and the media fail to acknowledge that the violence experienced by one caste could have been perpetrated by members of the same caste; they also take acts of violence against some castes out of context and downplay violence against other castes. He interprets this evidence to mean that political actors deliberately use the occurrence of violent acts in India to support their various political ideologies and to justify state authority. Brass then does an interpretive study in five villages where these acts of violence have occurred, three for which he has statistical evidence and two for which he does not. In these villages, he finds, incidents of caste violence have local meanings that are not consistent with inter-caste hostility; they are understood as ordinary crimes, not hate crimes. The finding that the local meanings of caste violence differ from the politicians' interpretations does not converge with his study of politicians; it is not an identical finding, which is what triangulation demands. But it is consistent with his analysis of the politicians' actions , and so helps the reader understand how these different findings can fit together.

Proponents of mixed paradigm research, then, should not find these categories of mixed analysis difficult to accept. The value of these distinctions, however, transcends beliefs about the possibility or the value of mixing different research paradigms. Rather, it lies in the traction they give researchers when planning their studies and considering how to present the results. Instead of focusing on the epistemology of different types of methods, the classification presented here focuses on the type of research task – supplementary validation, convergent

---

from different epistemological approaches; both distinguish between epistemology and method.

validation, or theory generation – that the researcher wants to accomplish. This encourages clarity on the researcher's part about the argument that she is hoping to make, rather than fuzzy assertions that, by definition, a mixed method project is better than a single method study. It also clarifies, for the researcher, what she is not planning to accomplish – a knowledge that can be very important, when faced with one's own or one's critics' instinct to do and demand more. At the same time, readers and reviewers are also likely to benefit from these distinctions. When they know what a researcher is trying to accomplish, they can more easily judge whether the task has been completed well or appropriately. While different interlocutors are unlikely to share the same standards, the ability to base their judgments on a shared identification of the research task is a step forward in creating a disciplinary understanding of the value of mixed methods.

The discipline of political science has gained a great deal from the efforts of quantitative scholars to discuss the benefits and the drawbacks of their methods, and from the growing effort among qualitative scholars to do the same. As the discipline becomes more self-conscious and more knowledgeable about how and why scholars might choose to structure their studies using qualitative or quantitative techniques, the decision to use both will certainly become more frequent. But this decision should never be made out of a belief that more is necessarily better. The choice to combine methods in a study is a choice to expand one's understanding of useful evidence; to validate one's research through different studies; or to explore multiple aspects of a particular phenomenon in an attempt to develop richer concepts and clearer propositions. The question every researcher must ask is whether one of these goals should be the purpose of her study; and if so, whether mixing methodologies is the best way to achieve it.

Works Cited

Adcock, Robert and David Collier. 2001. "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research, " *American Political Science Review* 95 (3):529-546 .September.

Althiede, David L. and John M. Johnson. 1994. "Criteria for Assessing Interpretive Validity in Qualitative Research." In Denzin, Norman K. and Yvonna S. Lincoln, eds., *Handbook of Qualitative Research.* Thousand Oaks, CA: Sage Publications.

Bennett, Andrew and Bear Braumoeller. 2004. "Where the Model Frequently Meets the Road: Combining Statistical, Formal, and Case Study Methods," draft paper.

Blee, Kathleen. 1991. *Women of the Klan: Racism and Gender in the 1920s.* Berkeley: University of California Press.

Campbell, Angus, Philip E. Converse, Warren E. Miller, and Donald E. Stokes. 1960. *The American Voter: Unabridged Edition.* Chicago: University of Chicago Press.

Caracelli, Valerie J. and Jennifer C. Greene. 1997. "Crafting Mixed-Method Evaluation Designs." *New Directions for Evaluation* 74:19-32. Summer.

Collier, David, H. E. Brady, and Jason Seawright. 2004. "Sources of Leverage in Causal Inference: Toward an Alternative View of Methodology" in Henry E. Brady and David Collier, eds., *Rethinking Social Inquiry: Diverse Tools, Shared Standards.* Lanham, Md.: Rowman & Littlefield.

Dessler, David. 2003. "The Positivist-Interpretivist Controversy." *Qualitative Methods.* 1(2):21-24. Fall.

Downs, William M. 1998. *Coalition Government Sub-national Style: Multiparty Politics in Europe's Regional Parliaments.* Columbus: Ohio State University Press.

Eckstein, Alexander. 1975. *China's Economic Development: The Interplay of Scarcity and Ideology.* Ann Arbor: University of Michigan Press.

Geertz, Clifford. 1973. *The Interpretation of Cultures.* New York: Basic Books

Gerring, John. 2001. *Social Science Methodology: A Criterial Framework.* New York: Cambridge University Press.

Glaser, Barney and Anselm Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research.* Aldine de Gruyter.

Harding, Sandra. 1987. "Is There a Feminist Method?" In Harding, Sandra, ed. *Feminism and Methodology.* Milton Keynes: Open University Press.

Jick, Todd. 1979. "Mixing Qualitative and Quantitative Methods: Triangulation in Action." *Administrative Science Quarterly* 24(4):602-611. December.

Keck, Margaret. 1992. *The Workers' Party and Democratization in Brazil.* New Haven, CT: Yale University Press.

King, Gary, Robert Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research.* Princeton: Princeton University Press.

Kirk and Miller. 1986. *Reliability and Validity in Qualitative Research.* Newbury Park, CA: Sage Publications.

Laitin, David. 2003. "Interpretation." *Qualitative Methods.* 1(2):6-9. Fall.

Lijphart, Arend. 1971. *Class Voting and Religions Voting in the European Democracies: A Preliminary Report.* Glasgow: University of Strathclyde.

Lin, Ann Chih. 1998. "Bridging Positivist and Interpretive Approaches to Qualitative Methods," *Policy Studies Journal* 26(1):162-180.

Maxwell, Joseph A. and Diane Loomis. 2002. "Mixed Methods Design: An Alternative Approach" in Abbas Tashakkori and Charles Teddlie, eds., *Handbook of Mixed Methods in Social and Behavioral Research.* Thousand Oaks: Sage Publications, Inc.

McKeown, Timothy J. 2004. "Case Studies and the Limits of the Quantitative Worldview" in Henry E. Brady and David Collier, eds., *Rethinking Social Inquiry: Diverse Tools, Shared Standards.* Lanham, Md.: Rowman & Littlefield.

Mueller, John. 1994. *Policy and Opinion in the Gulf War.* Chicago: University of Chicago Press.

Munck, Gerardo L. 2004. "Tools for Qualitative Research" in Henry E. Brady and David Collier, eds., *Rethinking Social Inquiry: Diverse Tools, Shared Standards.* Lanham, Md.: Rowman & Littlefield.

Stoker, Laura. 2003. "Is It Possible to Do Quantitative Survey Research in an Interpretive Way?" *Qualitative Methods.* 1(2):13-16. Fall.

Tarrow, Sidney. 2004. "Bridging the Quantitative – Qualitative Divide" in Henry E. Brady and David Collier, eds., *Rethinking Social Inquiry: Diverse Tools, Shared Standards.* Lanham, Md.: Rowman & Littlefield.

Tashakkori, Abbas and Charles Teddlie. 2003. "The Past and the Future of Mixed Methods Research: From "Methodological Triangulation" to "Mixed Model Designs"" in Abbas Tashakkori and Charles Teddlie, eds.,*Handbook of Mixed Methods in Social and Behavioral Research.* Thousand Oaks: Sage Publications, Inc.

Yanow, Dvora. 2003. "Interpretive Empirical Political Science: What Makes This Not a Subfield of Qualitative Methods." *Qualitative Methods.* 1(2):9-13. Fall.

Ragin, Charles. 2004. "Turning the Tables: how Case-Oriented Research Challenges Variable-Oriented Research" in Henry E. Brady and David Collier, eds., *Rethinking Social Inquiry: Diverse Tools, Shared Standards.* Lanham, Md.: Rowman & Littlefield.

Figure 1: Developing Mixed Methods Research

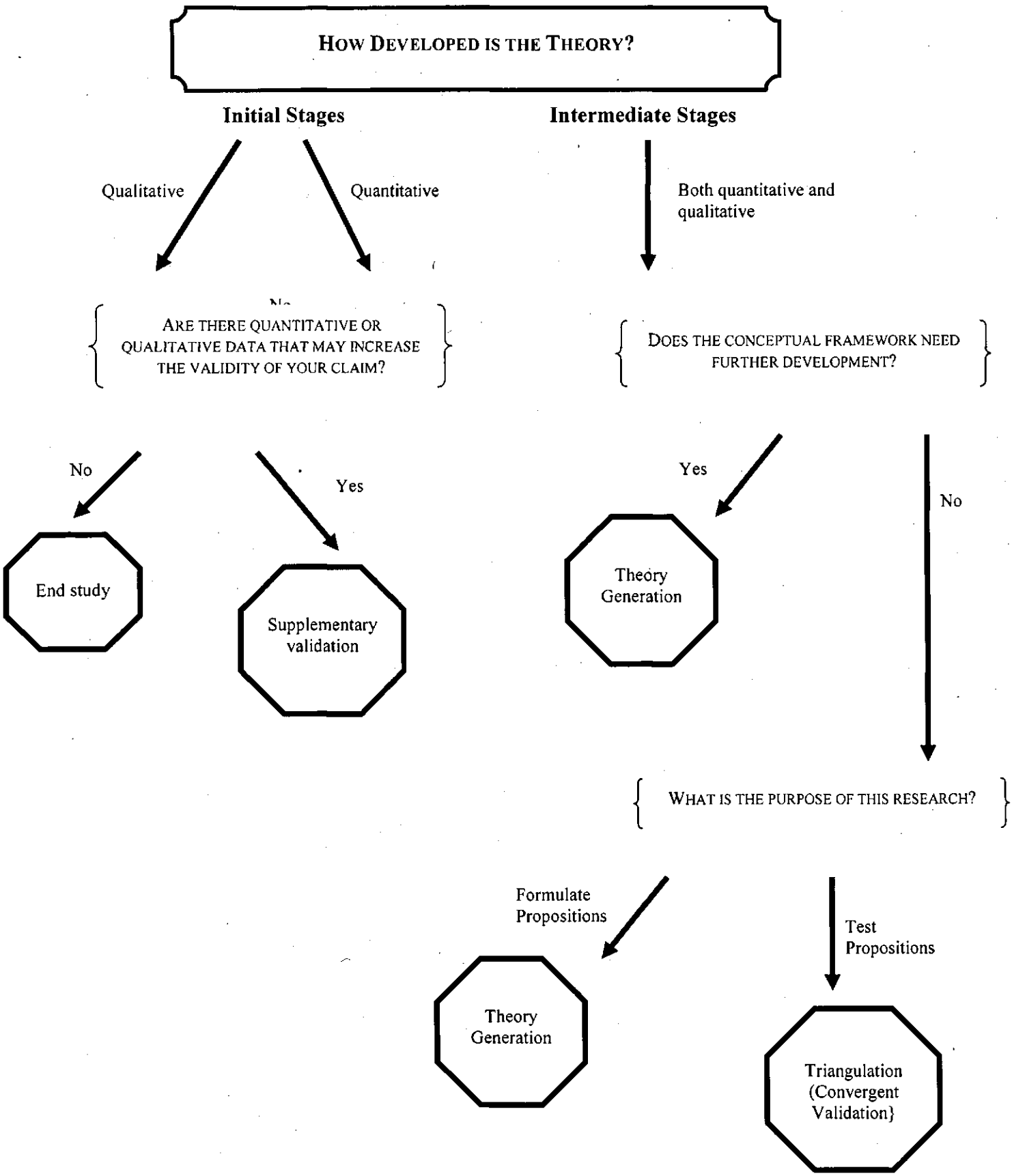
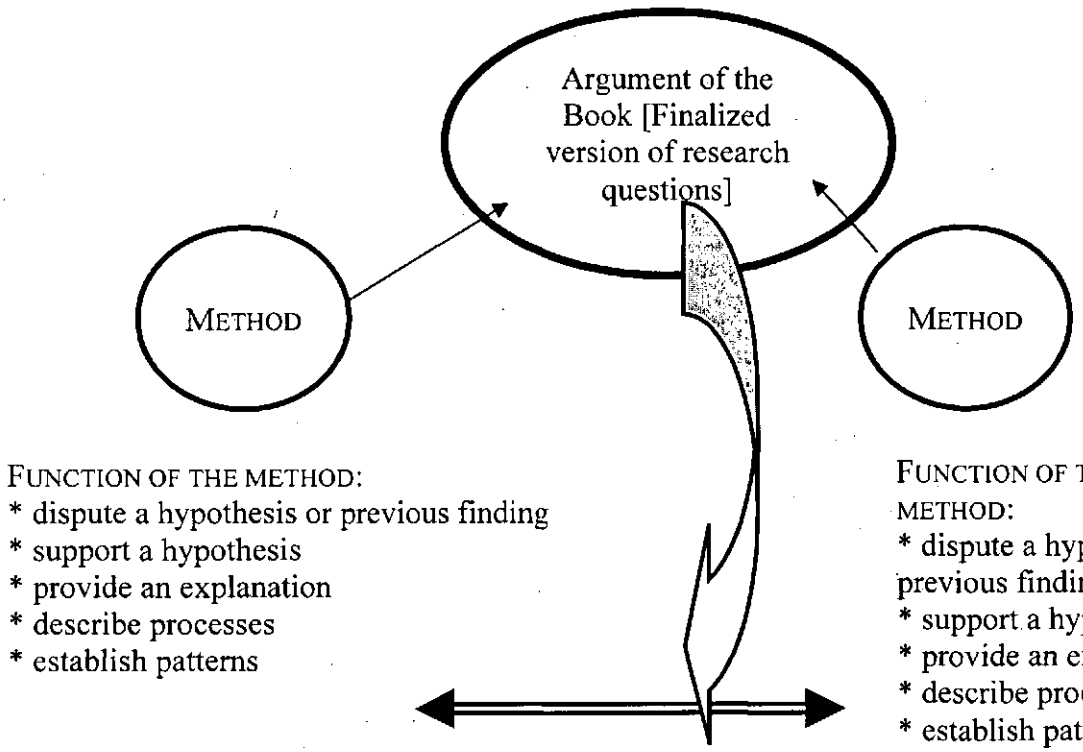


Figure 2: Evaluating Mixed Methods Research



SUPPLEMENTARY VALIDATION → • Both methods support a single hypothesis

TRIANGULATION → • Low level of methodological integration

TRIANGULATION → • One method supports a hypothesis and the second disputes alternatives

- Pattern establishment coupled with provision of explanations
- Pattern establishment coupled with process description
- Process description coupled with provision of explanations
- High level of methodological integration

→ THEORY GENERATION