Mixed Method Research: Fundamental Issues of Design, Validity, and Reliability in Construction Research

Deborah A. Abowitz[1] and T. Michael Toole[2]

Abstract: The fact that people play key roles in nearly all aspects of construction suggests that effective construction research requires proper application of social science research methods. This is particularly true for researchers studying topics that involve human actions or behavior in construction processes, such as leadership, innovation, and planning. In social science research, no single method of data collection (survey, experiment, participant observation, or unobtrusive research) is ideal. Each method has inherent strengths and weaknesses. Careful attention to the methodological ABCs of the design process, as discussed here, can enhance the validity and reliability of a given study. Combining quantitative and qualitative approaches in research design and data collection, however, should be considered whenever possible. Such mixed-methods research is more expensive than a single method approach, in terms of time, money, and energy, but improves the validity and reliability of the resulting data and strengthens causal inferences by providing the opportunity to observe data convergence or divergence in hypothesis testing.

DOI: 10.1061/(ASCE)CO.1943-7862.0000026

CE Database subject headings: Research; Methodology; Measurement; Data analysis; Construction management.

Author keywords: Research methods; Mixed methods; Social science; Concept measurement; Data analysis.

Introduction

Construction is essentially a "social" process. In effect, construction can be considered to be the application *by people* of technology developed *by people* to achieve goals established *by people* involving the erection or retrofitting of infrastructure and buildings. The fact that people play key roles in nearly all aspects of the construction process suggests that in order to understand the human or social factors, effective construction research requires the proper application of social science research methods. Even when the concepts in our research (such as alternative delivery methods) do not ostensibly involve individual human traits or behavior per se, social factors can complicate the research process and jeopardize the results. Whether these factors derive from specific actions of, or decisions by, researchers in the design stage or in the field, and/or from individual informants in organizational settings, the social nature of construction research, and the implications of it as such, need to be carefully considered.

The focus of this paper is on some of the hazards we face in the application of social science methods to construction research and on the utility of applying a mixed or multimethod approach to enhance the reliability and validity of our results. Using multiple methods, in particular mixing qualitative and quantitative techniques, on a given research project (such as surveys *and* participant observation, or interviews and archival data), allows us to balance the strengths and weaknesses of each approach and, if our theoretical premises are correct, converge on a common pattern or result. Combining multiple methods in this way, a form of triangulation takes place however, within a larger methodological context. Using multiple or mixed methods "affects not only measurement but all stages of research" (Brewer and Hunter 1989, p. 21). To utilize a mixed method approach properly, the key is to incorporate sound methodological principles at each stage of the design process.

The discussion here is intended for construction researchers whose particular methods intersect with those of social scientists (that is, those who use surveys, ethnographic research, and the like), whether as new doctoral students in the field or as construction faculty performing research in social science areas of construction for the first time. This paper represents a contribution to the literature not because it presents innovative construction research, but because it synthesizes the substantive and methodological expertise of a construction engineer with that of a social science methodologist to highlight and summarize critical social research problems in construction research and the utility of a mixed rather than singular methodological approach.

The importance of addressing social research problems in construction and the need for mixed research methods in particular can be supported by examining the papers published in the five issues of the *Journal of Construction Engineering and Management* (JCEM) between February and August 2008. Analyses of the entire manuscripts (not just the abstracts) indicated that social science constructs (defined shortly) were either the focus of the paper or were key parts of the background discussion for 22 of the 43 papers published, yet none of the papers used multiple methods. Additional relevant findings from the writers' analysis of the five issues of the JCEM will be mentioned later in this paper. Following a brief review of the existing literature on construction research methods, we review basic methodological prin-

[1]Professor, Sociology, Bucknell Univ., Lewisburg, PA 17837. E-mail: dabowitz@bucknell.edu

[2]Associate Professor, Civil and Environmental Engineering, Bucknell Univ., Lewisburg, PA 17837 (corresponding author). E-mail: mike.toole@bucknell.edu

Note. This manuscript was submitted on July 29, 2008; approved on December 29, 2008; published online on March 27, 2009. Discussion period open until June 1, 2010; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Construction Engineering and Management***, Vol. 136, No. 1, January 1, 2010. ©ASCE, ISSN 0733-9364/2010/1-108–116/$25.00.

J. Constr. Eng. Manage., 2010, 136(1): 108-116

Table 1. Four Types of Validity in Social Science Research (Cook and Campbell 1979)

Type of validity	Key issue
Construct validity	Do the indicators capture the expected relationships among the concepts being researched?
Statistical conclusiveness validity	Are the relationships between hypothesized independent and dependent variables statistically significant?
Internal validity	Has the research truly demonstrated a causal link between the hypothesized variables, or are there plausible alternative explanations for the statistical association between the independent and dependent variables?
External validity	Are the apparent relationships found within the sample's experimental subjects generalizable to the larger population assumed in the hypothesis?

ciples related to the application of a mixed methodological design.

Review of the Existing Literature on Construction Research Methods

The literature on construction research methods is rather sparse, which would seem to confirm the need for this special issue of the *JCEM*. Although the book is not well known in the United States, an excellent contribution is made by Fellows and Liu (2008), which focuses on the chronological processes that should be followed to ensure successful research. This book also focuses on the philosophical issues related to research methodologies. Three 1997 papers also discussed research methods at the broadest levels, debating the relative merits of theoretical versus empirical papers and qualitative versus statistical or quantitative research (Seymour et al. 1997; Raftery et al. 1997; Runeson 1997). Loosemore (1999) identified cultural differences, especially those involving communication, that make one category of research methods more appropriate than another. Walker (1997) discussed his doctoral research process as a case study on the challenges of obtaining data from thirty-three projects and analyzing them using linear regression. Similarly, El-Diraby and O'Connor (2004) used the collection and analysis of bridge construction data as a case study to summarize key methodological issues taken from Cook and Campbell (1979), which is one of the major works on experimental design in the social sciences [along with Campbell and Stanley (1963)] [for a thorough introductory level discussion of true experiments and the flaws associated with various preexperimental and quasi-experimental designs, see Babbie (2008) or Chambliss and Schutt (2006)]. Four types of validity concern Cook and Campbell in particular. They are briefly summarized in the Table 1.

In both the social sciences and construction research, a clear understanding of experimental design issues is critical (Fellows and Liu 2008). It is important not only to the quality (validity) of any experimental data collected but for a deeper understanding of basic research processes and problems. Careful study of experimental design texts, like Cook and Campbell (1979) can improve other research efforts—including surveys and ethnographic observation—by drawing increased attention to the methodological ABCs (basic principles) of research design. Ultimately, understanding issues of internal and external validity in experiments, as well as measurement (construct validity) and statistical significance (statistical conclusiveness validity), helps us understand the broader problems of causal inference, generalizability, measurement reliability and validity, and statistical inference in construction and other types of research. These problems in research design, and the application of multiple or mixed method approaches to address them, are discussed in the following sections.

Methodological ABCs

Although the emphasis in the examples discussed here is on the application of social science methods in construction research, these principles pertain whether we are studying social phenomena or aspects of the physical world. The lessons highlighted here draw from the first writer's extensive experience teaching social science research methods, conducting survey research and quantitative analyses, and writing peer reviews for journals, book publishers, and granting agencies both in the United States and abroad. They also draw from the second writer's experiences as a construction researcher, paper reviewer and JCEM specialty editor.

Surveys, questionnaires, experiments, ethnographic observation, and unobtrusive techniques are all valuable research tools, but each yields somewhat different perspectives on a research question. Each is subject to particular problems of validity and reliability; each has distinct limits of generalizability. Whether using one method to generate and collect data, or combining methods in a mixed-method approach, researchers need to be cognizant of the consequences of each of the methodological choices they make.

Proper Research Planning and Design Enables Successful Data Collection

As Babbie (2008, p. 122) notes, "Research design involves a set of decisions regarding what topic is to be studied among what population, with what research methods, for what purpose." Proper research planning and design therefore involve several distinct issues. For example, if the topic is leadership in the construction industry and its effect on firm performance, we have to determine which subjects to include from among firm leaders (the sampling technique and sample size) and the method(s) of data collection (an experiment, survey, observational study, and so on). Each of these decisions affects the quality and quantity of data collected.

With regard to the sampling plan, there are two initial questions: the type of sample or sampling procedure used and the sample size. Good science typically involves probability-based sampling (an example of which is a "random" sample) to minimize the chance of bias within the data; but probability based methods are rarely feasible in applied settings like ours and may not yield a reasonable response rate. This is why convenience samples (and snowball sampling) are so common in construction research. Approaching individuals and/or firms that we feel are likely to agree to participate enhances our response rates. Subgroups of interest in construction may include, for example, firms working primarily in the commercial market versus industrial/process, infrastructure or residential markets, or firms operating primarily in one geographic region versus another region. Experience has shown, however, that firms either will not agree to participate in the study or will not provide the needed documents unless they are truthfully told the research focus. Informed con-

J. Constr. Eng. Manage., 2010, 136(1): 108-116

sent documents, necessary for human subjects research, must therefore strike a careful balance—they need to contain enough information about the research methods and risks for subjects to make an informed choice about participating but do so without prejudicing the data collected.

Research based on nonprobability sampling techniques, such as that using convenience samples, can provide useful insights but it is limited with regard to the accuracy of estimates and its generalizability to larger populations (Fellows and Liu 2008). With nonprobability based samples in particular, we cannot estimate the size of the sampling error (random error due to chance) nor determine what, if any, sample bias (nonrandom or systematic errors) exists in the data. In probability-based sampling, the statistical power of the analysis increases as sample size increases (reducing sampling error and increasing confidence levels in a quantifiable way). By using convenience and other nonprobability samples, however, the researcher gives up the ability to calculate sampling error or meaningful confidence intervals.

The use of nonprobability samples appears to be common in construction research based on the analysis of the February–August issues of the JCEM. Nearly all of the 10 papers that presented findings based on surveys either acknowledged they used convenience samples or did not address at all whether the sample was probability based. Given that none of the 18 papers that used one or more case studies to illustrate or empirically support theoretical frameworks or tools stated the source of the case studies, it is reasonable to conclude that all of the case studies would be considered convenience samples.

To the extent that data derived from a convenience or snowball sample and data collected by an alternate means or method (from the application of a mixed method research design) converge on a common pattern or result, our ability to generalize from these types of samples increases. Mixed method research designs can therefore help us overcome some of the inherent limitations of any particular source/sample of data when (or if) using methods with complementary strengths (and weaknesses).

Theoretical Concepts Must Be Explicitly Defined before They Can Be Measured

The meanings of most engineering terms are unambiguous. Stress, flow rate, and viscosity, for example, are well understood across the engineering community. Many construction management terms, on the other hand, like many concepts in the social sciences, are somewhat imprecise, with meanings that can and often do vary between researchers and research contexts. For example, leadership may be considered the ability to articulate an organizational vision, to mobilize resources toward a set of goals, to motivate coworkers to perform at their best, to enact change, or to establish systems that enable operational efficiency. Similarly, technological innovation may be thought of as creating new technologies, adopting new technologies, transforming processes through a new type of system, or establishing a culture of embracing anything based on new technology. Effectively researching many construction management research topics thus requires researchers to define explicitly their theoretical concepts (referred to as "constructs" in the social sciences) at the beginning and to maintain a uniform definition throughout the research process.

In social science methods, the terms concept and construct, often used interchangeably, refer to the theoretical labels for traits or characteristics that exist analytically, at the abstract level, but are not directly observable, such as job satisfaction, leadership, or innovation. To the extent that researchers may attach somewhat different meanings and measures to abstract social concepts like these, they consist of "constructed" meanings, hence the use of the term construct rather than concept by some researchers. In order to develop valid and reliable measures, whether as survey questions, as indicators of experimental treatments, for behaviors observed in ethnographic study, or from archival sources, we therefore need clear and explicit conceptual definitions for each construct (variable) of interest (Fellows and Liu 2008).

The conceptual level of analysis is among the key attributes to be specified. Are we looking at a micro or macrolevel characteristic—that is, does the trait or characteristic in question pertain to an individual or to a group or organization as a whole? Emotional intelligence, for example, is clearly an individual trait. Effective teamwork is a group construct just as organizational financial performance is a group construct (in which the organization constitutes the social group and the quality pertains to the group as the unit of analysis). Some constructs, however, such as leadership, are often used to describe behaviors at more than one level. For example, we say an individual exhibits good leadership skills over his or her subordinates, a team can exercise leadership by suggesting improved procedures within a firm, and a firm can demonstrate leadership by introducing a series of new products in a market. Although using the same term to describe micro and macrobehaviors is common in construction research, this is problematic methodologically. In multilevel modeling, for example, apparent causal relationships can be spurious effects of uncontrolled and/or unmeasured multicollinearity between overlapping micro and macrolevel measures.

How we therefore define and specify each concept is critical for: first, constructing meaningful theoretical tests of our hypotheses; second, the design of valid and reliable (adequate and accurate) empirical measures in the operationalization stage; and third, useful statistical analysis of the data. Furthermore, the negative results of confounding the use and meaning of constructs, as discussed earlier, are compounded when applying a mixed method approach, yielding results that not only may lack theoretical meaning but which then cannot be empirically interpreted or compared. Thus careful attention to conceptualization is critical—even more so when using multiple methods and measures.

As previously noted, 22 of the 43 papers published in the February–August 2008 issues of the JCEM included social constructs. Social constructs that were either the focus of the paper or key terms underlying the background discussions in these papers included cultural and political impacts of a project, management commitment, employee training, employee and organizational knowledge and reasoning, site safety, safety consciousness, project management performance, employee dedication and attitude, employee work overload and stress, construction quality, personality attributes, collaboration and cooperation, and client satisfaction. While the reader may feel that most construction researchers share a common understanding of these constructs on a broad level, the need for providing explicit definitions becomes more critical when one attempts to empirically investigate them, as discussed under the next principle.

Theoretical Concepts Must Be Operationally Defined to Provide Valid and Reliable Measures

Most engineering terms have a standard method for measuring a value associated with them. Stress, strain, and viscosity, as examples, are measured using specific types of equipment and a range of procedures such that any engineer would agree the resulting values should be accurate. Many construction manage-

J. Constr. Eng. Manage., 2010, 136(1): 108-116

ment concepts are neither uniformly defined nor easily measured. Consider the hypothesis that successful leaders ensure their firms are technologically innovative. How do we identify a leader in an organization? Are all individuals holding certain titles (president, chief financial officer, water resources group manager) leaders? Do all leaders in an organization have one of these titles? If we focus our study on individuals who are vice presidents or higher, how should we measure the success of each leader? Should we ask the individuals themselves if they are successful leaders? Should we ask their bosses or subordinates? Should we seek objective archival metrics (use existing organizational performance data) that demonstrate their organizations, divisions or corporate segments were recently successful, such as unit profits, growth in sales, productivity, or reported customer satisfaction?

Operationalization, the process of choosing appropriate empirical indicators, is critical to the success of a project. This process involves specifying the exact procedures to be used to measure the attributes or properties of each construct, defining which specific observable indicators will be used to measure each variable, and explaining how to interpret each indicator relative to the theoretical property being measured. When complete, operational definitions provide other researchers with clear instructions on how we measured each construct. Since there are multiple ways to operationalize social constructs in a survey questionnaire, we need to specify the exact question or questions asked and make explicit how we are interpreting the answers in terms of the concept or construct in question. For example, to operationalize "client satisfaction," we could include a global or overall question asking clients to rate their satisfaction with the construction firm's performance: "Overall, how would you rate your satisfaction with the performance of this firm? Please choose a number from 1 to 10, where 1=completely satisfied and 10=not at all satisfied." In this case, the lower the mean value among responses, the higher the level of satisfaction. Alternatively, client satisfaction could be operationalized by the percent of respondents who answered "fairly" or "completely" satisfied in response to the question, "In general, how satisfied have you been as a client with the work performed by this firm? Would you say you have been completely satisfied, fairly satisfied, fairly dissatisfied, or completely dissatisfied?" As these examples show, providing the exact wording for survey questions (or completely specifying the measures used from archival data) and explaining how we interpret the answers insures consistency on our parts as researchers and provides clarity to readers. Moreover, with careful operationalizations, we can assess measurement validity and reliability and replication studies (another aspect of triangulation) become feasible.

Measurement validity is the determination that an indicator actually measures what it is supposed to measure. If an indicator is not valid, there will be systematic error or bias in the measurement process and the resulting data. There was a doctoral student, for example, who spent many hundreds of hours collecting data from a large sample of construction firms for the student's dissertation. The hypotheses were important and interesting, pilot interviews seemed to indicate the hypotheses were accurate, and a reasonably large number of firms provided survey data. The research plan seemed sound other than concern voiced by the doctoral committee about the quality of the indicators. It was the latter concern, however, that turned out to be critical after the fact. The results were ultimately inconclusive statistically, apparently because the indicators did not capture the underlying constructs. Researchers need to review their indicators carefully in the planning stages to ensure high levels of measurement validity. Where possible and appropriate, using operational definitions and measures developed by others and already reported in the literature gives us high levels of each. If such measures are not available, then pretesting any new measures we create in a pilot study is essential. Finding out that a measure does not capture the concept in question after the fact (once the data are collected) leaves no room for adjustment or correction.

In addition to concerns about measurement validity, careful operational definitions can reduce concerns about measurement reliability. Reliability is based on the application of uniform measurement rules and the uniformity of measurement results over time. Having and applying operational definitions produces consistent and stable results (all else being equal). The same indicator should produce the same results when other relevant factors remain unchanged. Problems of measurement reliability can arise from inadvertent changes in the measuring instrument (unplanned changes in question wording, for example), in the observer or mode of observation (changing from personal observation to video taping between subjects), or in the phenomenon itself (as in spurious changes in reported answers due to fatigue on the part of subjects during a long interview). Social scientists generally agree that using archival indicators from existing sources of statistical data are more reliable (consistent and stable) than self-reported data because the latter are inherently subjective and may reflect changing individual biases and inaccuracies. Self-reported indicators may reflect individuals' inflated opinions of their own abilities. Answers about other people or issues may be biased because respondents fear that answering the question truthfully may lead to trouble with their boss or coworkers or that they will disappoint the researchers. Archival data, too, is subject to biases and problems of reliability and validity. Often the specific measures we need do not match the data or indicators available and/or we do not have access to information on the possible measurement errors that pertain to the ones we do use. Changing organizational politics and practices, moreover, can affect the quality of such data particularly with regard to the comparability of measuring change over time.

Unfortunately, identifying accessible and reliable archival indicators is difficult for many constructs. In some cases, it is not a lack of access to relevant corporate data but the nature of the constructs we are studying that provide the challenge. In fact, some social scientists unapologetically identify some constructs as latent constructs, meaning they cannot be measured directly by one or more indicators (Loehlin 1998). Structural equation modeling is a relatively recent statistical tool that uses factor analysis of multiple indicators to measure latent variables. Molenaar et al. (2000) and provide a good explanation of the use of structural equation modeling in construction research.

Measurement validity and reliability are not our only concerns in the operationalization process. As researchers, we also need to pay careful attention to indicators' levels of measurement and the degree to which they can be quantified. Interval and ratio level data, for example, allow for the most complex mathematical manipulations. Since the level of measurement for each indicator depends on how it is operationalized, and the statistical tools that can be used with each indicator depend on the level of measurement, decisions made in this stage of research are critical for those that follow. More information is better than less, and if an indicator is truly interval or ratio level, it should be collected in that form. If we are collecting data on employee seniority with the company, as measured in years of service, for example, it is better to ask a question that allows the respondent to specify how long they have been with the company or firm in single years rather than have them check off a box for a category that collapses the

J. Constr. Eng. Manage., 2010, 136(1): 108-116

data into larger groupings. If the critical threshold point for the effect of seniority on perceptions of leadership falls in the middle of a large preset category, we are likely to miss the significant effect in the analysis. Smaller groupings can always be created during the analysis by temporarily recoding and collapsing a large amount of data into fewer categories when needed, but if the data are only collected in a less precise categorical form, we have no way to redress the problem after the fact.

A mixed method approach, using two or more data collection methods whose validity and reliability problems counterbalance each other, enables us to triangulate in on the "true" result. Without careful operationalizations, however, the use of mixed methods (or multiple measures within a single methodological approach) only serves to create additional noise in the data rather than help illuminate the underlying patterns of interest.

Most of the 22 JCEM papers with social constructs that were analyzed did not identify the specific ways constructs were operationalized. This may not be surprising given that manuscript length limitations often preclude including survey instruments in the manuscript. The writers believe, however, that confirming the operationalization of key variables was appropriate is an important part of a peer-review process that is not possible given the typical manuscript peer-review process used by the JCEM and other leading construction journals.

Hypothesized Causal Relationships Must Be Stated Explicitly but Are Difficult to "Prove"

Essentially all research should start with at least one meaningful research question, which typically leads to at least one hypothesis investigated through empirical study. Most hypotheses are causal in nature. Assuming there is a plausible theoretical explanation causally linking the two variables, three conditions must be met to establish causality. First, we need to establish temporal precedence, indicating that the hypothesized cause comes before the hypothesized effect. Second, we need constant conjunction, demonstrating that changes in the independent variable are associated with changes in the dependent variable (each time the value of the hypothesized cause changes, there is a predictable change in the effect, which we often measure through the Pearson correlation coefficient or some other appropriate statistical measure of association). Finally, we need to establish nonspuriousness, thus eliminating plausible alternative explanations for the statistical relationships found (Cook and Campbell 1979).

Many construction researchers typically are able to establish a plausible theoretical connection and demonstrate constant conjunction but do less well with temporal precedence and the problem of spuriousness. We all know that correlation does not equal causality, yet many fail to acknowledge that regression only presumes causality; it does not prove it. The challenge to establishing temporal precedence is that it is usually done by observation and could, in the strictest sense, require us to conduct controlled experiments involving individuals and companies, to observe what happens after we manipulate the value of the independent variable. This is impractical in real-world settings. Instead, leadership researchers typically measure both the independent and dependent variables simultaneously and presume, by the use of logic, that the former has influenced the latter. For example, we might measure a firm's innovation level and a firm's financial performance at one point in time and presume that innovation has contributed to the firm's financial performance. An astute colleague, however, might argue the causal relationship is reversed, that is, that a firm's financial success provides them with the necessary resources to be innovative. In this case, rather than presume the direction of causality, the better approach is to collect data on both current and past innovative behavior and current and past financial performance. Although we are collecting all these measures simultaneously, using a quasi-longitudinal design, in which data are collected on past and present values of all key variables, provides some evidence of the time order of change when direct observation is not possible.

The example earlier also illustrates the importance and difficulty of establishing nonspuriousness. To reject alternative explanations for any statistical relationship found between two variables we need to be able to test and eliminate them. Because we are rarely allowed to manipulate independent variables and/or physically control all extraneous factors when doing human subjects research, spurious relationships are a constant threat. For example, if we were to measure the technological innovation levels and financial performances of 100 construction firms and found the two variables to be highly correlated, it may be that this association merely reflects the fact that both variables are related to the size of the firm, i.e., large firms are more innovative and achieve higher profit margins, not that more innovative firms are more profitable.

Nonspuriousness is the most difficult condition of causality to establish. There is no way to actually prove that the relationship between the independent and dependent variables is not spurious, that is, that you can reject *all* other possible explanations. We can use observation and logic to establish time order and observation and statistical tools to establish and even quantify the degree of statistical association or constant conjunction between variables. But how do we prove that there are no other likely explanations or hidden third factors? In fact, unless we can actually determine and test all possible alternative explanations, we can never truly prove the causal connection. We use a range of methodological techniques and statistics to hold constant or control what potentially spurious factors we can (that is why random assignment in experiments and random sampling for surveys are so important), but the possibility always exists that there is some alternative explanation that we did not consider (a third factor we did not take into account). As a result, social scientists talk about *confirming* or *disconfirming*, *accepting* or *rejecting* research hypotheses, never "proving" them (Judd et al. 1991).

Some researchers acknowledge the difficulty of establishing causality and as a consequence merely hypothesize that two variables are associated in some way. Such hypotheses are more technically appropriate but make a rather limited contribution to the literature. Real contributions to our understanding come from attempting to overcome both the methodological and theoretical challenges that inevitably arise from compelling causal hypotheses. With careful attention to the conditions of causality and the use of appropriate controls and caveats, causal inferences can be drawn even if causality cannot be proven per se. Where applicable, tests of statistical significance and confidence intervals can increase the strength of causal inferences. To that end, choosing appropriate statistical techniques in data analysis is critical. Beyond this, mixed method studies can increase our confidence in causal inferences well beyond the results of one study. When different types of data, collected by different methods, converge (whether in support of the hypothesis or not), the result is likely *not* due to accidental correlations in a particular data sample, however carefully collected.

As stated earlier, 18 of the 43 JCEM papers analyzed applied the theoretical framework or tool presented in the paper to one or more case studies. Unfortunately, perhaps due to manuscript

J. Constr. Eng. Manage., 2010, 136(1): 108-116

length limitations, the amounts of information about the case studies were always so limited that the case studies could not begin to serve as preliminary empirical validation of the theoretical model. As such, the discussion of the case studies at most provided support that the model could be applied to actual projects, but could not provide meaningful support that the model was accurate or reliable tests of the hypotheses as such.

Appropriate Statistical Analysis Is Critical to Meaningful Results

In construction research, linear regression is most often the statistical tool of choice. Two mistakes involving linear regression are common. The first mistake is to use regression even when the data violate the fundamental assumptions underlying regression: first, that dependent variables are continuous and approximate a normal distribution; second, that independent variables approximate a normal distribution or are categorical; and third, that the relationships between dependent and independent variables are linear in nature. Linear regression is a statistically robust procedure, but experts suggest the need to perform a variety of tests before and after performing multiple regression analysis to verify that the data or the model do not significantly violate the underlying assumptions [see Belsley et al. (1980) for examples]. Construction research data, especially those using Likert scales, are often neither continuous nor normally distributed. When this is the case or residual diagnostics indicate violations of regression assumptions, researchers should consider supplementing their standard regression analysis with probabilistic regression, logistical regression or other appropriate statistical techniques.

The second mistake common among some construction researchers is to perform only bivariate analyses. Many relationships between variables may be statistically significant but spurious, as discussed in the previous section. When two variables are found to have a statistically significant association or correlation (i.e., the chance of a type I error being 5% or less), we cannot be sure the relationship between them is nonspurious unless we control for the confounding effects of antecedent variables. We can do this using partial correlation analyses (assuming we have interval or ratio level data) or some comparable nonparametric statistical analyses (if we have nominal and/or ordinal level variables). As we noted earlier, the choice of multivariate statistical technique depends on the levels of measurement of the variables in the model (which derive from the indicators chosen in our operational definitions). In examining the relationship between firms' levels of innovation and their financial performance, we would want, for example, to look at the relationship between firm innovation scores and financial performance while controlling for firm size and other relevant variables.

If the assumptions of the data hold, a relatively easy way to eliminate plausible alternative explanations is to perform multivariate linear regression using a model that contains other potentially confounding variables [for other appropriate multivariate statistical techniques for professional research, see Vogt (2007)]. This way, we can estimate how much each independent variable in the model contributes to each dependent variable, controlling for all others. Moreover, we can compare the relative strength of the relationship of each independent variable to the dependent variable and estimate the contribution each makes to the overall amount of variance explained. Some researchers focus solely on the ρ values to identify relationships that are statistically significant, but ignore the fact that the adjusted R^2 (the square of the Pearson product moment correlation coefficient adjusted for the degrees of freedom) of the model is so low as to be essentially meaningless. Meaningful research results need to explain sufficient variation for the significant relationships to matter. How much explained variation is meaningful, however, varies across levels of analysis. With microlevel survey data in the social sciences, for example, explaining more than 10% of the variance is considered good and more than 20% is excellent. With macrolevel data, such as organizational data commonly used in construction research, the adjusted R^2 tends to be much higher since there is often some smoothing of the data when aggregate measures are constructed.

A growing statistical analysis trend in construction research is the use of factor analysis. A substantial portion of the papers using factor analysis that the second writer has reviewed over the past 5 years seemed to have used this technique inappropriately. Factor analysis is appropriate for identifying latent variables that underlie two or more of the specific variables measured. The rational is that many social constructs (such as leadership, innovation, and project management performance) cannot be adequately captured by only one measurement. One of several advanced statistical software packages can be used to identify the apparent latent factors underlying the data based on relatively high loadings (that is, statistical relationships) on the associated measured variables. Unfortunately, it has been the second writer's experience that some researchers have used the resulting factors without confirming that the factors identified by the software make intuitive sense. That is, regression analysis has been performed using latent factors that are composed of measured variables that are theoretically unrelated to one another.

As anyone in the construction industry knows, you need the right tool for the job. The statistical tools available for any given problem are the product of the various decisions made at each stage in the research process—from the type of sampling done, to the sample size, the operational definitions used, the levels of measurement for each construct, to the causal conditions being tested. No methodological choice, even choosing the statistical tools applicable to a given problem, is truly independent of the others. Each has consequences; each has limitations and particular sources of error (problems of validity and reliability). The best way to address the inherent limitations of any singular methodological approach (be it a survey, experiment, observation, or archival research) or individual measure (self-report, archival indicator, participant observation) is to apply more than one method to the problem—to utilize a mixed methods research design.

Mixed Method Research: No Single Method Is "Best"

By using multiple methods to study the same problem, we can detect recurrent patterns or consistent relationships among variables, results that are independent of one particular data source or type of measurement and its inherent weaknesses. Triangulation, simultaneously using multiple research methods or measures to test the same hypothesis or finding, is a valuable strategy in the research process, but more so when we mix methods that have different but complementary strengths and weaknesses (Fellows and Liu 2008).

In fact, triangulation is a term used in social science research methods for decades. It was first used to describe convergent validation from multiple operationalism, where multiple tests of the hypothesis are conducted using different operationalizations or empirical indicators of the constructs in each test (Campbell

J. Constr. Eng. Manage., 2010, 136(1): 108-116

and Fiske 1959). It then came to be used to refer to the application of multiple methods of data collection (mixed methods), the use of multiple data-collection technologies per se, as well as the use of multiple theories and multiple observers or researchers (Denzin 1989). The use of multiple theories to generate and test competing hypotheses and the use of multiple observers, while also forms of triangulation, are not applications of a mixed method approach per se. As such, they are not discussed further here [see Brewer and Hunter (1989) for detailed discussion of these forms of triangulation].

The simplest form of a mixed method approach is the triangulation of measures. This involves the use of multiple operationalizations of the same construct in a study. It allows us to see more facets of the phenomenon in question. In surveys, this can involve the use of both closed-ended (forced choice) and open-ended questions to measure a key construct (behavior, trait, or attitude, for example). It can also involve the use of several distinctly worded questions (whether open or closed ended) for the same construct. To the extent that the results from different measures of the same concept converge, our confidence in the results increases. Different measures (indicators) of a given construct are not likely to reflect the same problems of measurement validity since question wording and format affect how a subject interprets its meaning as does the method of administration (if the question is asked by mail, by phone or face to face). If, however, the answers remain consistent (data converge) no matter how many different ways you ask the question, the result is reliable and you are likely circling around something "valid" in the data.

The use of multiple methods of data collection, mixing qualitative and quantitative approaches in a mixed method design, is a broader form of triangulation—one that costs more but yields greater utility. Mixing qualitative and quantitative methods allows us to combine research styles whose strengths and weaknesses are counterbalanced. If the methods chosen only partially overlap in style, a study using more than one method, applied either sequentially or simultaneously, will provide richer, more comprehensive, data (Neuman 2000).

What social scientists call "fieldwork," such as ethnography or participant observation, allows us to study natural, often complex behavior in its natural setting, and to study a problem holistically from the subjects' point of view while observing change over time. This approach, however, is confined to relatively small, localized groups, such as communities and organizations. We cannot measure past behavior and must be very careful when we generalize to large populations from such detailed case studies. Surveys, in contrast, allow us to collect a lot of information about large populations but only on topics that can be self-reported (whether verbally or in writing, face to face or via phone or computer). What surveys cannot do is allow us to study nonverbal behavior or observe change over time directly. Experiments, whether lab experiments or natural field experiments, provide the strongest controls over the conditions of causality, but many if not most of the topics studied by social scientists, let alone those of interest to construction researchers, are generally not open to direct experimental manipulation either physically or ethically. Finally, unobtrusive methods (like the use of archival data) are considered nonreactive (the responses are not affected by the measurement process or by interaction with the researcher) and may be broadly applicable. But they are limited to the study of those constructs and hypotheses for which data already exist and/or to those situations in which unobtrusive observation is ethical.

Field methods tend to yield highly valid but not highly reliable data. The results have limited generalizability and little control over the conditions of causality. Survey research often yields highly reliable measures and, if probability-based sampling is used, the results can be highly generalizable. But surveys are reactive methods and are subject to error in self-reports (deliberate and not) while providing a limited ability (via statistic controls) to establish causality. Experimental designs provide the strongest support for causal inferences. Lab experiments allow the greatest "control" over subjects and the experimental conditions and therefore have the highest internal (experimental) validity. Natural or field experiments, on the other hand, have lower internal validity but higher external validity (the results are more broadly generalizable to the real world application of the experimental treatments). Although unobtrusive measures are free from the problem of reactivity (subjects are not altering their behavior in response to being studied), a problem that pertains in varying degrees to different types of field work, surveys and experiments), it is the least flexible method. Unobtrusive techniques are often less precise in the available measures and provide little if any control over conditions of causality. As Brewer and Hunter noted:

> "... fieldwork's realism is restricted by its small scale ... The ability of survey research to generalize is limited by the ability to define theoretically relevant populations that can be readily sampled and questioned. The causal clarity of experimental research ends when there are confounding influences present that cannot be eliminated by control and randomization. And nonreactive studies' freedom from reactive measurement effects can never be taken for granted. Even in historical research, the possibility exists that the documents, statistics ... have been altered or edited in anticipation of scrutiny" (p. 47).

The debate over the relative value of fundamentally different research approaches has persisted in the social sciences for many decades and has generally focused on several sets of related dichotomies: quantitative versus qualitative research, nomothetic versus ideographic explanations, and positivist versus phenomenological approaches. To investigate a hypothesis involving leadership in technological innovation, for example, should we use a survey to obtain a relatively large random sample of objective data that can be subjected to rigorous statistical analysis, or do a 3–6 month ethnographic study in which we shadow a few top managers in several firms? The survey approach, using a stratified sample, would allow us to report statistical relationships with a high degree of confidence, yet we might omit key variables in our survey or be unable to explain some of the relationships that appear in the data. The ethnographic approach, on the other hand, would provide us with the richly detailed observational data and the opportunity to test highly nuanced explanations. Yet it opens us up to criticism that our writing merely reflects filtered observations and subjective interpretations rather than "facts" and that it is too limited in scope to be generalizable to the larger population of firms or leaders.

The mixed method approach is based on the premise that an effective body of research on a topic should include more than one research approach. A typical practice in construction is to distribute surveys to large samples via the postal service, e-mail or a website, perform statistical analysis on the data collected, and to supplement the survey with interviews before or after the survey. Another common method is to do a detailed case study of one issue within a firm, or minicase studies with four to twelve firms. While there have been a few "insider stories" published on specific construction trades and major projects, these do not meet

J. Constr. Eng. Manage., 2010, 136(1): 108-116

the standards of ethnographic research. Social science research suggests that construction research would benefit greatly from more ethnographic studies, especially in research areas still in the exploratory stage, which are then followed with larger systematic surveys. Although most construction researchers have neither the interest nor capability to do long-term ethnographic studies, other less time-consuming forms of observational or field methods could nonetheless lead to the development of better indicators, survey instruments, and sampling designs.

Aside from the use of field methods, a mixed method approach could incorporate unobtrusive research methods (the use of existing organizational or industry data, for example) to complement self-reported survey data. At a minimum, the principle of triangulation implies that even with a single method of data collection, multiple measures of key variables should be used. In this way, the researcher is not reliant on single indicators that individually could yield invalid or unreliable information. Within a standard survey questionnaire or interview schedule, including several different questions and versions of a question to measure the same theoretical concept meets the spirit of a mixed or multimethod approach.

When doing mixed method research projects, however, the researcher does face some new or additional challenges for cross-validation of data to be possible. One problem that can arise in mixed method research is that the response rate varies markedly among different approaches. Mail out/mail back surveys have very low response rates, typically, in comparison to experiments or personal, face-to-face interviews. Furthermore, it is unlikely that the subgroup of the population that does not respond to surveys is the same as the subgroup who avoids field researchers, who are likely omitted from official records, and/or fail to complete an experiment. In fact, subjects typically missing from each methodological approach are likely to differ systematically. The desire to counterbalance these particular sources of error and increase the overall generalizability of the study is one more reason to utilize a mixed method design.

Although the benefits of mixed method research designs include increased reliability and validity of the data and greater confidence in tests of the hypotheses (and the resulting conclusions), there are added costs for both the researchers and the research subjects. Since the research costs, generally counted in terms of time, money, and energy, are not identical for each approach, and the additional costs for additional methods of data collection enter into the project at different stages for different methods, there are tradeoffs between additional methods in terms of time and money of the researchers and the time and energy of subjects:

> " . . . while a field or nonreactive study plus a survey or an experiment is more expensive than either alone, the costs involved are not necessarily borne simultaneously, nor are they doubled for all parties, nor do they entail double consumption of the same resources. Rather, field and nonreactive studies most often chiefly consume researchers' time and energy in the search and analysis phase, while surveys and experiments consume researchers' time and energy more heavily in the design and instrumentation phase, and require more money and more of subjects' time and energy in the recording of observations . . . [mixed method designs are] no cheaper in an absolute sense. However it does increase its feasibility by spreading costs among different parties and budgets, and throughout the research process" (Brewer and Hunter 1989, p. 96).

Even with the additional costs, mixing methods with different strengths and weaknesses (e.g., a survey with an experiment, a survey with a field study, a survey with archival research, interviews, and archival data) helps us in the process of causal inference. If the data converge to support the hypotheses, we can have greater confidence that the results are not in fact spurious or mere artifacts of one particular body of data. If the results of different methods diverge, however, the lesson may be equally important in that unexpected findings should push us as researchers to reexamine our theoretical assumptions and improve our methods and measures. Whether a mixed method design results in data convergence or not, it is likely to provide more valid and reliable data and thereby allow us to have greater confidence in our conclusions—whatever they might be.

Conclusions

This paper has reviewed several basic social science principles that construction researchers should follow, along with a mixed method approach to research design, in order to increase their ability to draw credible and compelling conclusions from empirical research. The benefits of using a mixed method research design are especially applicable to areas of construction research that align with the social sciences rather than with, for example, operations research or information technology. Adhering to the methodological ABCs helps researchers reduce threats to validity and reliability while mixed method studies simultaneously increase the likelihood that the research will make a meaningful contribution to the literature.

It should be clear to the reader that the issues raised in this paper regarding research methods are only the proverbial tip of the iceberg. This paper is intended to serve as an introduction to issues in effective research methods in construction and to principles of triangulation in social science methods in order to encourage researchers to more fully study the topic. Effective research on topics involving human behavior in construction is hard work and requires expertise that is rarely provided in engineering or construction management curricula. To understand the full range of methodological challenges faced when doing research with and about human subjects, graduate students in construction should take at least one course on social science research methods, preferably one that covers both qualitative and quantitative approaches, and one course on multivariate statistical analysis before initiating their research. Including social science research methods courses in a graduate curriculum in construction management is unusual and perhaps controversial. But somehow integrating the knowledge derived from such courses is an important step if the construction research community is to maintain a reputation for effective research.

References

Babbie, E. (2008). *The basics of social research*, Thomson Wadsworth, Belmont, Calif.

Belsley, D. A., Kuh, E., and Welsch, R. E. (1980). *Regression diagnostics*, Wiley, New York.

J. Constr. Eng. Manage., 2010, 136(1): 108-116

Brewer, J., and Hunter, A. (1989). *Multimethod research: A synthesis of styles*, Sage, Newbury Park, Calif.

Campbell, D. T., and Fiske, D. W. (1959). "Convergent and discriminant validation by the multitrait-multimethod matrix." *Psychol. Bull.*, 56(2), 81–105.

Campbell, D. T., and Stanley, J. (1963). *Experimental and quasi-experimental design for research*, Rand McNally, Chicago.

Chambliss, D., and Schutt, R. K. (2006). *Making sense of the social world: Methods of investigation*, Pine Forge, Thousand Oaks, Calif.

Cook, T. D., and Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*, Houghton Mifflin, Boston.

Denzin, N. K. (1989). *The research act: A theoretical introduction to sociological methods*, Prentice-Hall, Englewood Cliffs, N.J.

El-Diraby, T. E., and O'Connor, J. T. (2004). "Lessons learned in designing research methodology in field-based construction research." *J. Profl. Issues Eng. Educ. Pract.*, 130(2), 109–114.

Fellows, R., and Liu, A. (2008). *Research methods for construction*, 3rd Ed., Blackwell Science, Oxford, U.K.

Judd, C. M., Smith, E. R., and Kidder, L. H. (1991). *Research methods in social relations*, 6th Ed., Holt, Rinehart, and Winston, New York.

Loehlin, J. C. (1998). *Latent variable models: An introduction to factor, path, and structural analysis*, Lawrence Erlbaum Associates, London.

Loosemore, M. (1999). "International construction management research: Cultural sensitivity in methodological design." *Constr. Manage. Econom.*, 17(5), 553–561.

Molenaar, K., Washington, S., and Diekmann, J. (2000). "A structural equation model of construction contract dispute potential." *J. Constr. Eng. Manage.*, 126(4), 268–277.

Neuman, W. L. (2000). *Social research methods: Qualitative and quantitative approaches*, Allyn and Bacon, Needham Heights, Mass.

Raftery, J., McGeorge, D., and Walter, M. (1997). "Breaking up methodological monopolies: A multi-paradigm approach to construction management research." *Constr. Manage. Econom.*, 15(3), 291–297.

Runeson, G. (1997). "The role of theory in construction management research: Comment." *Constr. Manage. Econom.*, 15(3), 299–302.

Seymour, D., Crook, D., and Rooke, J. (1997). "The role of theory in construction management: A call for debate." *Constr. Manage. Econom.*, 15(1), 117–119.

Vogt, W. P. (2007). *Quantitative research methods for professionals*, Pearson Education, Boston.

Walker, D. H. T. (1997). "Choosing an appropriate research methodology." *Constr. Manage. Econom.*, 15(2), 149–159.

J. Constr. Eng. Manage., 2010, 136(1): 108-116