

A Review of the Strengths and Weaknesses of Archival, Behavioral, and Qualitative Research Methods: Recognizing the Potential Benefits of Triangulation

Amy Hageman

University of Central Florida

ABSTRACT

This paper discusses the benefits, limitations and challenges in developing research projects that integrate a combination of archival, behavioral, and qualitative research methods. By demonstrating the inherent strengths and weaknesses of a using a single method in isolation, this paper aims to broaden our understanding of why and how research that examines various issues from the different perspectives is richer than employing any single method and enhances our understanding of a given accounting phenomenon. This paper also discusses how investigating an issue through multiple research methods can help researchers improve the generalizability of findings and present a panoramic view of a particular phenomenon.

INTRODUCTION

Accounting is an applied discipline, within which researchers rely on theories from a wide array of root disciplines to investigate research questions. Perhaps because of the variety of root disciplines that have been influential in helping to motivate accounting research (particularly economics, psychology, and sociology), accounting researchers employ a variety of research methods to conduct empirical studies of accounting phenomena. Some of the most common research methods can be broadly classified as *archival* (secondary sources of, mostly, numeric data at the organizational unit of analysis), *behavioral* (primary sources of, mostly, numeric data

at individual, group and organizational units of analysis), and *qualitative* (secondary and primary sources of, mostly, non-numeric data at individual, group and organizational units of analysis).

Currently, researchers using archival research methods vastly out-number those who use behavioral or qualitative techniques. A survey by Koonce and Mercer (2005) of five of the top accounting journals showed that from 1993-2004, over 94% of financial accounting studies used archival research methods, compared to less than 6% that employed behavioral methods. Similarly, a more comprehensive survey of 14 accounting journals from 1981-2004 by Merchant and Van der Stede (2006) found that very little accounting research used field research and other qualitative techniques. Part of the reason for the dominance of archival research methods concerns the teaching philosophies of accounting many doctoral programs – whereas virtually all U.S. accounting doctoral students take at least one course emphasizing archival research methods, fewer than half are typically exposed to behavioral methods (Koonce and Mercer, 2005) and only a small number are trained in qualitative research techniques (Merchant and Van der Stede, 2006).

Equally troubling is that doctoral students' training often emphasizes the *research method* rather than the *research question*. As stated by Koonce and Mercer (2005, p. 177), "Accounting doctoral students typically choose one of two, largely non-overlapping, fields of specialization...this choice is often determined by the type of method the student expects to use in his or her coursework [archival or behavioral]." Appropriate research methods should certainly match a study's research questions. However, given that most future accounting researchers are trained to examine a problem within the confines of a particular research method, most researchers rely on a singular method in examining their research questions, and limit their views of their particular paradigm based on the research method used (see Burrell and Morgan,

1979). This limitation is perpetuated by the inherent difficulties in developing and maintaining skill sets in multiple methods.

The purpose of this paper is to discuss the benefits, limitations and challenges in developing a well-designed study using any single research method in isolation. In particular, raising accounting researchers' awareness of the limitations of their "dominant" research method is important, as it can help foster deeper understanding of the potential contributions of other techniques. While researchers within the management accounting discipline have long recognized the importance of using integrative research methods (e.g. the call for cross-sectional field study research in Lillis and Mundy, 2005), this paper strives to show how *triangulation*¹ of research methods (using two or more research methods to study a given topic) can produce a richer, more complete understanding of all accounting phenomena. Currently, the academic accounting community emphasizes the use of archival research methods at the expense of behavioral and qualitative research. By discussing the shortcomings in each of these types of research methods, this paper aims to illustrate how multiple papers on the same research topic should use differing methods in order to overcome the weaknesses in collectively relying upon a single approach to research.

This paper is especially aimed at less-experienced academics (i.e. doctoral students and new faculty members) as such individuals can benefit from understanding how each research method presents its own set of challenges, and why investigations of accounting phenomena through multiple methods helps present a more comprehensive view of a specific phenomenon. The discussion of alternative research techniques within this paper can also be informative to more experienced researchers who wish to investigate research questions using methods that are outside of the focus of their initial training.

Many previous authors have certainly noted the strengths and weaknesses of archival, behavioral, and qualitative research methods in the social sciences in general (e.g. Kerlinger and Lee, 2000; Shadish et al. 2002; Berg, 2003; Greene, 2003) and in accounting research in particular (e.g. Kothari, 2001; Libby et al. 2002; Merchant and Van der Stede, 2006). While all of these articles and textbooks constitute important examinations of specific research methods, none provides a combined overview of *all* of these different methods in the context of accounting research. This paper therefore aims to synthesize information on the benefits and limitations of research methods as used in accounting research within the framework of a single paper, which overviews the research methods field and provides new scholars with a single resource for beginning their investigations of how to conduct research.

The remainder of this paper is organized as follows: The next three sections detail the challenges involved in conducting a well-designed study using archival, behavioral, and qualitative research methods. The subsequent section provides a brief example of how examining a research topic with multiple methods can help produce a more complete understanding of the underlying accounting phenomena. The final section concludes with a discussion of the importance of triangulation in accounting research.

ARCHIVAL RESEARCH METHODS

Strictly speaking, the use of an “archival” research method entails the use of secondary data sources, in which researchers analyze data contained in an archived record. However, as commonly used in the accounting discipline, “archival” denotes the use of mostly *numeric* data; in practice, many studies employing this method use large-scale, secondary numerical data in the positivist tradition (Burrell and Morgan, 1979).² This type of research method was not common in academic accounting research until the paradigm-shifting work of Ball and Brown (1968) and

Beaver (1968), and today represents the dominant research method of first choice when conducting accounting research (see Watts and Zimmerman, 1986; Kothari, 2001; Koonce and Mercer, 2005).

The archival research method is particularly useful in its ability to examine trends in large-scale data. Thus, external validity is particularly high in studies using archival research methods, as such studies use data pertaining to *naturally occurring* events. The archival research method is particularly appropriate for examining *macro-level patterns* (broad economic or societal trends), such as general economic trends over time, but is also commonly used for examining micro-level behavior in the *aggregate*. Many archival studies therefore use *econometrics*, or “the application of mathematical statistics and the tools of statistical inference to the empirical measurement of relationships postulated by economic theory” (Greene, 2003, p. 1). This analysis is useful to accounting researchers who examine large-scale trends of naturally occurring events, such as the stock market’s reaction to a new accounting standard.

Despite the usefulness of archival research methods, challenges remain. These relate to the general problems in using secondary numerical data sources to draw causal inferences, as well as to the more mundane nuances associated with conducting a quality archival study.

Difficulty in Establishing Causal Inferences

One of the primary challenges associated with archival research is that researchers *cannot easily test causal relationships*. Relying on secondary data often compromises an archival study’s internal validity. The researcher can analyze data trends, but it is difficult to establish that a particular factor causes another. In most archival studies, researchers examine the relationships between certain variables and attempt to control for other alternative explanations that may have affected the relationship. Thus, archival research designs have relatively low levels of internal

validity, as it is difficult for the researcher to properly control for all other plausible explanations for an observed relationship between phenomena. This is particularly true for *non-experimental designs*, which lack randomization, control groups, pre-tests, and other factors; researchers instead measure and statistically control for alternative explanations (Shadish et al. 2002). Establishing causality with archival studies is especially challenging with *cross-sectional studies*, in that it is unclear whether the purported cause actually precedes the event; it is therefore critical that all other potential causal factors are well-measured and controlled, and that the model is very well-specified (Shadish et al. 2002).

Since archival research cannot clearly indicate causality, *endogeneity* of independent variables is another common concern. Researchers may test the effect of an independent variable on a dependent variable, but ignore the antecedents of the independent variable itself. An analyst's affiliation is one such endogenous construct, as companies may choose analysts that are naturally optimistic (Kothari, 2001).

Archival researchers attempt to address the difficulties of establishing causal inferences in several ways. First, since establishing causality in archival research is difficult, the researcher must clearly *control for and explain other alternative explanations*. This manner of helping to mitigate threats to internal validity often requires the researcher to become immersed in statistical methods and techniques, since controlling for other explanations can be challenging.

Researchers conducting archival research may encounter difficulties when the model violates assumptions of multivariate analysis, such as normality, homoskedasticity, and linearity (Hair et al. 1998). Researchers who are aware of these common problems may try to transform their data before the analysis, such as *scaling variables* by the size of assets in order to help eliminate heteroskedasticity or by taking the natural log of a variable (such as audit fees in

Simunic, 1980). However, all of these statistical techniques represent a high barrier to entry for those not well-trained in these methods, as those not well-trained in statistical techniques are unlikely to succeed in conducting the extensive statistical testing and analysis required to assure the reliability of the results.

A second way of alleviating the problem in archival research of establishing causal inferences is the use of *natural experiments*, in which the researcher examines the relationship between a naturally occurring event and a comparison event (Shadish et al. 2002). Natural experiments help address causality concerns by providing built-in temporal precedence of the proposed cause in relation to its effect. Many archival studies of earnings management use natural experiments in testing their hypotheses, such as studying earnings management among firms undergoing or not undergoing import price relief investigation (Jones, 1991). Moreover, many econometric studies have shifted toward the use of natural experiments as a tool for establishing valid inferences; the use of this technique is particularly attractive in the econometric domain since researchers cannot easily manipulate variables of interest (Shadish et al. 2002).

A third compensating technique used by archival researchers is the *matching of observations*. In this case, the researcher is not able to achieve random selection and random assignment, but can match observations on certain characteristics. This method is not foolproof, as it is very difficult to know which dimensions should be matched, or even that an exact match has actually been made, but it helps to alleviate some of the problems with pure non-experimental designs by assuring that the control and treatment groups have equivalent matched dimensions (Kerlinger and Lee, 2000). Some studies of earnings management have used matching techniques. For example, Holthausen et al. (1995) studied the relationship between

management bonus plans and manipulation of earnings, and matched firm managers' performance on bonus plans characteristics; Klein (2002) used a matched portfolio matched by abnormal accruals in examining the relationship between earnings management and corporate governance structures. The differences between matched pairs help to establish some of the relevant antecedents of earnings management.

Difficulty of Secondary Data Reliance

Another major difficulty in conducting quality archival research arises from this method's reliance upon secondary data sources. This reliance brings about a host of additional problems that archival researchers must address. *Measurement error* of the variables is abundant with the use of secondary data sources and is one of the chief difficulties in using archival research methods to test the relationship among variables in a model. Specifically, threats to construct validity are particularly prevalent in archival research, as sampled variables rarely are exact proxies for the latent theorized constructs. Indeed, Greene (2003, p. 8) writes that, "...the difficulty of obtaining reasonable measures of profits, interest rates, capital stocks, or, worse yet, flows of services from capital stocks is a recurrent theme in the empirical literature."

Measurement error due to poor construct validity arises from two related reasons – the data may be very poorly measured, or data that is needed to operationalize a theoretical construct may simply be unavailable in a secondary dataset. Poorly measured data is a common concern in archival studies. For instance, earnings management studies may use the Jones (1991) model of discretionary accruals to approximate the accruals that management may manipulate, but since this construct is unobservable to those outside of the company, the approximation of "discretionary accruals" contains a great deal of error. A related challenge is that some variables may be inherently immeasurable in numeric form, hence, not available in an existing dataset.

Archival studies that attempt to measure constructs such as investors' "expectations" about future stock market prices (see Greene, 2003) suffer from extreme measurement error, since no secondary dataset truly captures "expectations" and the construct of expectations itself is difficult to describe numerically. Thus, nearly all archival studies in accounting doubtlessly suffer from measurement error due to threats to construct validity, and empirically observing the proposed theoretical relationship may be difficult. Researchers acknowledge these threats to construct validity by carefully selecting their operationalized variable from an existing data set and disclosing limitations.

The reliance on secondary data is particularly problematic if there are *errors in the datasets themselves*. For instance, AuditAnalytics is a popular dataset containing information on auditing information by company such as corporate auditor, auditor fees, and Sarbanes-Oxley (SOX) disclosures and compliance. However, investigation of this dataset has shown that there are numerous inconsistencies between the information reported in AuditAnalytics and in the actual 10-K filings (Canada et al. 2007). This means that prior studies using information from AuditAnalytics may have relied on incorrect data. Alternatively, information may not match across datasets. For instance, unlike Compustat, I/B/E/S adjusts earnings for one-time events and special items (Kothari, 2001). Whether researchers are cognizant of these discrepancies or make the necessary transformations or adjustments is not always clear.

Another problem with relying on secondary datasets is that much of this research, particularly capital markets events studies, uses secondary data from a *long time period* in order to enhance the generalizability of the findings. However, with longer time horizons, controlling for all confounding factors is difficult. Another problem particularly prevalent in capital markets research is survivorship bias; observations not available for all periods are excluded from the

analysis, which in capital markets research biases the results toward older, more successful firms. One way to solve the problem of data availability is to use data available at greater intervals, such as quarterly earnings announcements; quarterly data helps increase the power of statistical tests, alleviate survivorship bias, and can potentially expose seasonality (Kothari, 2001).

Other problems arise from the use of *large-scale data*, particularly in capital markets studies. Variables in these studies may be serially dependent or biased by correlated omitted variables. These concerns may be mitigated by the use of *first differences*, in which data is lagged over a period; some researchers may choose to lag the data for additional periods. Controlling for cross-correlation in the data may also address these concerns (Kothari, 2001). With the use of large-scale data, archival researchers often practice data truncation, whereby outliers are removed from the analysis. This practice biases the results toward average values and away from the extremes.

Archival researchers must also contend with a number of *practical data limitations*. First, researchers that utilize publicly available secondary data, such as information from databases such as CRSP or Compustat, must compete with a large pool of other researchers also using the same data source. It may be harder to develop an original testable hypothesis that uses these widely held databases. Second, researchers that use data from a propriety source (such as Holthausen et al. 1995 and its use of a propriety compensation database) must incur large monetary or temporal expenses. This ensures that researchers may need a variety of resources to conduct high-quality archival research. Third, researchers are limited by the availability of data; some data is only available in limited periods. For instance, Omer and Shelley (2004) were unable to test their tax competition hypothesis before 1977 because a great deal of state-related data was not available before this time.

Finally, archival studies are limited by what has *actually occurred*. In addition to the fact that this enhanced external validity comes at the cost of reduced internal validity, archival studies are not as useful in testing the consequences of proposed policy action or regulatory change.

Overall

Overall, archival research methods are very useful in helping researchers examine macro-level patterns in naturally occurring events, but suffer from poor internal and construct validity, along with a host of other problems associated with using large-scale secondary data sources. In particular, the use of questionable proxies in pure archival research often limits researchers' attempts at understanding a phenomenon. Triangulation of archival methods with behavioral and qualitative research methods may help alleviate some of these concerns.

BEHAVIORAL RESEARCH METHODS

Behavioral research methods are particularly useful in understanding human behavior. These methods can focus on individuals, groups or organizations as units of analysis. "Behavioral" research is a broad umbrella encompassing a variety of research methods, including laboratory experiments (where participants engage in contrived tasks under controlled, yet artificially created, experimental conditions), surveys (where respondents reflect their beliefs, attitudes, cognitions and motivations under general conditions), experimental economics (where buyers and sellers interact in an artificial market setting), and field experiments (where participants work in their natural environments under various conditions of interest). However, as used in the academic accounting paradigm, the term "behavioral" typically denotes a laboratory experiment. Despite the use of behavioral research methods across all functional areas of accounting, however, the use of this research method is not as common as the archival method (see Koonce and Mercer, 2005).

The single greatest advantage of behavioral research that uses laboratory experiments is the ability to *establish causality* in relationships between and among phenomena. As stated by Shadish et al. (2002, p. 18), “The strength of experimentation is its ability to illuminate causal inference.” This is because in an experiment, the researcher can isolate and control for other potentially confounding factors, which provides evidence against alternative explanations. To establish causality, researchers need to demonstrate temporal precedence (that a cause occurred before an effect), co-variation (that the cause and effect are correlated), and that alternative explanations can be ruled-out (Shadish et al. 2002). Thus, a great benefit of experiments is rooted in the concept of control. This offers the experimental method a high degree of internal validity in that tightly-designed and well-executed experiments can illustrate causal relationships. While experiments in which the independent variables are manipulated and participants are randomized among experimental conditions have the highest degree of internal validity, quasi-experiments and experiments that measure the independent variables can still investigate causality to some extent.

In general, behavioral methods are also useful for helping to understand *individual differences*. Some archival studies seek to examine questions of individual behavior (such as how analysts incorporate information, e.g. O’Brien, 1988), but behavioral methods are particularly effective in shedding light on individual-level phenomenon. This is especially the case in studies of factors for which the use of archival methods is impossible, such as understanding knowledge structures or other cognitive elements (e.g. Nelson et al. 1995). In that sense, behavioral methods are often superior to archival methods in their ability to use *primary source data*.

In summary, Kerlinger and Lee (2000) characterize laboratory experiments (the most common behavioral method used in accounting research) as having the primary strengths of

control, randomization, precision, and manipulation and measurement. Because of these benefits, a true experiment is considered a scientific ideal, particularly when testing and building theory. However, laboratory studies, experimental economics, surveys, and field studies all have a host of limitations.

Laboratory Experiments

An informal survey of the journal *Behavioral Research in Accounting* indicates that laboratory experiments are the most common research method used by behavioral accounting researchers. While there are benefits from using laboratory experiments, they also have their own set of specific challenges.

One of the greatest strengths of the laboratory experiment is its high degree of control, such that researchers can learn about human cognitions and behavior by systematically varying certain factors and controlling for alternative explanations. Ironically, this very strength is also one of the primary challenges faced by behavioral experiments. In order to carry out studies with a high degree of internal validity, researchers must conduct studies in *contrived, artificial environments* that remove participants from their natural settings. This process results in a loss of external validity and can hamper the generalizability of the study. It is difficult for a tightly-designed experiment to also tap into the richness of the natural environment. For example, Ashton (1990) is an example of a study with a high degree of internal validity (all of the independent variables are tightly manipulated), but in which the generalizability is compromised (for instance, “incentives” in this study were operationalized as a tournament-type scheme that would not apply to practicing auditors). In summary, high internal validity and control come at a cost, the lack of generalizability and external validity; therefore, researchers must exercise caution when generalizing the results of laboratory studies to non-laboratory settings.

Researchers must balance internal and external validity concerns when designing experiments. However, it must be reinforced that the main purposes of experiments are to test and build theory, and that external validity can be achieved to some degree by extrapolating experimental results through theory, not through tasks or settings.

A related challenge is that laboratory manipulations often produce relatively *weak effects* (Kerlinger and Lee, 2000). This is related to the fact that the laboratory is a contrived environment, where reproducing the precise effects that may be seen in the natural setting is quite difficult. Kennedy (1993) is an example of a weak manipulation; the study tested order effects, but auditor participants did not exhibit order bias in the experiment. In order to determine whether the manipulation has the intended effect, experiments generally include manipulation checks to ensure that participants view the experimental treatment in the same light as the researcher. Behavioral researchers typically conduct extensive pre-tests to ensure that the manipulations will work, and make changes to the experimental design if effects are not detected. In most instances, participants that do not pass the manipulation check are excluded from further analysis, although some studies only examine whether participants in experimental conditions passed the manipulation check in the aggregate. Researchers must carefully design experiments to avoid the limitations of weak effects.

Beyond the limitation of weak effects, laboratory experiments *cannot estimate the magnitudes of effects*. In an experiment in which the researcher manipulates one or more independent variables, the results can provide information on whether an effect occurred, but it is difficult to quantify the absolute magnitude of such an effect. Thus, it is safe and appropriate for experimental researchers to interpret their findings as ordinal, not interval, effects.

Laboratory experiments often require participants to perform a *task* during the experiment, which means that the researcher is responsible for developing an accurate, realistic task that can be used to test the study's hypotheses. Developing a task that is both appropriate for participants and can be used to adequately test the study's hypothesis is challenging, particularly when the study is examining a sensitive issue. For example, Hunton et al. (2006) required participants to determine which available-for-sale security to sell as a proxy for earnings management for the study's investigation of whether more comprehensive disclosures help to mitigate earnings management behavior. While the ability to generalize about earnings management behavior within the confines of this particular task is somewhat limited, through theory, their study does indicate that when given sufficient discretion in accounting standards and personal incentives, managers will likely use such discretion to achieve personal gains. On a related note, determining how to *operationalize the study's constructs* of interest may be difficult. Behavioral studies often examine individual behavior in terms of psychological constructs that may be difficult to define or properly measure, such as "procedural knowledge" (e.g. Bonner and Walker, 1994).

Furthermore, proper laboratory and other behavioral studies require adequate *participants* who are an appropriate match for the experimental tasks. Accounting studies in all functional areas tend to use practitioners when appropriate, such as the use of auditors for much of the judgment and decision-making literature stream (Gramling et al. 2001). This means that researchers must recruit practitioners, which can require significant amounts of time and money, and may be difficult if CPA firms refuse to cooperate due to liability or other concerns. Unlike the psychology discipline, accounting behavioral studies have access to a limited pool of potential participants (see Gibbins, 1992), so it is very difficult to replicate an experiment if

something goes amiss. Due to the difficulties in obtaining accounting practitioners as participants, many accounting researchers conduct experiments with undergraduate or graduate students. While the use of student subjects may be appropriate if the experimental task is properly constructed (Elliott et al. 2007), many studies relying upon student participants do not satisfy this criterion (see discussion in O'Neil and Samelson, 2001).

Experimenters must also ensure that participants are properly *motivated* and are engaged in the task. Subjects that are not properly motivated by the degree of experimental realism or via attention to the task will not produce usable data. Some researchers use incentives in order to increase attention to the task at hand; however, using incentives to motivate participants can introduce its own series of problems (Bonner and Sprinkle, 2002). Overall, there is extensive pressure in each experiment to make sure that all procedures are properly implemented.

Another challenge for the researcher is that participating in an experiment exists as part of a *social situation* (Shadish et al. 2002). Within a laboratory experiment, participants may respond differently due to social cues from the researcher. Likewise, many experimental findings are dependent upon participants' perceptions of the task or of measurement items. This means that researchers must rely upon individual-level data that is often perceptual and may not adequately measure the construct of interest (Kerlinger and Lee, 2000).

These inherent limitations in experiments lead to the constraint that most experiments also require *extensive work prior to the collection of data*, through theoretical development, experimental design, task development, pre-tests, human subjects' approval, and so forth. Since a tight experimental design is critical to achieving experimental control, researchers must ensure that their task, measurements, and procedures are perfectly developed. While all research methods require an extensive, pre-data collection time commitment, making changes to the

research design after data collection has begun is extremely difficult when using behavioral research methods. This restriction can add to considerable development time; for instance, the software Insolve, which is used in many experiments that study insolvency practitioners' judgments (e.g. Arnold et al. 2000), took nearly seven years to develop (Leech et al. 1998). Furthermore, experimenters also need to control for the fact that human information processing is subject to bias (Tversky and Kahneman, 1974). As an example, individuals are influenced by the order of presentation (Hogarth and Einhorn, 1992), so experimenters must take care to randomize the presentation of experimental materials. Taken together, the time to develop an experiment and the factors that must be considered are quite challenging.

Finally, experiments are not immune from problems with *measurement error*, including threats to construct and internal validity. These threats are especially prevalent for the experimental variables that are challenging to manipulate in the laboratory and are instead measured. Many studies that examine "experience," for example, tend to measure participants' years or level of experience (e.g. Tubbs, 1992). While this measurement can be surmounted (for instance, Hampton, 2005), independent variables that are measured provide weaker results than those that are manipulated, and may not necessarily be appropriate proxies for the higher-level constructs of interest.

Experimental Economics

Experimental economics may be thought of as a special type of laboratory experiment. In this research method, experimenters create an artificial market in order to study the behavior of buyers and sellers. Some experimental economics studies utilize multiple periods to study how behavior may change over time (e.g. Sprinkle, 2000). This specificity of focus is one of the primary strengths – researchers are able to reproduce a marketplace to determine how purchasers

and vendors behave. Many of these studies rely on student participants, as students are appropriate proxies for vendor and consumer behavior.

Like the laboratory experiment, experimental economics benefits from very tight experimental design and high internal validity. This high internal validity comes at a price of *reduced realism and generalizability*. Indeed, some commentators have criticized experimental economics studies for developing such a low level of experimental realism as to be practically meaningless. An example is criticisms of studies of taxpayer behavior using this method and removing all social and moral contexts (O'Neil and Samelson, 2001).

Practically, experimental economics studies are *expensive* to run. Since most of these studies examine buyer and seller behavior within an artificial market, researchers must provide funds to mimic a marketplace; many studies guarantee a minimum payment to all participants, but the determination of how much each participant is awarded depends on his/her performance (e.g. Sprinkle, 2000). Given that this method is expensive to conduct, the researcher must ensure that the experiment is perfectly executed.

Surveys

The use of the survey method uses participants to respond to questionnaires; these questionnaires contain scales with items that measure certain constructs of interest. There is no manipulation involved; instead, the researcher measures all of the variables of interest. A review of *Behavioral Research in Accounting* indicates that the use of surveys has waned in most recent behavioral accounting research, although their use continues to be more popular in the management accounting domain (as evidenced by an examination of management accounting journals such as *Journal of Management Accounting Research* or *Management Accounting Research*). Still, surveys do have some advantages over experiments, particularly in that they

may be cheaper to conduct and may garner a higher degree of realism and external validity than found in an artificial laboratory environment. A well-administered survey may also generate a greater volume of information than in a single experiment (Kerlinger and Lee, 2000).

One of the greatest challenges associated with survey research, however, is that the data collected from this method is entirely *perceptual*. Thus, the value of survey data is only as valid as the reporting of participants' perceptions. Many people may not respond, tell the truth, or know the answer to the question, which further biases the results. Some techniques help to alleviate these concerns. For instance, following the techniques of survey design in Dillman (2006) can help to alleviate non-response bias, whereas techniques such as the randomized response technique (see Bailey et al. 2001) can help compensate for individuals' tendencies to answer in an inaccurate, socially desirable manner. Nevertheless, the information provided by survey respondents is often impossible for researchers to verify (Kerlinger and Lee, 2000); and, these self-reports may threaten the study's construct validity (Shadish et al. 2002).

Practically, even when attempting to control for non-response bias, most surveys still suffer from *low response rates*; response rates of around 10% are not uncommon when randomly surveying accounting practitioners (e.g. Bobek and Radtke, 2007). Furthermore, like an experiment, a survey is a social event, in which results may suffer from *social desirability bias* (participants respond in a manner that is socially desirable, but does not reflect their actual outcomes or attitudes).

Field Experiments

The final primary category of behavioral methods is the field experiment, in which researchers undertake a company investigation to study relationships among variables. Typically, researchers utilize a realistic setting and manipulate an independent variable to the degree

allowed by the organization or other setting (Kerlinger and Lee, 2000). This type of research design is similar to a laboratory experiment, but is typically carried out in a more *realistic*, naturalistic environment. This type of method enables the examination of *richer*, more complex problems than seen in laboratory experiments.

Despite these strengths, several challenges remain in conducting reliable field experiments. The first is that field experiments have a lower degree of internal validity than laboratory experiments, because “the independent variables are contaminated by uncontrolled environmental variables” (Kerlinger and Lee, 2000, p. 582). This means that the researchers should implement as many controls as possible over other potential variables to be assured of the reliability of the results. Furthermore, true *randomization* is very difficult in a field experiment. Many companies may tend to resist having their employees randomly assigned to conditions; therefore, in most field experiments, the independent variable of interest cannot be disentangled from other potential factors. Threats to internal validity due to field experiments’ lack of control introduce more random noise in the results than in laboratory experiments.

Overall

Overall, behavioral research methods attempt to collect primary data from individuals in order to gain knowledge about human behavior. Whereas all methods grant the researcher much more flexibility in investigating research problems than do archival methods, all types of behavioral research methods suffer from challenges. In short, because researchers collect data from individuals in behavioral studies, all data is in some sense self-assessed. While some methods have more control over this than others (i.e., laboratory experiments are better controlled than surveys), all research is dependent upon individual responses for data. Thus, behavioral and archival research methods can complement each other.

QUALITATIVE RESEARCH METHODS

Qualitative research methods rely mostly on non-numerical data; Kerlinger and Lee (2000, p. 588) formally define this method as, "...social and behavioral research based on unobtrusive field observations that can be analyzed without using numbers or statistics." As compared to archival and behavioral studies, qualitative studies use an interpretive approach that is more focused on understanding *meaning* than on assessing causal relationships. Thus, qualitative research is often used to investigate more complex 'how' and 'why' research questions.

Qualitative research is a broad category with research traditions in fields such as sociology, anthropology, or education (Kerlinger and Lee, 2000). As used in academic accounting, qualitative studies of data are typically either *field studies* (involving direct contact with real-world participants) or a *content analysis* (involving non-numerical analysis of primary or secondary communications). While this research method is less common in the North American academic accounting community and is not generally taught in doctoral education programs (Merchant and Van der Stede, 2006), its use is generally more accepted in European and Australian universities.

The greatest advantage of qualitative research methods is in the ability to analyze *naturalistic environments* (Merchant and Van der Stede, 2006). Such studies are able to examine more complex research questions, without sacrificing the *richness* or complexity of the natural environments. The process of carrying out qualitative research also has a high degree of flexibility; in most qualitative studies, the research questions and coding schemes can change during the research process (see Berg, 2003). Researchers can also select which samples to present at the end of their data collection period, rather than in the beginning (Kerlinger and Lee,

2000). While these characteristics of qualitative studies help strengthen the research, these same characteristics can also represent methodological challenges.

Qualitative Research in General

One challenge to qualitative research is in mitigating the potential challenge of *researcher bias*. Whereas bias can affect research design in all methods, qualitative research methods are particularly prone to this challenge due to the potentially subjective nature of data collection and analysis. However, qualitative researchers argue that if the qualitative researcher is a well-trained observer, then, “If done properly, the data collected from qualitative research can yield more information and less spurious variability than other research methods” (Kerlinger and Lee, 2000, p. 590). Thus, the training of the researcher to function as unobtrusively as possible and the disclosure by the researcher of all of his or her assumptions can help to mitigate this concern. As an example, many qualitative researchers will quote liberally from their interview notes or consulted text in order to assemble evidence in support of propositions (e.g. Greenwood et al. 2002).

Related to the potential of researcher bias is the concern that qualitative data requires much more *interpretation* on the part of the researcher. As an example, Anderson-Gough et al. (2005) used in-depth interviews of audit trainees to help understand the embedding of gender relationships in CPA firms. Their analysis required considerable interpretation in the developing of themes (e.g. temporal commitment) and the assignment of portions of interviews with informants to these themes. The challenge of interpretation can be addressed by using the technique of “analytic interpretation” that attempts to find a negative case within the set of hypothesized relationships (Merchant and Van der Stede, 2006), or to assume initially that there are no relationships among the data in order to evaluate emerging patterns with an open mind

(Berg, 2003). Problems with interpretations of construct measurement can also be mitigated by using multiple coders of the study's data.

Explaining and predicting causal relationships with the use of qualitative research is difficult. Whether this is perceived as a weakness of the method depends on the researcher's assumptions. Many researchers, particularly those from a functionalist paradigm, view the difficulty in investigating causal relationships in qualitative studies as a serious deficiency, since the data examined in most qualitative studies is too complex to garner cause and effect relationships (Shadish et al. 2002). However, other researchers do not see qualitative research as striving to explain or predict relationships among phenomena, but see its goal as helping to explain and understand constructs, such as the nature of the "public interest" (Baker, 2005). Researchers wishing to explain causal relationships would probably be better advised to use an experimental method.

From a pragmatic standpoint, qualitative research entails numerous *ethical issues*; confidentiality and acting with high ethical conduct are particularly important when the researcher is personally involved with the study's participants (Kerlinger and Lee, 2000). Conducting qualitative research is also extraordinarily *time intensive*. Some research may take years solely to gather data; for example, Broadbent and Laughlin (1997, 1998) conducted in-depth investigations of physicians' and schools' responses to changing accountability requirements, and collecting interview data initially took several years to complete. Furthermore, researchers often need organizational contacts in order to obtain access to certain types of information, such as historical records (e.g. Chua and Poullaos, 2002) or interviews with audit partners (e.g. Greenwood et al. 2002). This type of extensive organizational cooperation may not always be forthcoming.

Accounting research that uses qualitative methods generally employs either field studies or content analysis of secondary narrative documents. In addition to the general concerns outlined above, each of these specific methods contains its own unique set of challenges.

Field Studies

A field study is similar to a field experiment, but does not employ randomization, manipulation, treatment groups, control or any other elements that are similar to laboratory experiments. Merchant and Van der Stede (2006, p. 118) specify that field studies entail the “in-depth study of real-world phenomena through direct contact with organizational participants.” Kerlinger and Lee (2000) treat field studies as involving non-experimental methods that seek to discover relationships and interactions present in real social structures. Both of these perspectives emphasize that field studies are *naturalistic* investigations in which the researcher acts as an observer and investigator. Field research is particularly helpful in understanding the *context* in which events occur. Within accounting, the use of this method has increased in the past few decades, but its use is still primarily confined to managerial accounting (Merchant and Van der Stede, 2006). This type of investigation can be particularly helpful in building theory.

In addition to their unique strengths, field studies also face several unique challenges. One is the process of *field selection* (sample selection). Researchers must carefully choose fields that accomplish the goals of the study, such as choosing a mix of organizations that have experienced both success and failure in the implementation of enterprise systems (e.g. Nicolaou, 2004). Since field studies often use analyses of a small number of organizations, proper field selection is critical.

The next challenge deals with *generalizability* and external validity. Qualitative field studies tend to use much smaller samples, but study the chosen samples in much greater depth in

order to generalize the study's findings to *theory*. This type of analysis permits the user to gain a richer understanding of a particular field's context, but potentially sacrifices the ability to generalize the findings to other contexts. For example, Seal et al. (2004) used a case study to attempt to theorize about the inter-organizational role of the supply chain, but only used a single firm for analysis. This limits the ability to apply findings from the study to subsequent works.

Another specific limitation of field research is that within North America, many accounting academicians are not well trained in how to conduct quality field research. In particular, most novice field researchers do not tie in theory or prior accumulated knowledge in their analysis of results (Merchant and Van der Stede, 2006). This limitation can be overcome with specific training in field research techniques.

Finally, the richness in a field study may make it difficult for a researcher to consider all of the possible variables and factors that may affect an outcome. For example, a field study on interorganizational networks may provide rich detail that networks tend to function like enterprises, but cannot precisely specify all of the factors that may influence this result (e.g. Mouritsen and Thrane, 2005). The goal of most qualitative research is therefore to understand, rather than to predict and explain (Llewelyn, 2003).

Content Analysis

Another common type of qualitative method involves the content analysis of secondary, non-numerical communications. This type of method is unobtrusive, cost-effective, and particularly useful for studying long historical trends or general societal changes (Berg, 2003). For instance, Chua and Poullaos (2002) used historical documents to study the relationship between professional associations at the center and periphery of empire in the 19th century.

Like all other research methods, content analysis has its drawbacks. According to Berg (2003, p. 288), “The single serious weakness of content analysis may be in locating unobtrusive messages relevant to the particular research questions.” This is particularly a challenge when content analysis is used as the primary research method. The challenge of *locating unobtrusive messages* arises because the documents analyzed have already been recorded and do not contain data pertaining to the initial research question. Therefore, performing a content analysis on research questions such as the history of professional associations in Trinidad & Tobago (Annissette, 2000) requires the researcher to extensively search documents in order to develop themes related to the research question.

Secondly, content analysis is “virtually useless” in investigating causal relationships (Berg, 2003, p. 288). A content analysis of public documents is useful for carrying out an exploratory or descriptive study; despite the prevalence of themes that might be shown, however, researchers should not assume that causality has been established. This challenge is not easily mitigated; therefore, researchers should not use content analysis if the aim of their study is to test a causal relationship.

Overall

Overall, qualitative research relies upon the researchers’ examination of naturalistic environments in order to understand the context in which accounting phenomena occur. This presents the researchers with flexibility and enhanced external validity, but must be balanced against researcher biases and the inability to test causal relationships. When used in conjunction with archival and behavioral methods, qualitative methods can provide a richer picture of the setting’s context and environment.

EARNINGS MANAGEMENT: TRIANGULATION OF RESEARCH METHODS

Archival, behavioral, and qualitative research methods all have limitations in their ability to explore research questions. Archival research methods are useful in explaining general large-scale phenomena, behavioral methods are well-suited for testing causal inferences, and qualitative methods afford a deeper understanding of context. Thus, the limitations inherent in a particular method may be addressed by using one or more complementary methods. For instance, qualitative and archival research methods cannot adequately assess causal relationships, whereas behavioral experiments can, due to the higher internal validity afforded by this method. Behavioral experiments may have limited external validity, however, while archival studies are useful for examining naturally occurring phenomena and qualitative research methods afford a rich examination of context. Hence, the use multiple methods to examine common research questions is more robust than a single method in isolation.

The lack of superiority of a single research method points to the need for *triangulation*. Simply stated, triangulation of research methods involves the use of multiple research methods to investigate a research question.³ Triangulation can help researchers improve the *generalizability* of findings, while presenting a *richer* picture of a particular phenomenon. Several researchers have called for the joint use of in-depth field studies with cross-sectional surveys to improve both theory and empirical testing (e.g. Arnold, 2006). Such a multi-method approach – particularly a mixture of qualitative and quantitative methods – helps to overcome the weaknesses of using a single research method in isolation (Kerlinger and Lee, 2000). This could best be accomplished by fostering a dialogue across different studies, such that synthesizing multiple papers relying on multiple methods to investigate a common phenomenon could provide a richer, panoramic view of the topic in question.

The advantages of triangulation are evident if researchers' goals are to understand a broad phenomenon. One area of research in accounting, *earnings management*, provides an interesting lens through which to view the usefulness of triangulation. Earnings management is defined by Schipper (1989) as "purposeful intervention in the external reporting process, with the intent of obtaining some private gain to managers or shareholders." Most of the archival literature on earnings management has sought to *identify* instances of earnings management, and ultimately to *explain under what conditions* firms manage their financial earnings (Koonce and Mercer, 2005). Overall, archival studies on the earnings management phenomenon have failed to provide convincing evidence that managers explicitly misstate earnings. Part of the reason for this lack of support is due to the inherent limitations of archival research. Since archival methods use secondary-databases (none of which contain a convenient "earnings management" dataset), researchers have had to identify proxies to test for instances of earnings management. While a variety of models have been developed to proxy for earnings management, all of these demonstrate low power and an inability to properly distinguish between true performance and opportunistic behavior (DeChow et al. 1995). Some newer models have made improvements (e.g. Francis et al. 2005), but archival researchers are faced with an uphill battle in solely using secondary data to study earnings management if they cannot determine when such behavior exists.

A second limitation of archival studies of earnings management is that it has been very difficult to tap into the *motivation* involved with this behavior (Schipper, 1989). Archival studies have used economics-based theory to provide a number of explanations for why earnings management may occur, including political reasons (e.g. Jones, 1991), bonuses (e.g. Watts and Zimmerman, 1986), satisfying debt covenants (e.g. DeChow et al. 1996), or optimally timing

equity transactions (e.g. Kothari, 2001). However, since archival studies cannot easily establish causal inferences, it is difficult to conclude whether firms with these characteristics engage in earnings management because of the above-mentioned motivations, or whether another correlated omitted variable is responsible for the connection. Likewise, the motivations used in archival studies of earnings management ignore the nuances of human behavior that may also be responsible for such actions (Koonce and Mercer, 2005).

Behavioral studies of earnings management have complemented the findings in the archival literature by addressing some of these shortcomings and provide evidence that earnings management behavior is very prevalent. Whereas archival studies have failed to convincingly demonstrate that managers overtly manage earnings, behavioral studies' use of individual participants provides stronger support for the occurrence of this behavior. Synthesizing the results of archival and behavioral studies can therefore help overcome the limitations of relying upon studies of a single method. Part of the reason for this difference in findings in the archival and behavioral research method streams is due to the varying strengths of these two methods. From a behavioral perspective, "earnings management" behavior has stronger construct validity (i.e., is more easily defined and measurable) than from an archival standpoint. However, behavioral methods are not immune from threats to construct validity. In laboratory studies, researchers must design a task for participants to complete, and the operationalization of "earnings management behavior" may prove similarly challenging.

Behavioral studies on earnings management have often used *auditors* as participants to help gather information about this phenomenon. One such study is Nelson et al. (2002, p. 184), in which the researchers surveyed auditor partners and managers about "experiences with companies that attempt material earnings management". This study provided evidence on the

effect of the preciseness of accounting standards on earnings management behavior among both managers and auditors. Specifically, auditors identified that managers were most likely to attempt earnings management behavior by either structuring transactions to satisfy very precise accounting standards (e.g. reserves), or by leaving transactions unstructured to satisfy imprecise standards (e.g. leases); auditors reported they were also less likely to require financial statement adjustments of such behavior. The behavioral method in Nelson et al. (2002) results in stronger construct validity than would attempts to develop an empirical proxy of earnings management, which could be why the study provided stronger evidence of earnings management behavior than seen in the archival literature.

Another manner in which behavioral studies can complement the findings of archival studies is through the strong internal validity of a laboratory experiment, which makes this a very useful method for establishing causal inferences. For instance, Lee et al. (2006) conducted an archival study that found that firms with a history of earnings management behavior tended to use less transparent reporting. Hunton et al. (2006) designed an experiment to test the causality of this relationship, and showed that greater transparency of standards helped to mitigate earnings management behavior. These findings demonstrate that archival studies are useful for establishing an overall pattern of behavior among naturally occurring events, whereas behavioral studies can provide evidence on the causality of relationships and are hence able to demonstrate the overt nature of earnings management behavior. Another laboratory experiment in this vein is Beeler and Hunton (2002), which used audit partners as participants and found that auditor judgment was biased by the presence of contingent economic rents. Extrapolating these findings suggests that earnings management behavior may be more common in these circumstances; again, such a causal relationship could not be gleaned from an archival study.

Overall, evidence from behavioral studies demonstrates that purposeful earnings management behavior is indeed common. In particular, evidence from auditors suggests that managers are more likely to manage earnings given fewer opportunities for detection and external constraints. The weaker construct and internal validity found in archival studies is the likely reason why results from behavioral studies provide stronger evidence of earnings management behavior. Thus, utilizing behavioral methods in conjunction with archival research affords a deeper understanding of the topic of earnings management.

Despite the broadened view given by a joint examination of a topic from archival and behavioral viewpoints, both of these methods ignore the *context* within such behavior occurs. The use of a qualitative field study to address this topic could help researchers “drill deeper” into the underlying rationale behind earnings management and provide a better understanding of the findings from the archival and behavioral studies thus far. In the words of Merchant and Van der Stede (2006, p. 129), “...[I]f more researchers would engage in field research and ask management how and why they would manage earnings...then a richer theory would emerge with details about the methods used to manage earnings and multiple relevant behavioral factors.” Archival research tends to truncate outliers; a qualitative field study could focus on extremes of behavior and help to assess how corporate environments and/or individual management factors interacted to exacerbate or constrain earnings management behavior (Merchant and Van der Stede, 2006). Such a qualitative study could involve interviewing managers to determine how they defined and recognized “earnings management,” along with ascertaining its contributing factors. This would help to tie together past archival and behavioral studies on earnings management and develop a more complete picture of the process.

Overall, triangulating research methods in investigating a complex accounting phenomenon such as earnings management affords a richer, panoramic view of the motivations, causality, context, and contributing factors of this occurrence. Despite these potential benefits, however, many issues may arise when researchers attempt to synthesize separate studies that have used differing research methods to investigate a common phenomenon. Two of the most salient issues concern the role of research methods in different stages of investigation, and the reconciliation of conflicting results.

First, an unanswered question is whether all research methods are appropriate for all stages of investigating a phenomenon, or whether specific methods have a distinct role at different phases. For example, in the study of earnings management, archival methods may be useful for investigating patterns of behavior associated with the presence of earnings management, behavioral methods can follow up with more rigorous tests of causality of the purported relationships, and qualitative methods can build upon such work by delving in to more complex ‘how’ and ‘why’ relationships. However, these methods may still have a role at other stages – for example, qualitative methods may be useful in the initial stages of investigation in helping to define a phenomenon’s constructs. Researchers should therefore think carefully about the aims of their study before deciding on a research method.

Second, using multiple research methods across research studies opens the door to the possibility of conflicting results. For example, a behavioral experiment that follows up on an archival study may fail to find the hypothesized causal claim between constructs. Moreover, qualitative field studies may prove that the construct is more complex than initially anticipated, suggesting that prior proxies used to operationalize this construct may have been poor choices. While there is no clear answer for reconciliation, the conflicting results achieved with multiple

research methods could suggest that further research be conducted on the phenomenon – perhaps by re-examining the theory suggesting causality, or by developing better proxies. In all, triangulation of multiple research methods does not guarantee a perfectly aligned picture of a phenomenon, but could lead to advances in incremental knowledge,

CONCLUSION

This paper presents an overview of the limitations of relying on research that exclusively uses archival, behavioral, or qualitative research methods. Triangulation of studies that have used differing research methods will help researchers overcome the limitations of reliance on a single method and achieve a deeper understanding of accounting phenomena. In particular, recognizing the oft-neglected value of behavioral and qualitative research would be beneficial to the academic accounting community, particularly as such methods may help compensate for some of the inherent weaknesses in archival research methods.

Triangulation of multiple research methods has the potential to increase both the generalizability and richness of findings for a given phenomenon. Research methods are the tools researchers utilize to test research questions; the joint use and understanding of multiple types of research methods affords a broader view of appropriate research questions and of accounting research in general. Some evidence suggests that researchers are beginning to realize the benefits of triangulation. For example, the use of an experiential questionnaire enables practitioners to provide in-depth information about real-world phenomena within a structured environment; this method thus provides the richness of an interview (qualitative) with the structure of a questionnaire (behavioral) (Gibbins and Qu, 2005). Unfortunately, the use of multiple research methods is rare, as are attempts to connect studies of the same topic that were investigated with different research methods (Lillis and Mundy, 2005; Merchant and Van der Stede, 2006). This

suggests that the accounting research community has much to gain by broadening its collective paradigmatic lenses (e.g. Lewis and Grimes, 1999).

Triangulating across methods and conducting multi-method research are not without their own limitations. A broader view of research entails much larger time and resource commitments for researchers to become familiar with studies using different methods, not to mention the huge start-up costs researchers must invest in learning how to use different research methods themselves. Due to the doctoral education process, much research is conducted in “silos” that may be difficult to eradicate overnight (Koonce and Mercer, 2005). Moreover, triangulating studies that have used differing research methods to examine the same research question may produce conflicting results that point the way for further work. Nevertheless, being aware of multiple methods and recognizing the value and role of each are important early steps in broadening the research process. This awareness is critical so that researchers can look beyond their “native” research stream in helping to motivate their own studies. While it may not be feasible for researchers to conduct research using different types of methods, they should be open to incorporate the results gained from other research methods in order to better understand the holistic nature of a research phenomenon.

In conclusion, accounting is essentially an inter-disciplinary area formed at the crossroads of several root disciplines and research methods. To borrow from Slemrod (2003), an attempt to study accounting phenomenon through the singular lens of a particular research method is like the parable of the elephant and the blind men – each are incapable of recognizing that they are faced with an elephant due to their preoccupation with a narrowly defined realm. By breaking down the barriers between research methods, accounting researchers can attempt to examine the “larger creature” behind accounting phenomena.