Points to consider when self-assessing your empirical accounting research

Points to Consider When Self-Assessing Your Empirical Accounting Research*

JOHN HARRY EVANS III, *Katz School of Business, University of Pittsburgh*

MEI FENG, *Katz School of Business, University of Pittsburgh and Visited at Cheung Kong Graduate School of Business*

VICKY B. HOFFMAN, *Katz School of Business, University of Pittsburgh*

DONALD V. MOSER, *Katz School of Business, University of Pittsburgh*

WIM A. VAN DER STEDE, *London School of Economics*

1. Introduction

Converting an initial research question into a thoughtful and carefully designed empirical accounting study is a complex, multifaceted process. Certain broad issues, particularly concerning the study's contribution and research design, arise repeatedly in seminars and review reports. This suggests that researchers can benefit from anticipating these issues and systematically self-assessing whether their study addresses them effectively. The Appendix provides a list of points to consider (PTCs) when conducting empirical accounting research. Figure 1 provides readers with a "big picture" summary of the main topics covered in the list. The text describes a set of more challenging issues on the list, offers suggestions for how to address them, and provides illustrative examples.[1] Our goals are to help accounting researchers, especially those who are less experienced, improve the quality of their research and more clearly communicate how they address key issues. Although improved quality and communication will not ensure publication, it can help readers better appreciate a study's contribution and increase the likelihood of publication.

Our paper is intended to help researchers in two ways. First, our list of PTCs serves as a decision aid to help researchers anticipate key questions and self-assess whether they have adequately addressed these issues. Accounting research requires extensive knowledge and skill, with researchers typically working on several projects at a time over a number of years. One implication of this complex, multifaceted process is that researchers can easily overlook or inadequately address key components of a project. Our PTCs can help researchers review important items systematically to compensate for limitations in their memory and attention. Second, our discussion offers suggestions for how to address some

* Accepted by Michel Magnan. We thank Jongwoon (Willie) Choi, Michele Frank, Weili Ge, Jennifer Joe, Steve Kachelmeier, Bill Kinney, Sarah McVay, Bryan Stikeleather, PhD students at the University of Washington, workshop participants at the University of Pittsburgh, the 2012 ABO Doctoral Consortium, and the 2012 AAA Annual Meeting for their helpful comments and suggestions. We are also grateful to Steve Salterio and Michel Magnan, as well as three anonymous reviewers, for their suggestions and advice. We appreciate the research assistance of Eric Chan and Amy Yurko.

1. To demonstrate specific applications of the concepts on our list of PTCs, we refer to a variety of accounting studies throughout this paper. We have chosen these particular studies because they address topics familiar to most empirical researchers and because we believe that the examples chosen are good illustrations of the issues we are discussing. The choice of studies does not reflect any assessment of the quality of the research itself.

Contemporary Accounting Research Vol. 32 No. 3 (Fall 2015) pp. 1162–1192 © CAAA
doi:10.1111/1911-3846.12133

Figure 1 Summary of list of points to consider (PTCs) when self-assessing your empirical accounting research

1. Research Question: What is your precise research question?
   Be clear in your mind and in your manuscript concerning:
   a. The meaning of key terms in the research question
   b. The scope of the research question
   c. Whether the research question involves a causal relation or an association
2. Theory: Is your study's conceptual framework logically consistent and credible?
   a. Develop your conceptual framework by applying theoretical arguments and/or previous empirical findings to a specific setting
   b. Explain how your study distinguishes among alternative explanations for your predictions, or discuss why it is not important to do so in your setting
3. Contribution: What does your study add to our understanding of accounting?
   a. Establish how your study is new and interesting relative to the prior literature
      i. The topic itself is interesting and important
      ii. The study is differentiated from and extends prior literature
   b. Identify important implications in terms of actions or beliefs that will change based on your study
4. Research design and analysis
   a. Research design and analysis for *Archival Studies*
      i. Select the appropriate sample, proxies, and empirical models
      ii. Consider alternative explanations
      iii. Conduct tests to support your theory and rule out alternative explanations
   b. Research design and analysis for *Experimental Studies*
      i. Consider *in advance* which design and operationalization of variables provide the best test of your theory and help rule out alternative explanations
      ii. Consider potential mediating or moderating variables that would be consistent with your theory and inconsistent with alternative explanations
      iii. Consider incorporating a predicted interaction to control for potential omitted variables
      iv. Conduct tests to support your theory and rule out alternative explanations
   c. Research design and analysis for *Field Studies*
      i. Establish field data reliability
      ii. Ensure the validity of the field study inferences
   d. Research design and analysis for *Survey Studies*
      i. Match the design of the survey with the purpose of the study
      ii. Establish survey data generalizability
      iii. Ensure the internal validity of the survey study inferences
5. Interpretation of results and conclusions
   a. Interpretation of empirical results
      i. Describe the statistical significance, and the economic magnitude or effect size of the results, if applicable
      ii. Discuss and justify the pattern and magnitude of the results based on your story and findings in prior studies
      iii. If data support only part of the predicted pattern, or if different analyses lead to different statistical inferences, consider possible reasons for this to evaluate the reliability of your findings
   b. Conclusions
      i. Relate your conclusions back to your motivation and research question
      ii. Avoid over-generalizing or over-concluding
      iii. Discuss the contribution and implications of your finding

of the more challenging issues on the list, which we expect to be most beneficial to less experienced researchers.

Our list of PTCs has five main sections. Because all successful research projects must be both conceptually sound and empirically valid, the first three sections address conceptual issues and the last two sections address empirical issues. The three conceptual sections are the same for all types of studies, and offer suggestions about (1) refining a preliminary research question and communicating the question clearly; (2) developing the study's "theory" and associated logic;[2] and (3) evaluating and communicating the study's contributions and implications. The fourth section includes separate subsections regarding the research design and analysis of archival, experimental, field, and survey studies because these different research approaches have some unique issues.[3] The fifth section focuses on the interpretation of results and conclusions.

Our PTCs can provide useful self-assessment at various stages of a project, and can help authors clarify their thinking, which is a precondition to communicating their ideas effectively to others. For example, when developing an initial idea, the PTCs can help researchers refine the idea and evaluate the likelihood that the study will ultimately provide convincing empirical results and make an incremental contribution. Later, the list can help researchers self-assess whether their manuscript adequately addresses fundamental questions that workshop participants are likely to have. This should help workshop participants focus more effectively on other substantive issues specific to the study. Researchers can consider the entire list or focus on specific issues of concern, either individually or as a team. Using the list as a team can potentially improve communication among coauthors by adding structure to the process.

Previous accounting articles have discussed how to ask better research questions (Maines, Salamon, and Sprinkle 2006); design better empirical studies (Kinney 1986, 2011); write more readable manuscripts (Zimmerman 1989); produce better reviews (Kachelmeier 2004); and increase the likelihood of publishing manuscripts (Brown 2005). Although doctoral students and emerging accounting scholars will find each of these articles useful, none is likely to fully prepare them for their first research presentation or journal referee report in which they are asked seemingly fundamental questions that they may never have considered, or that they once considered but failed to include in their study. The purpose of this paper is to help accounting researchers systematically analyze the important aspects of a good empirical accounting study. The paper identifies a set of key issues that arise often in empirical accounting studies and offers potential reasons why researchers often overlook such issues. We note that many of the same issues arise in archival, experimental, survey, and field studies, but also discuss differences across these research methods. Examples from published research are used to illustrate how critical issues have been effectively addressed in prior studies. Specifically, our paper identifies and elaborates on ways to (1) clearly specify a research question; (2) develop a credible and logically consistent "theory" or "story"; (3) identify a study's contribution; (4) design a

2. We use the term "theory" to refer to a study's general theoretical structure or conceptual framework. While this could be a formal mathematical model, we also include in this notion less formal appeals to existing theories, results of prior empirical studies, or even simply a set of logically consistent, integrated, and persuasive arguments for addressing new questions. What we refer to as theory could in some cases be referred to more colloquially as the study's "story."

3. Our paper seeks to offer practical guidance as opposed to advocating a position regarding more fundamental debates regarding how the philosophy of science relates to accounting as a social science (e.g., Christenson 1983; Chua 1986; Hines 1988; Watts and Zimmerman 1990). Our paper has an inherent positivistic approach. Therefore, with regard to qualitative research, we include studies with a positivistic orientation that use qualitative methods such as field studies. However, given our positivistic orientation, we are unable to comment meaningfully concerning other types of qualitative research that employ a different ontological set of assumptions, such as papers in the interpretivist and critical research traditions.

study to test the research question(s); and (5) interpret the results and draw appropriate conclusions. Thus, our list of PTCs complements previous work by organizing many critical components of the research process in a structured and concise way.[4]

2. Accounting research "list of points to consider"

The three conceptual sections of the list emphasize the importance in empirical accounting research of addressing the following primary issues: What is the research question? What is the theory? What is the contribution? We next describe how the PTCs address each of these questions.

1. Research Question: What is your precise question?

Items 1.a–c on the list indicate that specifying the research question precisely requires researchers to be clear in their own minds and in their manuscript regarding the definition of key terms, the scope of the study, and the assumed relation among concepts.[5] The research question should be stated explicitly and precisely in the introduction of the study. If the research question is vague, readers and reviewers will likely adopt ideas about the question that differ from what the researchers intend. Stating which issues the study does and does not address, and why, helps readers better understand and appreciate the study's key points. A clearly defined research question also helps researchers design their study and differentiate it from prior studies.

1.a. The meaning of key terms in the research question

The accounting literature includes many broad terms that should be defined precisely, such as earnings quality, auditor independence, and optimal contracts.[6] For example, Dechow, Ge, and Schrand (2010) argue that "earnings quality" can be defined only in the context of a specific setting because this concept has different meanings for different information users, and has been operationalized in various ways in prior research (e.g. earnings response coefficients and discretionary accruals). As a result, researchers analyzing earnings quality should define the precise meaning of this concept and explain how it is operationalized for purposes of their study.

A second example of the importance of defining key terms concerns alternative definitions of "optimal contracts." In an analysis of the form and efficiency of executive contracts, Bebchuk and Fried (2004) adopt a "managerial power" perspective of corporate governance and use "arm's-length bargaining" between the executive and a board as their benchmark for "optimal contracting." They then provide evidence that executives routinely exert significant influence over the board makeup and board decisions, including the design of executive compensation contracts. Using their benchmark, they conclude that the resulting executive compensation contracts, and corporate governance in general, are suboptimal. In contrast, Core, Guay, and Thomas (2005) argue that "optimal contracts" minimize the firm's overall agency costs, not just executive compensation costs. Therefore,

4. Although there are research decision aids in other areas such as psychology (e.g., Rogelberg 2005), we are not aware of any that target issues specific to accounting research. Because the psychology decision aid is tailored to psychology research, it includes some items irrelevant to accounting and excludes many items important to accounting researchers.
5. Our list of PTCs is designed to help researchers carefully *consider* each point so that they can subsequently *communicate* it clearly in the manuscript. Therefore, when we discuss the need to "specify," "state," or "explain" something explicitly in the manuscript, we mean that researchers should first clarify the issue in their own minds, and then express that clear thinking to the reader in the manuscript.
6. Typically, studies that identify an entirely new phenomenon, as opposed to these examples, will provide an explicit definition to specify the boundaries of the new issue being addressed.

shareholders can benefit from arrangements that allow the CEO to influence the board, for example, by including insiders on the board. Because informed insiders can facilitate such decisions as project selection, the benefits of having insiders on the board may outweigh the associated costs related to executive compensation contracts. The different definitions of "optimal contracts" in the two approaches lead to significantly different conclusions concerning how to improve corporate governance.

1.b. The scope of the research question

No study can examine all aspects of an issue, and therefore it is important to decide on the appropriate scope of the study.[7] Researchers should consider how their choice of scope affects the focus, validity, and contribution of the study. Here we discuss the focus and validity dimensions, deferring the issue of contribution until the third section of our list.

Regarding focus, the conventional length of accounting research papers constrains how broadly an issue can be addressed effectively. A study that attempts to cover too much is often unable to provide compelling empirical evidence about all of those issues. For example, one study might investigate the choice and relative weight of performance measures in incentive contracts, while another study examines the use of external benchmarks in setting executive pay. Most studies can effectively address only one or two key issues, but referees or workshop participants frequently suggest ways to expand a study's scope. Researchers should think carefully about whether expanding the scope will provide significant new insight regarding the main issues or simply add one more issue. The former is often a good idea, whereas the latter could dilute the focus of the paper and detract from the main issues.[8]

The scope of a study can affect the validity of the conclusions. The rationale for excluding an issue usually is that its effect is less important or that it can be treated as exogenous; that is, determined outside the study.[9] However, assuming that a particular aspect is exogenous could threaten the validity of the study. For example, Francis and Schipper (1999) investigate the changing relevance of financial statements to investors by examining the association of stock returns with earnings and balance sheet information. They find that between 1952 and 1994, earnings became less value relevant and balance sheet information became more value relevant. Thus, if they had limited their scope to only the relation between stock returns and earnings or only the relation between stock returns and balance sheet information, they would have reached opposite conclusions about how the relevance of financial statements has changed.

1.c. Whether the research question involves a causal relation or an association

Many accounting research questions imply causal relations rather than associations, where a causal relation represents a stronger and more definitive finding. For example, concepts A and B could be associated because A causes B, because B causes A, or because

7. Examples of how the scope of a study might be defined in different ways include the following: (1) A causes B, B causes C, and C causes D, but the study only addresses the relation between B and C; (2) A is associated with B, C and D, but the study only addresses the relationship between A and B; (3) players A, B and C are involved in the research question, but the study only addresses the roles of players A and B; and (4) a key player needs to make multiple decisions, A, B and C, but the study only addresses decision A.
8. This discussion also applies to questions regarding when to conduct additional empirical analyses or to run other experiments.
9. The study can also incorporate an exogenous factor as a control variable rather than as a concept to be explained or determined within the study.

C causes both A and B. In most archival and field studies, the simultaneous and endogenous nature of the data limits the ability to draw causal inferences. Experiments are better suited to establishing causality because the independent variable (A) can be manipulated before measuring the dependent variable (B), while holding other factors constant. Conversely, the limitation of experiments is that other factors in the natural environment that are not measured, manipulated, or controlled for in the experiment could offset or amplify the effect of A on B in the field. Researchers must carefully evaluate whether their research design supports causal inferences, and clearly specify the nature of the conclusions they reach (Ittner 2014; Van der Stede 2014).

2. Theory: Is your study's conceptual framework logically consistent and credible?

The second section of the list focuses on the study's theory, logic, or *story*, by which we mean the conceptual framework within which the study's research question is analyzed. Studies that document statistical associations without a compelling conceptual framework for establishing how these associations are related typically have limited impact on the literature because they are often viewed as less reliable. To develop a compelling conceptual framework, researchers need to apply and synthesize prior theoretical and empirical findings consistently and credibly (item 2.a), and to differentiate their "story" from alternative explanations (item 2.b).

2.a. Develop your conceptual framework by applying theoretical arguments and/or previous empirical findings to a specific setting

A study's conceptual framework often draws on a well-established theory from prior literature. However, in some cases it may be based on other forms of logical reasoning that are not yet sufficiently developed to qualify as a theory, but can provide a reasonable basis for an empirical study. Common inputs to such studies would include systematic findings from prior empirical work, adaptations or components of well-established theories, or even intuitive and logical arguments.[10] Regardless of the basis for the study, conceptual coherence is crucial for a study's credibility.

2.a.i. Ensure that your literature review is thorough enough to identify all major theories related to your story that lead to the same or opposite predictions

It is important to clearly explain how existing theory and evidence relate to the research question, rather than simply summarizing them. The literature review can then flow naturally into an explanation of how the current study differs from prior literature and how any differences represent a contribution to the literature, as we discuss further in the Contribution section.

2.a.ii. Consider whether your theory holds when analyzed from all key actors' points of view

The theory's key actors should behave in a consistent manner unless a compelling explanation is offered for why certain actors may behave differently. For example, in a study in which managers make two types of decisions, if the managers are subject to cognitive bias X in making the first decision, then the same effect should influence their

10. For example, Sloan (1996) documents accrual anomalies arguing that, while the accrual component of earnings is less persistent than the cash flow component of earnings, investors appear to fixate on earnings, failing to distinguish fully between the differential persistence of the accrual and cash flow components of earnings. As a result, firms reporting a high accrual component of earnings are systematically overvalued and firms reporting a high cash flow component of earnings are systematically undervalued. Sloan's "story" combines intuitive arguments based on market efficiency and on individual biases and heuristics.

second decision unless the study explains how differences in the second decision's context eliminate the bias. Similarly, in a study of managers and analysts, if the managers are assumed to be sophisticated rational decision makers, then analysts should be too, unless the study offers a rationale for a distinction.

2.a.iii. Ensure that all hypotheses/predictions in your setting are consistent with each other

The study's hypotheses and predictions should reflect consistent assumptions. For example, if the study assumes that a firm's competitors can anticipate the firm's strategic behavior along one dimension, then the study should make similar assumptions with respect to other dimensions of strategic interaction.

2.a.iv. Specify the mechanism by which the concepts in your setting relate to each other

It is important to explain the mechanism through which, or *how*, one event, decision or action is related to another event, decision, or action. Specifying the details of such mechanisms by documenting how they actually operate in practice provides readers with a more complete appreciation for the study's logic and should make the study more persuasive. One way to acquire such practical knowledge is to talk with relevant practitioners. For example, researchers studying the role of accounting data in bond-ratings agencies' decisions may gain considerable new insight from discussions with rating agency analysts to confirm or disconfirm important aspects of the study's logic. Another way to develop a more complete and persuasive conceptual framework is to conduct formal survey research, case studies or field research.

A study by Campbell, Epstein, and Martinez-Jerez (2011) provides an example of a field study that relies on interviews and observation to provide a compelling basis for how the tightness of the control system influences employee learning and firm profit. To understand the role played by gaming resort employees who act as "hosts" for customers at the resort, the researchers interviewed resort employees from multiple locations and shadowed "hosts" for several days. Hosts must learn how to achieve the proper balance between maximizing current resort profit versus sacrificing some current profit to build customer relationships that will increase the customer's contribution to future resort profit. Hosts must decide the extent of complimentary benefits (comps) to provide to each customer during a given resort visit. Allowing more comps builds goodwill with the customer, but reduces the resort's current profit.

Hosts base their comps decisions on a combination of historical data on the resort's profitability from the mix of games each customer plays and on their subjective judgment from interacting with the customer during each visit. To monitor the level of comps awarded by individual hosts, individual resort properties use different management control systems, which vary from very tight to quite loose monitoring systems with less frequent reviews and meetings and far weaker threats of termination for "abnormal" comps percentages. Campbell et al. (2011) find that employees learn best under the looser control systems by "experimenting" with a variety of levels of comps and noting which levels are most effective with particular customers. This, in turn, allows the hosts to improve the resort's lifetime profit from that customer. Conversely, hosts who operate under the tighter control systems learn to apply more rigid standardized rules to each customer, which increases current profit, but reduces future profit.

2.a.v. Specify additional effects one should or should not observe if the theory is correct

Specifying additional implications can significantly increase a study's credibility. For example, researchers can identify settings in which their theory is more or less likely to

hold (item 2.a.v.1); specify and document additional consequences of their theory (item 2.a.v.2); or analyze the path by which the independent variable affects the dependent variable by examining intervening effects (item 2.a.v.3). By identifying and testing such additional effects, researchers can sharpen and enrich their study, and more effectively rule out alternative explanations.

Bernard and Thomas's (1990) post-earnings-announcement drift study provides an example of items 2.a.v.1 and 2.a.v.2. They propose that investors naively believe that quarterly earnings follow a seasonal random walk when the earnings actually follow a more complicated time-series pattern. This specific characterization of investors' beliefs enables the authors to develop predictions about the timing, magnitude, and sign of market reactions subsequent to earnings announcements. Specifically, the authors make a cross-sectional prediction that the post-earnings-announcement drift in stock returns will be concentrated in three-day windows around the subsequent earnings announcements (item 2.a.v.1). They also predict that stock returns will drift in the same direction as the original earnings news and that this drift will decrease in magnitude in the subsequent three quarters. Then in the fourth quarter the drift will reverse sign and be in the opposite direction of the original earnings news (item 2.a.v.2). Bernard and Thomas find support for each of these predictions, which significantly increases the credibility of their explanations. Moreover, the specificity of their predictions and their supporting evidence make it difficult to generate convincing alternative explanations for their findings.

2.b. Explain how your study distinguishes among alternative explanations for your predictions

Identifying and analyzing alternative explanations for the study's results enhances the credibility of the findings. This requires determining which predictions of the study's maintained theory are also predicted by an alternative theory, and which predictions differ between the two. Even if it is not possible to completely distinguish between competing explanations, it is important to acknowledge plausible alternative explanations and to discuss the extent to which it is critical to distinguish between the researchers' explanation and these plausible alternative explanations for the study's findings.

3. Contribution: What does your study add to our understanding of accounting?

The contribution of a study is critical because it determines the study's impact on the field. Therefore, it is important to develop and clearly articulate the study's contribution. Doctoral students and junior faculty members often find this difficult. In part, this may reflect the intense competition for new and important research insights, but it also may reflect the greater emphasis on validity in most doctoral programs. Deciding what constitutes a significant contribution requires judgment that can take years to develop. However, readers are more likely to agree that a study makes an important contribution when the researchers clearly articulate the contribution rather than leaving the reader to infer the contribution from the study's results.

Contribution is typically assessed in terms of what is new and interesting given the prior literature (item 3.a), and what are the implications of the study (item 3.b). Sometimes addressing a single PTC item effectively may be enough to make a meaningful contribution, but studies that contribute on several dimensions are more likely to convince readers of the paper's merit. Researchers should draw on the study's conceptual framework in section 2.a as an important source of ideas for the study's contribution.

3.a. Establish how your study is new and interesting relative to the prior literature

Readers might find a study interesting if they consider the topic itself to be interesting (item 3.a.i), or because the study extends prior literature in a new and important way (item 3.a.ii).

3.a.i. The topic itself is interesting and important

Items 3.a.i.1–4 of the list operationalize the concept of an "interesting" study by noting that a topic could be viewed as interesting because it relates to a pervasive phenomenon, an emerging trend, a controversial regulatory issue, an economically significant phenomenon, a fundamental accounting question, or a critical factor related to financial or managerial reporting, auditing, tax, or other aspects of an organization's performance.

Barton and Mercer (2005) is an illustration of how providing new insight about a critical factor related to an essential accounting process, namely, the credibility of management's financial reporting disclosures, can make a study interesting (item 3.a.i.4). Their experiment extends previous disclosure literature regarding managers' tendency to blame poor performance on external factors. They show that when management offers a plausible explanation for poor performance, analysts forecast higher earnings than when management offers an implausible explanation or does not offer any explanation, suggesting that management's explanations can backfire. Rather than viewing implausible explanations favorably or ignoring them as "cheap talk," financial analysts view them negatively, and consequently lower their earnings forecasts and anticipate a higher cost of capital than if management had not offered any explanation.

3.a.ii. The study is differentiated from and extends prior literature

A study can extend the prior literature in several ways (items 3.a.ii.1–3). For example, it could expand our knowledge of a previously studied issue by examining the issue from a new perspective or by identifying a new setting, method, or theory (item 3.a.ii.1). Alternatively, a study could reconcile previously mixed results or resolve a puzzle identified in the prior literature (item 3.a.ii.2). In addition, tension can arise when different theories or arguments make conflicting predictions. Studies with greater tension are typically judged as making a greater contribution because they answer questions for which the answer is less predictable (item 3.a.ii.3).

Bamber, Jiang, and Wang (2010) is an example of a study that effectively communicates how it extends the prior literature in a new and important way (item 3.a.ii.1). The authors note that the theories used in most prior research posit little or no role for manager-specific characteristics or preferences to affect corporate reporting. Bamber et al. (2010) draw on an alternative management theory that allows top managers' personal values and styles to influence their reporting choices. Given the relatively compelling empirical evidence they provide in support of this innovative perspective, the study identifies a new determinant of disclosure that will likely influence future disclosure research.

3.b. Identify important implications in terms of actions or beliefs that will change based on your study

Beyond what is new and interesting, item 3.b indicates that a study can make a significant contribution by demonstrating that it has important implications. Readers often ask, "Who cares?" about the answer to a research question. That is, whose beliefs or actions would be affected if they knew the results of the study? A study can make a contribution if it has important implications that help individuals improve their investment decisions;

regulators better understand the implications of an existing or planned regulation; or managers design better incentive systems. Alternatively, a study may be primarily of interest to researchers. The contribution of this latter type of study requires careful evaluation, but is most likely to make a contribution if it affects the nature or direction of future research.

Frederick and Libby (1986) is an example of a study that makes an item 3.b contribution because it had important implications for future research. Before this study, researchers expected more experienced auditors to perform better than less experienced auditors, but experiments found mixed results regarding the effect of experience on audit judgments. Frederick and Libby's main contribution was to demonstrate that to identify performance differences between experienced and inexperienced auditors researchers need to consider when auditors acquire certain types of audit knowledge. Then the researcher can match the auditors' level of knowledge with the level of specific knowledge needed to perform the experimental task. If a task requires only a novice's knowledge, both experienced and inexperienced auditors should perform well and no experience difference would be found. This insight helped explain why previous auditing studies that did not specify the level of requisite knowledge were unable to consistently document differences in performance based on experience. While this study was not designed to have immediate implications for audit practice, it quickly changed the way researchers approach the study of expertise, subsequently enabling researchers to better answer many questions important to audit practice.

4. Research design and analysis

Researchers who have carefully thought through their research question(s), theory, predictions, and potential contributions will be more effective in their research design and analysis. We next discuss research design and analysis for archival, experimental, field, and survey studies separately because of the distinctive features of these research methods.

4.a. Research design and analysis for archival studies

This subsection discusses three important steps to help support an archival study's findings: (1) select appropriate samples, proxies, and empirical models; (2) consider alternative explanations; and (3) conduct tests to rule out alternative explanations. We focus on issues we believe accounting researchers frequently encounter and for which our PTCs can provide useful guidance.

4.a.i. Select the appropriate sample, proxies, and empirical models

Items 4.a.i.1–4 indicate that researchers need to carefully select a proper sample, identify good measures for the treatment and dependent variables, check data validity, and measure the independent and dependent variables during the appropriate time period. Before conducting initial statistical analyses, researchers should first think carefully about all of the items in subsection 4.a.i. The first item, 4.a.i.1, relates to sample selection. Researchers prefer samples that more closely represent the population so that the findings based on a specific sample generalize to the population. Researchers should carefully consider using randomization and stratified sampling with hand-collected data in order to generate a representative sample that also offers the maximum statistical power. However, when smaller and less representative samples are the only option because of limited data availability, costly data collection, and self-selection, researchers must be very cautious in generalizing from their study to other settings.

Item 4.a.i.2 emphasizes the validity of the study's measures, an issue that frequently draws particular attention from readers and reviewers. If a study develops a new measure, researchers should check the measure's construct validity. To the extent that a measure is noisy, it becomes more important to use alternative measures to check the robustness of the results.

4.a.ii. Consider alternative explanations

Many readers will search carefully for alternative explanations for documented findings, whereas the researchers' attachment to their own explanation can often blind them to considering such alternatives. We group alternative explanations into two main categories. First, item 4.a.ii.1 applies when variables x and y are intended to capture concepts A and B in empirical tests, but do so with considerable error. As a result, a statistical association between variables x and y does not necessarily imply that a corresponding association exists between concepts A and B. Second, item 4.a.ii.2 applies when there is a relatively good mapping from the concepts to the variables, but the documented association between variables x and y is biased. For example, assume that the initial model to support the researchers' theory takes the following form:

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_i x_i + \varepsilon, \tag{1}$$

where y is the dependent variable, x_i are independent variables ($i = 1, 2, \ldots, n$), and ε is the random error term. Given a positive estimated value of α_1, proposing an alternative explanation is equivalent to arguing that the significantly positive estimated coefficient α_1 is biased. A biased coefficient typically results from a violation of the linear regression assumption that x_1 is independent of ε. Such a violation can occur for three reasons: (a) omitted variables (item 4.a.ii.2.a); (b) simultaneity (item 4.a.ii.2.b); and (c) measurement error in x (item 4.a.ii.2.c—also see Wooldridge 2001, 50–51). We next discuss and illustrate the first two reasons, omitted variables, and simultaneity, because both occur frequently in archival research.[11]

An omitted variables problem (item 4.a.ii.2.a) arises when researchers are unable to obtain measures of certain variables that are correlated with both the dependent and independent variables. In this case, the effect of the omitted variables becomes part of the error term, resulting in a correlation between the error term and the independent variable. Self-selection, a common source of an omitted variable problem, can occur when the value of the dependent variable is observed only for a subsample that depends on some endogenous choice (Tucker 2010; Lennox, Francis, and Wang 2012). For example, in a study investigating the determinants of management forecast precision, the sample will contain only firms that have chosen to issue management forecasts. Self-selection can also occur when an independent variable is an endogenous indicator variable. For example, Kasznik and Lev (1995) examine the relation between firms' issuance of earnings warnings (x, an indicator variable) and the firms' stock returns (y), where a firm's decision to issue an earnings warning is endogenous. Without proper controls, both types of self-selection can lead to biased estimated coefficients because certain factors influencing the self-selection decisions could be related to the dependent variable.

11. Assume that a model a researcher is interested in is $y = a_0 + a_1 x_1^* + \varepsilon$. However, the researcher can observe only an imperfect measure of it x_1^*, $x_1 = x_1^* + u$, where u is the measurement error. The estimated model is $y = a_0 + a_1 x_1 + (\varepsilon - a_1 u)$. If u is related to x_1, x_1 will be correlated with the error term, $\varepsilon - a_1 u$, in the estimated model, leading to a biased estimate of a_1. This illustrates how a measurement error in an independent variable can cause a biased coefficient estimate.

Kasznik and Lev document a negative association between firms' issuance of earnings warnings and the firms' stock returns, which is consistent with investors penalizing firms that issue guidance. However, a firm's decision about whether to issue an earnings warning can be affected by some unobservable factors, including management's future profit expectations. Because these factors are often related to stock prices, excluding these factors creates an omitted variable problem. That is, the difference in stock returns between firms that issue earnings warnings and firms that do not can be driven by either the issuance of a warning itself or by other unobservable factors. Consistent with this concern, Tucker (2007) shows that firms that have more negative non-earnings news tend to issue more earnings warnings. She finds that after using the Heckman selection model to control for the unobservable negative news associated with warning issuance, the negative association between issuing earnings warnings and stock prices generally disappears. This example illustrates how self-selection can create an omitted variable problem that becomes an alternative explanation for the observed results.

The second potential source of biased estimated coefficients is simultaneity (4.a.ii.2.b), which occurs when x, the independent variable, influences y, the dependent variable, but y also simultaneously influences x. Because x is also a function of y, x is correlated with the error term. For example, prior studies argued that firms issue earnings warnings to reduce the firms' litigation risk. However, Francis, Philbrick, and Schipper (1994) document a positive association between issuing earnings warnings (x) and litigation risk (y), consistent with the warnings increasing litigation risk rather than decreasing it. Field, Lowry, and Shu (2005) explain this puzzle by arguing that firms with high litigation risk (y) are more likely to issue earnings warnings (x), making x also a function of y. Consistent with this reasoning, Field et al. show that after controlling for the simultaneity, issuing earnings warnings actually reduces litigation risk, which illustrates how failing to consider simultaneity can generate inappropriate inferences.

4.a.iii. Conduct tests to support your theory and rule out alternative explanations

Item 4.a.iii offers three suggestions (items 4.a.iii.1–3) for how researchers can support their theory by supplying evidence counter to alternative explanations. Our fourth recommendation (item 4.a.iii.4) involves econometric techniques, but the range and complexity of these techniques preclude our covering them here. Instead, we refer readers to econometrics textbooks and several recent accounting and finance papers that address specific issues that frequently arise when applying econometric methods in accounting and finance settings (Lennox et al. 2012; Larcker and Rusticus 2010; Roberts and Whited 2011; Tucker 2010).

The three additional tests that help distinguish a paper's story from alternative explanations are (1) tests of additional predictions (item 4.a.iii.1); (2) change analysis (item 4.a.iii.2); and (3) analysis based on exogenous shocks (item 4.a.iii.3). First, tests of additional predictions can bolster the paper's story and help to rule out alternative explanations for reported results. Specifically, cross-sectional analysis (item 2.a.v.1) often tests whether an association is significantly stronger in one setting than in others. Such analysis helps rule out alternative explanations caused by omitted variables as long as the omitted variables are not related to the cross-sectional conditions. For example, Bernard and Thomas (1990) find that the post-earnings-announcement drift is concentrated in three-day windows around the subsequent earnings announcements. This finding helps rule out the alternative explanation that the drift is driven by risk because risk is unlikely to be unusually high during the earnings announcement windows.

The second category of tests, change analysis (item 4.a.iii.2), investigates an association between changes in the dependent variable and changes in the independent variables.

Change analysis helps mitigate the effect of stable omitted variables when changes in the omitted variables are approximately zero, so that no further control for these variables is needed. For example, an alternative explanation for the negative association between internal control quality and cost of equity documented in Ashbaugh-Skaife, Collins, Kinney, and LaFond (2009) is that such an association is driven by certain systematic differences between firms with and without internal control deficiencies, rather than by the internal control deficiencies themselves. To rule out this alternative explanation, Ashbaugh-Skaife et al. (2009) conduct a change analysis and demonstrate that firms that remediate their internal control deficiencies, thereby changing their internal control quality, exhibit a significant decrease in their cost of capital around the disclosure of the remediation. Because any systematic differences between firms with and without internal control deficiencies are likely to remain unchanged around the disclosure of the remediation, this analysis adds support for the researchers' explanation by ruling out alternative explanations related to such systematic differences between firms with and without internal control deficiencies.

The third form of additional tests is exogenous shock analysis (item 4.a.iii.3), which identifies an exogenous event that causes changes in one or more independent variables, x. Such "natural experiments," seek to exploit environmental changes that are beyond the control of firms, investors, and other strategic players. By focusing on exogenous "shocks," researchers can eliminate the possibility that the observed change was itself driven by actions of the strategic players in the study's analysis, thus reducing endogeneity concerns. By construction, if y changes in the predicted direction with the exogenous changes in x, then x likely causes y instead of y causing x. Thus, an analysis based on an exogenous shock can help to establish causality.

Fang, Noe, and Tice (2009) document a positive association between a firm's stock liquidity and firm value, using an exogenous shock to help rule out an alternative explanation for their result. They argue that greater liquidity increases the information content of market prices, thereby enhancing the value of performance sensitive managerial compensation, and increasing firm value. However, an alternative explanation for their findings would reverse the causality by arguing that institutional investors are more likely to hold stocks with high firm value, and that this investment choice by the institutional investors would increase the liquidity of such firms. To help rule out this alternative explanation, Fang et al. (2009) identify an exogenous shock to firm liquidity that occurred in 2001 when the NYSE, AMEX and NASDAQ all switched from listing stocks in fractional prices with 16 price points to decimal prices with 100 price points. This switch reduced bid-ask spreads and the cost of trading, and thus increased liquidity, particularly for more actively traded stocks. Fang et al. (2009) document that the increase in stock liquidity caused by this exogenous event is positively associated with a change in firm value. This finding is consistent with the researchers' theory that a firm's stock liquidity affects its value, but is inconsistent with the alternative explanation that a firm's high value increases its liquidity, thus providing stronger support for their theory.

4.b. Research design and analysis for experimental studies

This subsection discusses important design and analysis issues for experimental studies. We provide additional guidance on items for which this is likely to be most useful. For expediency, we sometimes refer back to issues already discussed in the archival subsection.

Experimental researchers must plan *in advance* how to support the study's theory and rule out alternative explanations. Advance planning is important because once the experimental data are collected, obtaining additional data requires conducting additional experimental sessions or a new experiment. Items 4.b.i–iii suggest points to consider in advance

when designing an experiment, and items 4.b.iv.1–3 suggest tests that can provide empirical support for the theory and help rule out alternative explanations. Of course, the subsections of 4.b are related because it is important to consider which statistical tests will be conducted later to ensure that the necessary data are collected when the experiment is administered. Our discussion below relates primarily to points to consider in advance (items 4.b.i.1–4, item 4.b.ii–iii), with limited discussion of the statistical tests typically used to test a study's theory and to rule out alternative explanations.

4.b.i. Consider in advance which design and operationalization of variables provide the best test of your theory and help rule out alternative explanations

Item 4.b.i.1 notes the need to operationalize the independent/treatment variables in a manner that captures the theoretical constructs of interest. Archival researchers rely on variables that exist in the natural environment as proxies for the theoretical constructs of interest. In contrast, experimental studies can substantially reduce concerns about construct validity by developing their own measures that more precisely capture the theoretical constructs of interest.

Item 4.b.i.2 stresses the importance of selecting levels of the treatment variable that capture critical points along a continuum in order to test for predicted relationships among the variables. Pilot testing can help ensure that participants perceive the selected levels of the independent variables as intended. For example, if the goal is to measure the effects of low/medium/high incentives, a pilot study can help determine whether participants perceive the incentive treatment levels in the experimental instrument as three distinct levels of incentives.

In experiments, some variables are manipulated, while others are measured, held constant, or randomized across treatment conditions. Item 4.b.i.3 indicates that when deciding whether to operationalize a variable as exogenous or endogenous, it is important to consider how that variable occurs in the natural environment. Exogenous random assignment of participants to treatment conditions is a primary benefit of experiments (Shadish, Cook, and Campbell 2002) because this makes it unlikely that any observed difference in the dependent variable across treatment conditions is due to differences in specific characteristics of the participants (such as age, gender, or experience) assigned to conditions. Moreover, random assignment eliminates the self-selection problem and helps overcome the causality, omitted variable, and simultaneity problems discussed in the previous subsection. Because only the treatment variable changes across conditions, simultaneity issues are not a concern. Furthermore, because the independent variable is manipulated *before* the dependent variable is measured, the concern that some other omitted variable can explain the change in the dependent variable is reduced.

Despite the benefits of randomly assigning participants to manipulated levels of an independent variable, there can sometimes be a cost. Random assignment abstracts away from, and therefore cannot capture, certain important strategic interactions that occur in many actual settings of interest to accounting researchers. For example, randomly assigning employees to either a bonus or penalty contract precludes the opportunity to gain insights regarding how often or why superiors choose to offer each type of contract or how employees react to an employer's choice. In addition, assigning employees to a contract that they would not voluntarily accept limits the generalizability of the results to more realistic settings in which employers choose which type of contract to offer and employees choose which type to accept. Employee behavior may depend on both the contract and on the process by which employees find themselves facing the contract. If this is the case, exogenous assignment precludes gaining this insight and could produce misleading evidence regarding behavior in actual employment settings.

Item 4.b.i.3 suggests that when examining a variable that is endogenously determined in the natural environment, researchers should consider and acknowledge the potential implications of exogenously manipulating it in the laboratory. Similarly, when participants endogenously self-select into different levels of the independent variable in the laboratory, researchers should explicitly consider and acknowledge the potential implications of the loss of experimental control.[12] Researchers can use both approaches in the same study, or when an initial study uses exogenous random assignment, a follow-up study could examine whether the reported results hold when the participants are allowed to strategically interact.[13] A combination of both approaches typically provides a more complete understanding of the issues of interest.

Item 4.b.i.4 discusses using a within-subject versus between-subjects design. In within-subject designs, the same participant receives multiple-treatment conditions. This has the statistical advantage of using each participant as his/her own control group to minimize variability in responses. However, such designs can potentially introduce experimental demand because the manipulation may be transparent and salient to participants.[14] In between-subjects designs, different participants receive different treatment conditions and are unaware of the other treatment conditions, making the treatments less transparent and salient to participants. However, because participants vary on other characteristics unrelated to the treatment, between-subjects designs add variability within each treatment condition, making it harder to identify statistically significant differences across treatment conditions even when they exist.

In item 4.b.i.4.a, we suggest that researchers could use both a within-subject and between-subjects design. Schepanski et al. (1992) explain how to do so by counter-balancing the order in which participants receive the treatment conditions. For example, all participants receive Treatment 1 and Treatment 2 (T1 and T2), but one group receives T1 followed by T2, and the other group receives T2 followed by T1. Because each participant receives both treatments, this is a within-subject design. However, in the first stage, each participant receives only one treatment, either T1 or T2, so the first stage is effectively a between-subjects design. Koonce (1992) provides a good example of such a design. One group of auditors generated potential explanations for how a hypothesized cause could account for an unusual financial statement fluctuation, and then generated potential counter-explanations for why the hypothesized cause might not explain the fluctuation. A second group of auditors first generated counter-explanations and then generated explanations for the fluctuation. Counter-balancing the order of explanation/counter-explanation between participants allowed Koonce to obtain the advantages of both a within-subject and between-subjects design.[15]

12. Participants endogenously self-select into levels of all variables in Fehr, Kirchsteiger, and Riedl (1993), which introduced a two-stage experimental game to investigate the fair wage-effort hypothesis, often referred to as "gift exchange." Similarly, the "investment" or "trust game" introduced by Berg, Dickhaut, and McCabe (1995) focuses on strategic interaction, allowing participants to self-select into the levels of the variable of interest.

13. For an example of a study that uses both random assignment and self-selection in the same study, see Kachelmeier and Williamson (2010).

14. Schepanski, Tubbs, and Grimlund (1992) discuss why using a within-subject design need not necessarily result in experimental demand.

15. In addition to using both a within-subject and between-subjects design in the same experiment, researchers could conduct two separate experiments, one that manipulates the independent variable within-subject and a separate second experiment that manipulates the same variable between subjects. Another way that experimenters use both approaches in a single study is to manipulate one independent variable within participants, so that each participant acts as his/her own control group (thereby reducing variability), while manipulating a *different* independent variable between participants. Two examples of such designs are Lipe and Salterio's (2002) balanced scorecard study and Nelson and Kinney's (1997) study on the effect of ambiguity on loss contingency reporting judgments.

Reffett (2010) illustrates how within-subject information can be used to supplement the findings of a between-subjects experiment. He examines a setting in which a fraud was not detected by auditors. His main experiment uses a between-subjects design in which one group of evaluators is told that the auditors performed additional audit procedures designed to detect fraud, and another group of evaluators is told that the auditors had not performed such additional procedures. He finds that evaluators judge the auditors who performed the additional procedures as more liable for not detecting the fraud than auditors who had not performed such procedures. Then in a post-experimental question, each evaluator judges how liable they find the auditor under the condition that they had not seen in the main experiment, which constitutes a within-subject manipulation. Reffett (2010) finds that when evaluators are aware of both treatment conditions, they do not believe that auditors who performed additional audit procedures are more liable than auditors who had not performed such procedures. Thus, he concludes that evaluators in his main between-subjects experiment had not intentionally judged auditors to be more liable because they had performed additional audit work to detect fraud.

4.b.ii. Consider potential mediating or moderating variables that would be consistent with your theory but inconsistent with alternative explanations

As noted earlier, researchers should carefully consider in advance what data they will need to conduct the types of empirical tests indicated in item 4.b.iv.1–3. In addition to the items already discussed, deciding in advance to include a mediating or moderating variable (item 4.b.ii) enables researchers to collect the data necessary to conduct additional analysis designed to provide further support for their theory and to rule out alternative explanations.[16]

Coletti, Sedatole, and Towry (2005) illustrate the benefits of using mediation analysis. They hypothesize that the initial cooperation induced by a strong control system increases the mediating variable of trust, which has a positive effect on subsequent cooperation. This challenges the view in the prior literature that control systems reduce trust and therefore decrease cooperation. Coletti et al. (2005) randomly assigned participant pairs to one of two between-subjects conditions, a strong control system condition or a no control system condition. They find higher levels of initial cooperation, trust (the mediating variable), and subsequent cooperation in the strong control condition, indicating that the increased trust and higher subsequent cooperation in that condition result from the initial presence of the strong control system.

To further support this interpretation of the results, they conducted mediation analysis to show that the increased trust resulting from the initial strong control system mediated the effect of the initial strong control system on subsequent cooperation. This supports the underlying theory that an initial strong control system fosters initial cooperation, which increases trust among employees, which subsequently leads to employee cooperation even when the strong control system is no longer in place. The specificity of this process helps rule out alternative explanations because any such explanation(s) would need to explain the higher levels of initial cooperation and trust and the mediating effect of trust on subsequent cooperation in the strong control condition.

4.b.iii. Consider incorporating a predicted interaction to control for potential omitted variables

As was the case for mediating or moderating variables, predicting an interaction can help rule out alternative explanations. An interaction occurs when the effect of an inde-

16. Archival researchers often refer to mediating variables as intervening variables, as discussed earlier.

pendent variable (I_1) on a dependent variable (D) differs under different levels of a second independent variable (I_2). Assume that researchers predict and find an association between I_1 and D. It is helpful to also predict and document an interaction between I_1 and I_2 that is consistent with the researchers' theory, but is inconsistent with potential omitted variables that only predict a main effect of I_1 on D, but not the interaction. This approach provides additional support for the theory while also providing evidence to rule out omitted variables.

4.c. Research design and analysis for field studies

This subsection discusses design and analysis issues related to field studies. Merchant and Van der Stede (2006) show that the use of field research methods in accounting has grown during the period from 1981 to 2004, but that its use has been confined mostly to the management accounting area. Field studies' most significant contributions have been in discovering leading-edge practices and engaging in scholarly investigations of these practices. We follow Merchant and Van der Stede (2006) in delineating field research as involving direct contact with practitioners (e.g., employees, managers, regulators, consultants) in their natural working environments as a primary research approach of the study. Although field research often involves the use of multiple methods, such as administering a survey or collecting archival data, our PTCs focus on the characteristic of field research that distinguishes it from archival or survey research—contact with participants in the field. Also, as mentioned in footnote 3, our PTCs apply to field research with a positivistic orientation. While papers in the interpretivist and critical research traditions employ qualitative methods, they have a different ontological set of assumptions (see Ahrens and Chapman 2006).

Anderson and Lillis (2011, 1353) is a recent example of a field study that falls within our scope because it uses direct contact with organizational participants as one of its primary methods "to elicit rich descriptions of the phenomenon [of corporate frugality] from a limited set of field sites." Whereas some researchers distinguish case studies from field research because case studies involve only a single field site or company, we include "case studies" as an example of field research. Finally, although fieldwork is often exploratory and used in theory development, it can also be employed to test theory. In keeping with our positivistic approach, our guidance assumes that the field research relates closely to existing literature, and that it contributes to that literature by either developing theory or by using field data to examine testable propositions from the literature.

Next, we provide key PTCs to achieve the requisite reliability and validity of field research. In this context, reliability relates to steps the researcher can take to ensure the reader trusts the data and can draw valid inferences from them. Validity relates to the theoretical strength of the field work for theory development or theory testing.

4.c.i. Establish field data reliability

Interviewing is a common method field researchers use to interact with and collect data from organizational participants in the field, but other means of data collection such as direct observation and inspection of documents may also be used. Items 4.c.i.1–3 discuss methods for establishing the reliability of the data collected through these methods.

To establish data reliability, the field researcher must maintain good records of data collection protocols and analysis procedures, such as data recording and coding, so that others could replicate the results by following the same steps (item 4.c.i.1). For example, Phua, Abernethy, and Lillis (2011, 1803) explain that to ensure the integrity of their data, they "(1) used a pretested interview guide; (2) audio recorded all interviews after providing assurances of confidentiality; (3) checked transcripts for accuracy; and (4) asked partici-

pants to focus on outsourced activities with which they were most familiar." Phua et al.'s (2011) final point recognizes that although field research may be exploratory in nature, it must maintain a clear topical focus. Therefore, the researcher must carefully define the scope of the study and identify the sources relevant to addressing the research question. This implies developing a consistent approach to both selecting and classifying observations (item 4.c.i.2).

Because readers do not see the field data, they must rely on the field researcher's inferences from the data. Any reasonable evidence that field researchers can provide to corroborate their inferences enhances the study's credibility. Item 4.c.i.3 notes that using and triangulating multiple-data sources such as interviews and documentary evidence can enhance credibility. When the field method requires coding of interview transcripts or other field data, using multiple coders, including some who are blind to the study's theoretical context, can also enhance the data's reliability (item 4.c.i.4).

4.c.ii. Ensure the validity of the field study inferences

The most important criterion to judge a field study's validity is that it compellingly incorporates relevant literature in developing or testing theory. The researcher's direct involvement in the collection, classification, and interpretation of the field data makes it essential that the study satisfies the reader that the data collection is as free from researcher bias as possible. It is particularly important to avoid the confirmation bias that arises from selective attention to data patterns that support the proposed theory. Item 4.c.ii.1 emphasizes the need to guard against the natural tendencies to focus on patterns the researcher expects to see in the data and to dismiss observations that follow a pattern that appears inconsistent with the theory. The latter is particularly crucial to avoid when engaging in theory development because unexpected but systematic patterns in the data (while maintaining the chosen topical focus) are often where the field work can contribute to the literature. In the positivist spirit of field research, researchers can avoid biased attention and documentation to data by relying on one of several approaches to comprehensively document all data, including the matrix approach in Lillis and Mundy (2005) and causal maps in Abernethy, Horne, Lillis, Malina, and Selto (2005).

An effective field study must not only be based on the relevant literature; it must also exploit the field setting in developing or testing theory. To do so, field research must go beyond providing mere descriptions of practice. At the same time, field researchers cannot employ experimental controls or manipulations of the settings in which they conduct their field study. Guarding against the type of bias mentioned in 4.c.ii.1, field researchers can select their sample of site(s) to be maximally informative about their focal research question (item 4.c.ii.2). For example, the field researcher may want to conduct the study in settings believed to be either exemplars or outliers in order to exploit contrasting settings that focus on the theory's important relationships. In this way, the field researcher aims to establish validity through "theoretical sampling" rather than statistical sampling. The aim is not to generalize the field study findings *to* a population but rather to substantiate a theory by confirming or refuting it *across* a clearly-defined and carefully selected context (Yin 2003). If a theory does not hold in a particular case where conditions are ideally suited to the theory, then that theory is either falsified or is at least incomplete. In such instances, the field researcher must juxtapose an in-depth analysis of the context with a thorough understanding of the literature to suggest modifications of, or qualifications to, the existing theory, thereby contributing to theory development (Merchant and Van der Stede 2006).

For example, examining whether social connections within the hedge fund industry affect investment decisions, Simon, Millo, Kellard, and Engel (2012) observed an "unu-

sual" consensus trade over its lifecycle—the Volkswagen-Porsche trade—where Porsche had been buying VW stock for some time and had accumulated a significant stake. To take advantage of the discrepancy between the stock prices of VW and Porsche that this had caused, hedge fund managers chose a "popular" long-short trading position, going long in Porsche (purchasing Porsche stock) and short in VW (borrowing VW stock and selling it with the plan to buy it back later and return it to the lender). To be able to profitably unwind these positions, enough stock must be available in the market. However, Simon et al.'s (2012) field analyses revealed "several unusual characteristics" of the VW-Porsche trade that made the stock available for trading fall short of the demand by the short-sellers, resulting in a sharp increase in the market price of VW stock. The researchers used this "unusual" case to analyze how investors miss or ignore relevant information, contributing to theory about how the structure of ties among agents affects the transfer and interpretation of information, the resulting decision-making process, and wider market behavior.

4.d. Research design and analysis for survey studies

Survey research has been used more often in management accounting research than in financial accounting research due to a lack of publicly available data to proxy for variables of interest to management accounting researchers (see Luft and Shields 2003 for examples of such variables). It is not surprising that survey data based on perceptions, such as the respondent's self-assessed performance, raise concerns with reviewers and readers that typically do not arise for more verifiable and seemingly objective publicly-available data. Therefore, as with field research, establishing data credibility is a critical point to consider when conducting survey research. Survey research needs to pass the scrutiny of skeptical readers who want reasonable assurance that the survey is well-designed and well-executed, and thus yields valid data to address the research question.

Although we cannot provide all of the details here that a careful survey researcher should consider when designing and executing a survey, Dillman (1978, 1999), Fowler (2009), and Van der Stede, Young, and Chen (2005), among others, provide useful guides in this regard. Instead, we organize our discussion around the three main points of the suitability, generalizability (i.e., external validity), and internal validity (which is often referred to as just "validity") of survey data. Regarding suitability, researchers should ask whether a survey is the best way to examine their research question or whether other adequate proxy data are available from existing data sources (item 4.d.i). Sometimes a survey may be necessary to obtain some, but not all, of the required data for a study. In such cases, it may be possible to link the survey data with archival data.

Regarding generalizability, readers are more likely to accept survey data as externally valid when they are sufficiently confident that knowledgeable respondents from an unbiased sample of the target population have completed a well-designed survey (item 4.d.ii). Regarding validity, survey questions should provide careful, unbiased measurement of all the key variables, including a reasonable set of control variables, to allow robust inferences about the relationships among the variables. One threat to the internal validity of survey data is that the measures of both the dependent and explanatory variables for each sample observation are typically collected at one point in time from a single respondent (item 4.d.iii). It is difficult to establish causality using survey data because the measurement of variables is not separated in time. Because each survey sample observation reflects a single respondent's perceptions about the constructs of interest, for example, the performance of a given practice as well as its various design characteristics, survey variables may be spuriously related as a result of common-method bias. Limiting potential bias in

both the *respondents* (item 4.d.ii.3) and their *responses* (item 4.d.iii.2) is therefore critical for survey research.

4.d.i. Match the design of the survey with the purpose of the study

When adequate proxy data are unavailable or incomplete, the survey method can help address important research questions. However, the appropriateness of the survey method should be judged in light of the objectives of the study. Thus, it is important to explicitly state the objective of the study, and then design the survey to fulfill that purpose. For exploratory studies, the survey should be designed to yield responses that describe the phenomenon of interest in a representative way that can potentially inform follow-up studies (e.g., Dichev, Graham, Harvey, and Rajgopal 2013; Graham, Harvey, and Rajgopal 2005). For such studies, the necessary measures are unlikely to be available from prior research, and thus the researchers will need to carefully develop many new questions and measures. For survey studies intended to test theory (e.g., Fullerton, Kennedy, and Widener 2013), it is critical that the survey-based results can either reject or provide support for the theory. In this case, readers will expect researchers to rely, to the extent possible, on existing measures with established psychometric properties. Failing to do so will raise concerns about the survey measures' validity. Although both generalizability and validity are essential for any survey study, exploratory surveys will have limited value if they have weak generalizability or fail to describe the phenomenon of interest in a representative way, whereas theory-testing surveys will have limited value if they have weak internal validity due to measurement problems.

4.d.ii. Establish survey data generalizability

An important issue for survey studies is to choose and recruit the appropriate participants (4.d.ii.1). Selecting the appropriate respondents increases the study's generalizability, which is generally viewed as a strength of the survey method. The challenge is how to gain access to the appropriate respondents and to persuade them to participate in the survey. When there is a low response rate, surveys are subject to nonresponse bias (4.d.ii.3). The concern is that those who respond may not be representative of the population from which they are drawn. In addition to concerns about such self-selection in the types of individuals who respond, it is also important to be aware of characteristics of the survey itself that may reduce all individuals' likelihood of responding to the survey. Item 4.d.ii.2 suggests that pretesting the survey instrument can reduce the chance that issues such as survey length, question order (especially the placement of sensitive questions), question type (closed or open-ended), uncertainty regarding anonymity, and distribution method (e.g., mail versus online)[17] will cause either a general decrease in response rate or result in a biased sample.

After the study is pretested, survey participants can be selected from such sources as professional organizations, customer or trade publication lists, and training sessions. However, researchers should be aware that selection from these subpopulations could introduce response bias even when there is a high response rate. For example, CFOs who enroll in certain types of executive education or who participate in quarterly industry polls may differ from other CFOs in important ways, such as in their propensity to engage in earnings management.

17. For example, spam filters may prevent target participants in certain firms from ever receiving an electronic survey.

4.d.iii. Ensure the internal validity of the survey study inferences

To establish valid inferences from survey data, the data must be reliable and the statistical analyses must be appropriate for examining the research question. Regarding variable measurement (4.d.iii.1), a single survey question can rarely reliably capture complex, latent constructs such as a manager's decision horizon or job-related performance pressure. Instead, researchers need to draw from the literature to develop multiple-survey questions, and then perform psychometric tests to assess measurement quality and to construct reliable measures. The goal is to increase the measures' reliability, which is the internal consistency among survey items, and its validity, which is the degree to which the variable measures the construct it purports to measure. The validity of a measure can be econometrically examined through factor analysis. A properly validated measure will be more credible if it correlates as expected with alternate test variables. For example, Fullerton et al. (2013) correlate their measure of lean manufacturing strategy with a measure of whether the firm identifies itself as a world-class manufacturer based on the idea that firms that implement lean practices are more likely to be world-class manufacturers (see Fullerton et al. 2013, 69 for other examples). Ideally, measures are validated against external measures, but when this is not possible, such as when the survey is anonymous, the next best alternative is to correlate related variables within the survey.

Item 4.d.iii.2 focuses on threats to validity that are important in survey studies. We first focus on threats due to common-method bias and single-informant bias. Common-method bias occurs when there is a spurious association between variables due to poor measurement of variables stemming from question ordering, grouping, labeling, or formatting. Dichev et al. (2013) discuss how an advantage of online survey administration is the ability to randomly scramble the order of choices within a question to mitigate potential order-of-presentation effects. Single-informant bias may arise when, as is commonly the case, all of the questions in a survey are answered by the same respondent. Obtaining some of the data from different respondents can help mitigate this bias. For example, it could be useful to collect a supervisor's assessment of the survey respondent's performance rather than the respondent's self-assessment of his or her performance. However, this would require a multirespondent survey, which is often impractical and can lead to lower response rates. This type of bias sometimes can be reduced by checking that participants responded consistently to similar or related survey questions or by linking survey data with archival data.

Other threats to validity are biases such as social desirability bias or bias introduced by leading or suggestive questions. For example, respondents who manage earnings may be reluctant to admit it because earnings management could be perceived as socially undesirable. Conversely, a survey question could be phrased in a way that causes respondents to agree that earnings management is acceptable because the survey question suggests that "everyone does it." Social desirability bias can be mitigated by providing respondents with anonymity. Pretesting can also help to identify and remedy questions that may be prone to bias.

In addition to procedural remedies, statistical checks of survey bias include exploratory factor analyses on the survey items used to measure the variables of interest. Setting aside the many complications of employing exploratory factor analysis in an econometrically robust way, the aim is to extract factors from the data. Specifically, when survey items are strongly intercorrelated, but weakly correlate with items outside their group of intercorrelated items, they are viewed as a factor. If such analysis indicates that all items merge into a few factors—or fewer factors than anticipated—that account for most of the variance across all the items, this suggests that the study suffers from common-method bias. For example, common-method bias could occur if respon-

dents consistently choose the middle point of the answer scale regardless of the question asked. In contrast, if the exploratory factor analysis reveals, say, six factors that correspond well with the postulated constructs of interest, then there is less concern about common-method bias. But even when there is a mapping from the factors onto the intended variables, if only a small number of these variables account for most of the variance among all the survey items, readers may still be concerned about common-method bias.

Survey studies by Bouwens and van Lent (2007) and Abernethy, Bouwens, and van Lent (2013) provide good examples of both procedural remedies and statistical checks used to alleviate validity threats. To increase readers' confidence that the inferences drawn from the survey are not unduly affected by threats to validity, the survey should be designed and implemented to mitigate such threats, and then subsequently validated using robust econometric checks.

5. Interpretation of results and conclusions

After reporting the statistical significance of their results, researchers should evaluate the importance of the findings by estimating the economic or practical magnitude and the effect size of their results (item 5.a.i). If the effect size and economic significance are not consistent with the researchers' theory or the findings from prior studies, further consideration and analysis is required (5.a.ii). If the economic effect is smaller than expected, understanding why the effect is small will help researchers assess whether their results remain important. Similarly, if the economic magnitude of the estimated effect is much larger than expected, understanding why the effect is so large will help researchers assess whether their results are credible.

Our list suggests various empirical analyses to test a study's predictions. In most empirical studies, some results support the study's theory while others do not. Item 5.a.iii indicates that for the latter cases, researchers need to consider possible explanations, including (1) data error, (2) low power due to small sample size, (3) important omitted factors, and (4) incomplete or invalid theory. Considering why some tests fail can help researchers revise or refine their theory, improve their research design, or evaluate the reliability and robustness of their empirical findings.

The concluding section of a paper describes the implications of the findings and how they contribute to the literature. Item 5.b.i. emphasizes the importance of relating the conclusions back to the questions or puzzles that motivated the study in the paper's introduction. Researchers should acknowledge all of the study's findings, even those that do not support the paper's main thrust, and qualify the study's results appropriately. It is also important to discuss the settings to which the findings would likely apply, and not to overgeneralize them to settings to which they are less likely to apply (5.b.ii).

3. Discussion and conclusion

We provide a list of PTCs designed to help empirical accounting researchers improve the quality of their research, and offer suggestions and illustrative examples for how to address some of the more challenging points. Several issues regarding our purpose and scope warrant further discussion. First, for practical reasons we are unable to address all aspects of every issue related to the research process. We selected issues that arise frequently and that we believe are important and challenging. Specifically, as detailed in footnote 3 and in the field study research section, a limitation of our approach is that we do not address critical and interpretivist research, which employs qualitative methods, but from an ontological vantage point different from our positivistic approach. We also do not address how to generate an initial research idea, the more detailed mechanics of writ-

ing the paper, or strategies for publishing the manuscript.[18] While we do not directly address how to maximize the probability of publishing a study, we believe that by addressing our objective of producing "a thoughtful and carefully designed study," a researcher simultaneously increases the odds of publication. We recognize that not all thoughtful and properly executed studies are published in the first (or even the first few) journal(s) to which they are submitted, and that there is an element of chance in the publication process. However, because journals compete to publish good work, particularly papers that draw readers and that will be widely cited, in the long run it is likely that the better the quality of the research, the more likely it is to be published in a good journal. Researchers obviously need to consider the match between their study and a journal's editorial policies, but this and other considerations related to publishing are beyond the scope of the current study.

Second, we acknowledge that our paper differs from most accounting research in that we do not provide new knowledge concerning accounting phenomena. Rather, our intention is to help less experienced accounting researchers execute the research process more effectively. We introduce our list of PTCs as a decision aid to help such researchers avoid overlooking certain key points or treating them inadequately. Third, readers should note that it is unlikely that any one study would involve all of the issues we raise or remedies we suggest.

We believe our list of PTCs, or adaptations of it, can help accounting researchers in two other ways. First, recent graduates who have been working on their dissertation for years often underestimate the need to clearly define their question and its importance to new readers. The fact that a dissertation committee has already endorsed their study could prevent recent graduates from anticipating the importance of convincing new readers of the study's contribution. Recent graduates have typically devoted a great deal of time to justifying the validity of their study. This focus can result in insufficient attention to establishing the study's general importance or incremental contribution. Our PTCs can help such less experienced researchers anticipate the types of questions that are likely to be raised at job interview workshops, by journal reviewers, or by other new readers.

Second, we believe that our list of PTCs can also improve communication among research teams, particularly when different team members take primary responsibility for particular aspects of the project. Because a team member focused on one aspect of the study can lose perspective on the entire project, the PTCs provide a framework to facilitate understanding of the entire project and communication among team members.

Although our list of PTCs is a broader decision aid than conventional checklists, it shares a few characteristics with them. For example, like most checklists, our list reminds users to devote sufficient attention to critical tasks that they might overlook due to cognitive limitations or time constraints. However, unlike many checklists, our list includes items that require detailed contemplation over an extended period of time. Further, many items on our list involve considerable uncertainty and subjectivity such that, even after serious consideration, researchers cannot be certain that they have been resolved satisfactorily (Luft and Shields 2014). By comparison, many simpler checklists are intended to prevent short-term, operational mistakes, such as a surgeon leaving a sponge in a patient. While such operational considerations are obviously very important, such checklists (e.g., Gawande 2009) do not offer any further suggestions for how to perform the overall task,

18. The "more detailed mechanics of writing the paper" refers to the manuscript's exposition. This includes the organization and integration across and within various sections of the paper, as well as across and within the paper's paragraphs and sentences. Good mechanical exposition provides a clear, direct, logical, and efficient flow of reasoning throughout the manuscript.

for example, surgery, more effectively. In contrast, our list of PTCs provides many such suggestions for how to address the important issues noted on the list.

Because both our list of PTCs and checklists are intended to serve as memory aids, we considered the potential disadvantages of using checklists when designing our list. In particular, prior auditing research has documented that checklists can lead to mechanistic processing and interference effects (Asare and Wright 2004; Heiman-Hoffman, Moser, and Joseph 1995; Frederick 1991; Pincus 1989). We therefore designed our list to encourage researchers to think more deeply about the associated issues rather than mechanistically checking off items on the list. Clearly, our list does not address all possible research issues and no decision aid could offer a mechanical formula for completing a quality accounting research project. Further, we recognize that our PTCs cannot replace the education, feedback from colleagues and seminars, and hard work necessary to successfully complete high quality accounting research.

In summary, we provide critical points for researchers to consider in the process of converting an initial research idea into a carefully designed study. We elaborate on selected issues that arise frequently in seminars and in reviewer comments, and provide suggestions to address them. We offer a framework that senior doctoral students, junior faculty, and other accounting researchers can use to self-assess how effectively they have addressed key issues that we believe are often not addressed adequately in the initial versions of empirical accounting studies.

Appendix

List of points to consider (PTCs) when self-assessing your empirical accounting research

1. Research Question: What is your precise research question? Be clear in your mind and in your manuscript concerning:
   a. The meaning of key terms in the research question
   b. The scope of the research question. No study can examine all aspects of an issue, but it is important to justify why the part you are focusing on is appropriate by considering how the scope affects the focus, validity, and contribution of the study.
   c. Whether the research question involves a causal relation or an association
2. Theory: Is your study's conceptual framework logically consistent and credible?
   a. Develop your conceptual framework by applying theoretical arguments and/or previous empirical findings to a specific setting
      i. Ensure that your literature review is thorough enough to identify all major theories related to your story that lead to the same or opposite predictions
      ii. Consider whether your theory holds when analyzed from all key actors' points of view. Apply your study's assumptions about incentives and preferences in a consistent manner. If the assumptions for key actors are not consistent, justify them.
      iii. Ensure that all hypotheses/predictions in your setting are consistent with each other
      iv. Specify the mechanism by which the concepts in your setting relate to each other —the more detailed the explanation, the more testable and credible is your theory
      v. Specify additional effects one should or should not observe if the theory is correct
         1. Discuss whether your theory should hold or should not hold for different settings (cross-sectional analysis)
         2. Discuss the potential consequences if your theory holds
         3. Discuss the intervening effect if your theory holds
   b. Explain how your study distinguishes among alternative explanations for your predictions, or discuss why it is not important to do so in your setting
3. Contribution: What does your study add to our understanding of accounting?
   a. Establish how your study is new and interesting relative to the prior literature
      i. The topic itself is interesting and important
         1. It relates to a pervasive phenomenon or emerging trend
         2. It relates to some controversial, regulatory issue, or because it is a hot topic
         3. It has a large effect size or economic significance
         4. It addresses a fundamental accounting question or a critical factor related to financial or managerial reporting, auditing, tax, or other aspects of an organization's performance
      ii. The study is differentiated from and extends prior literature
         1. The study expands our knowledge of a previously studied issue by investigating the issue from a new perspective or identifying a new setting, method or theory
         2. The study reconciles previously mixed results or resolves a puzzle in the literature

3. The results cannot be directly derived from previous studies or one's intuition. If the findings could appear obvious to readers, discuss what else could have occurred to explain why they are not so obvious.

b. Identify important implications in terms of actions or beliefs that will change based on your study. Whose beliefs or actions (i.e., researchers, investors, financial analysts, managers, auditors, tax preparers, regulators, policymakers) might change as a result of the study?

4. Research design and analysis

a. Research design and analysis for *archival studies*

i. Select the appropriate sample, proxies, and empirical models

1. Consider the size and representativeness of your sample carefully when selecting the sample
2. Check that the independent and dependent variables capture the behavior of interest
3. Check that your data appears reasonable given existing knowledge in the field
4. Ensure that the timing of the independent and dependent variables matches the theoretical relationship

ii. Consider alternative explanations

1. The dependent or treatment variable captures a construct other than the behavior of interest
2. The documented association between the dependent and treatment variables is biased because

a. There are omitted variables
b. The dependent variable also affects the treatment variable (simultaneity)
c. There is measurement error in the treatment variable

iii. Conduct tests to support your theory and rule out alternative explanations

1. Conduct comprehensive analyses to test various predictions in the fully developed story as listed under items 2.a.v.1 to 2.a.v.3
2. Conduct change analysis to hold constant and remove the effect of omitted factors that are fairly stable across time
3. Consider incorporating exogenous shocks into analysis to control for the order of causality by having the shock act as a manipulated variable
4. Use the appropriate econometric technique developed to address the particular concern (e.g., use the Granger Causality test to determine that the causality is not the reverse of what you predict)

b. Research design and analysis for *experimental studies*

i. Consider *in advance* which design and operationalization of variables provide the best test of your theory and help rule out alternative explanations

1. Consider whether the operationalization of the independent variable captures the theoretical construct
2. Determine whether the selected levels of the treatment variable provide sufficient variability

3. Consider whether the treatment variable is exogenously or endogenously determined in the world and how best to operationalize the variable in the experiment

   a. If an endogenous variable is operationalized as exogenous to achieve experimental control (e.g., via random assignment or matching), consider the costs and benefits of using this approach

4. Decide whether to use a between-subjects or within-subject design

   a. Consider using both a within- and between-subjects design so participants can act as their own control group without creating experimental demand and to minimize the risk of omitted variables

ii. Consider potential mediating or moderating variables that would be consistent with your theory and inconsistent with alternative explanations
iii. Consider incorporating a predicted interaction to control for potential omitted variables
iv. Conduct tests to support your theory and rule out alternative explanations

1. Conduct statistical analyses (e.g., regression or ANOVA) to provide initial support for your theory
2. Conduct mediation analysis (e.g., ANCOVA, path analysis) to test for intervening effects
3. Consider using any background post-experimental data (e.g., years of experience, industry experience) as a control variable to reduce variability

c. Research design and analysis for *field studies*

i. Establish field data reliability

1. Document and keep a record of data collection protocols and analysis procedures
2. Clearly define your focal construct(s) and develop decision rules about how to classify observations
3. Use multiple-data sources (such as interviews and documentary evidence) to corroborate findings
4. Use multiple coders who are "blind" to the theory

ii. Ensure the validity of the field study inferences

1. Avoid confirmation bias by paying attention to data patterns that are not explained by your theory
2. Pertinently engage the field setting in theory development or testing (e.g., employ theoretical sampling rather than statistical sampling)

d. Research design and analysis for *survey studies*

i. Match the design of the survey with the purpose of the study
ii. Establish survey data generalizability

1. Target and select knowledgeable participants as appropriate for the research question
2. Pretest the survey instrument with a small group of target participants
3. Avoid nonresponse bias

iii. Ensure the internal validity of the survey study inferences

1. Use appropriate survey data reduction analyses to construct variables and establish robust psychometric measurement properties
2. Avoid response biases, such as common-method bias and bias due to leading questions or socially desirable responses

5. Interpretation of results and conclusions

a. Interpretation of empirical results

i. Describe the statistical significance, and the economic magnitude or effect size of the results, if applicable
ii. Discuss and justify the pattern and magnitude of the results based on your story and findings in prior studies
iii. If data support only part of the predicted pattern, or if different analyses lead to different statistical inferences, consider possible reasons for this to evaluate the reliability of your findings

b. Conclusions

i. Relate your conclusions back to your motivation and research question
ii. Avoid over-generalizing or over-concluding
iii. Discuss the contribution and implications of your findings

References

Abernethy, M., M. Horne, A. Lillis, M. Malina, and F. Selto. 2005. A multi-method approach to building causal performance maps from expert knowledge. *Management Accounting Research* 16 (2): 135–55.

Abernethy, M., J. Bouwens, and L. van Lent. 2013. The role of performance measures in the intertemporal decisions of business unit managers. *Contemporary Accounting Research* 30 (3): 925–61.

Ahrens, T., and C. S. Chapman. 2006. Doing qualitative field research in management accounting: Positioning data to contribute to theory. *Accounting, Organizations and Society* 31 (8): 819–41.

Anderson, S. W., and A. M. Lillis. 2011. Corporate frugality: Theory, measurement and practice. *Contemporary Accounting Research* 28 (4): 1349–87.

Asare, S. K., and A. M. Wright. 2004. The effectiveness of alternative risk assessment and program planning tools in a fraud setting. *Contemporary Accounting Research* 21 (2): 325–52.

Ashbaugh-Skaife, H., D. W. Collins, W. R. Kinney, and R. LaFond. 2009. The effect of SOX internal control deficiencies on firm risk and cost of equity. *Journal of Accounting Research* 47 (1): 1–43.

Bamber, L. S., J. X. Jiang, and I. Y. Wang. 2010. What's my style? The influence of top managers on voluntary corporate financial disclosure. *The Accounting Review* 85 (4): 1131–62.

Barton, J., and M. Mercer. 2005. To blame or not to blame: Analysts' reactions to external explanations for poor financial performance. *Journal of Accounting and Economics* 39 (3): 509–33.

Bebchuk, L., and J. Fried. 2004. *Pay without performance: The unfulfilled promise of executive compensation*. Cambridge: Harvard University Press.

Berg, J., J. Dickhaut, and K. McCabe. 1995. Trust, reciprocity, and social history. *Games and Economic Behavior* 10 (1): 122–42.

Bernard, V. L., and J. K. Thomas. 1990. Evidence that stock prices do not fully reflect the implications of current earnings for future earnings. *Journal of Accounting and Economics* 13 (4): 305–40.

Bouwens, J., and L. van Lent. 2007. Assessing the performance of business unit managers. *Journal of Accounting Research* 45 (4): 667–97.

Brown, L. 2005. The importance of circulating and presenting manuscripts: Evidence from the accounting literature. *The Accounting Review* 80 (1): 55–83.

Campbell, D., M. Epstein, and F. Martinez-Jerez. 2011. The learning effects of monitoring. *The Accounting Review* 86 (6): 1909–34.

Christenson, C. 1983. The methodology of positive accounting. *The Accounting Review* 58 (1): 1–22.

Chua, W. F. 1986. Radical developments in accounting thought. *The Accounting Review* 61 (4): 601–32.

Coletti, A., K. L. Sedatole, and K. L. Towry. 2005. The effect of control systems on trust and cooperation in collaborative environments. *The Accounting Review* 80 (2): 477–500.

Core, J., W. Guay, and R. Thomas. 2005. Is U.S. CEO compensation inefficient pay without performance? *Michigan Law Review* 103 (6): 1142–85.

Dechow, P. M., W. Ge, and C. Schrand. 2010. Understanding earnings quality: A review of the proxies, their determinants and their consequences. *Journal of Accounting and Economics* 50 (2–3): 344–401.

Dichev, I., J. Graham, C. Harvey, and S. Rajgopal. 2013. Earnings quality: Evidence from the field. *Journal of Accounting and Economics* 56 (2–3): 1–33.

Dillman, D. 1978. *Mail and Telephone Surveys: The total design method*. New York: Wiley.

Dillman, D. 1999. *Mail and Internet Surveys: The tailored design method*. New York: Wiley.

Fang, V. W., T. H. Noe, and S. Tice. 2009. Stock market liquidity and firm value. *Journal of Financial Economics* 94 (1): 150–69.

Fehr, E., G. Kirchsteiger, and A. Riedl. 1993. Does fairness prevent market clearing? An experimental investigation. *The Quarterly Journal of Economics* 108 (2): 437–59.

Field, L., M. Lowry, and S. Shu. 2005. Does disclosure deter or trigger litigation? *Journal of Accounting and Economics* 39 (3): 487–507.

Fowler, F. J. 2009. *Survey research methods*, 4th ed. Thousand Oaks, CA: Sage.

Francis, J., and K. Schipper. 1999. Have financial statements lost their relevance? *Journal of Accounting Research* 37 (2): 319–52.

Francis, J., D. Philbrick, and K. Schipper. 1994. Shareholder litigation and corporate disclosures. *Journal of Accounting Research* 32 (2): 137–64.

Frederick, D. M. 1991. Auditors representation and retrieval of internal control knowledge. *The Accounting Review* 66 (2): 240–58.

Frederick, D. M., and R. Libby. 1986. Expertise and auditors' judgments of conjunctive events. *Journal of Accounting Research* 24 (2): 270–90.

Fullerton, R. R., F. A. Kennedy, and S. K. Widener. 2013. Management accounting practices and control in a lean manufacturing environment. *Accounting, Organizations and Society* 38 (1): 50–71.

Gawande, A. 2009. *The checklist manifesto: How to get things right*. New York: Metropolitan Books.

Graham, J., C. Harvey, and S. Rajgopal. 2005. The economic implications of corporate financial reporting. *Journal of Accounting and Economics* 40 (1–3): 3–73.

Heiman-Hoffman, V. B., D. V. Moser, and J. A. Joseph. 1995. The impact of an auditor's initial hypothesis on subsequent performance at identifying actual errors. *Contemporary Accounting Research* 11 (2): 763–79.

Hines, R. D. 1988. Popper's methodology of falsificationism and accounting research. *The Accounting Review* 63 (4): 657–62.

Ittner, C. D. 2014. Strengthening causal inferences in positivist field studies. *Accounting, Organizations and Society* 39 (7): 545–49.

Kachelmeier, S. J. 2004. Reviewing the review process. *Journal of the American Taxation Association* 26 (Supplement): 143–54.

Kachelmeier, S. J., and M. G. Williamson. 2010. Attracting creativity: The initial and aggregate effects of contract selection on creativity-weighted productivity. *The Accounting Review* 85 (5): 1669–91.

Kasznik, R., and B. Lev. 1995. To warn or not to warn: Management disclosures in the face of an earnings surprise. *The Accounting Review* 70 (1): 113–34.

Kinney, W. R. 1986. Empirical accounting research design for Ph.D. students. *The Accounting Review* 61 (2): 338–50.

Kinney, W. R. 2011. Planning for research success. Presentation at AAA Annual Meeting in Denver, Colorado.

Koonce, L. 1992. Explanation and counter-explanation in audit analytical review. *The Accounting Review* 67 (1): 59–76.

Larcker, D. F., and T. O. Rusticus. 2010. On the use of instrumental variables in accounting research. *Journal of Accounting and Economics* 49 (3): 186–205.

Lennox, C. S., J. R. Francis, and Z. Wang. 2012. Selection models in accounting research. *The Accounting Review* 87 (2): 589–616.

Lillis, A., and J. Mundy. 2005. Cross-sectional field studies in management accounting research: Closing the gaps between surveys and case studies. *Journal of Management Accounting Research* 17 (1): 119–41.

Lipe, M. G., and S. Salterio. 2002. A note on the judgmental effects of the balanced scorecard's information organization. *Accounting, Organizations and Society* 27 (6): 531–40.

Luft, J., and M. D. Shields. 2003. Mapping management accounting: Graphics and guidelines for theory-consistent empirical research. *Accounting, Organizations and Society* 28 (2–3): 169–249.

Luft, J., and M. Shields. 2014. Subjectivity in developing and validating causal explanations in positivist accounting research. *Accounting, Organizations and Society* 39 (7): 550–58.

Maines, L. A., G. L. Salamon, and G. B. Sprinkle. 2006. An information economic perspective on experimental research in accounting. *Behavorial Research in Accounting* 18 (1): 85–102.

Merchant, K. A., and W. A. Van der Stede. 2006. Field-based research in accounting: Accomplishments and prospects. *Behavioral Research in Accounting* 18 (1): 117–34.

Nelson, M. W., and W. R. Kinney. 1997. The effect of ambiguity on loss contingency reporting judgments. *The Accounting Review* 72 (2): 257–74.

Phua, Y., M. Abernethy, and A. Lillis. 2011. Controls as exit barriers in multi-period outsourcing arrangements. *The Accounting Review* 86 (5): 1795–834.

Pincus, K. V. 1989. The efficacy of a red flags questionnaire for assessing the possibility of fraud. *Accounting, Organizations and Society* 14 (1–2): 153–63.

Reffett, A. B. 2010. Can identifying and investigating fraud risks increase auditors' liability? *The Accounting Review* 85 (6): 2145–67.

Roberts, M. R., and T. M. Whited. 2011. Endogeneity in empirical corporate finance. Available online at ssrn.com/abstract = 1748604, retrieved January 24, 2013.

Rogelberg, S. G. 2002. *Handbook of research methods in industrial and organizational psychology*, Appendix 22.1, 466–78. Malden, MA: Wiley-Blackwell.

Schepanski, A., R. M. Tubbs, and R. A. Grimlund. 1992. Issues of concern regarding within- and between-subjects designs in behavioral accounting research. *Journal of Accounting Literature* 11 (1): 121–50.

Shadish, W. R., T. D. Cook, and D. T. Campbell. 2002. *Experimental and quasi-experimental designs for generalized causal inference*. Wadsworth: Cengage Learning.

Simon, J., Y. Millo, N. Kellard, and O. Engel. 2012. Close connections: Hedge funds, brokers and the emergence of a consensus trade. Working paper, London School of Economics and Political Science.

Sloan, R. G. 1996. Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review* 71 (3): 289–315.

Tucker, J. W. 2007. Is openness penalized? Stock returns around earnings warnings. *The Accounting Review* 82 (4): 1055–87.

Tucker, J. W. 2010. Selection bias and econometric remedies in accounting and finance research. *Journal of Accounting Literature* 29 (Winter): 31–57.

Van der Stede, W. A. 2014. A manipulationist view of causality in cross-sectional survey research. *Accounting, Organizations and Society* 39 (7): 567–74.

Van der Stede, W. A., M. Young, and C. Chen. 2005. Assessing the quality of evidence in empirical management accounting research: The case of survey studies. *Accounting, Organizations and Society* 30 (7–8): 655–84.

Watts, R. L., and J. L. Zimmerman. 1990. Positive accounting theory: A ten year perspective. *The Accounting Review* 65 (1): 131–56.

Wooldridge, J. M. 2001. *Econometric analysis of cross section and panel data*. Boston, MA: MIT Press.

Yin, R. K. 2003. *Case study research, design, and methods*, 3rd ed. Newbury Park, CA: Sage.

Zimmerman, J. L. 1989. Improving a manuscript's readability and likelihood of publication. *Issues in Accounting Education* 4 (2): 458–66.