Theory and Experimentation in Studies of Audit Judgments and Decisions: Avoiding Common Research Traps

Mark E. Peecher* and Ira Solomon
University of Illinois at Urbana-Champaign, US

In this article, we identify and discuss four research traps into which behavioral audit scholars sometimes fall: (1) The Rush to the Explanatory Research Stage Trap; (2) The Mundane Realism Trap; (3) The Measured Variable Trap; and (4) The Audit Practitioner as Participant Trap. Throughout our discussion of these traps, we emphasize the role of theory and highlight the comparative advantages of experiments. Our discussion is designed to help aspiring researchers carefully design and execute auditing research programs.

Key words: auditing; experiment; explanatory research; internal validity; manipulated variables; mundane realism; measured variables; random assignment

INTRODUCTION

There is a rich heritage of experimentation by scholars investigating audit judgments and decisions (see Trotman, 2001 for examples of the research questions investigated). In this paper, we discuss the role of theory and the comparative advantages of experiments. Subsequently, we leverage the initial discussion into identification and discussion of research traps into which we have seen audit scholars fall. Our goal is to help aspiring auditing scholars sort out some of the vexing issues that one faces when planning and executing experiments. Perhaps, however, we also will say a few things along the way that will catch the interest and attention of the seasoned scholar.

DEFINITIONS

The next section of this paper presents the foundation for our discussion of research traps by providing annotated definitions of key terms, including auditing, judgments, decisions, research, theory, experiment, random assignment, internal and external validity, and manipulated and measured variables. In the third section we identify and discuss four research traps: (1) The Rush to the Explanatory Research Stage Trap; (2) The Mundane Realism Trap; (3) The Measured Variable Trap; and (4) The Audit Practitioner as Participant Trap. The fourth section completes the paper by providing concluding remarks.
interested parties (see Solomon and Shields, 1995). There are many types of audits, but the best known is an examination of the conformance of financial statements with generally accepted accounting principles conducted by independent auditors. There also are many types of auditors (e.g., internal auditors). Further, the subject matter of an audit could be other than financial statements (e.g., conformance with international standards for product and service quality, individual web-site’s standards for customer privacy; see Solomon and Peecher, 2001).

Irrespective of the subject matter being examined, the auditor makes numerous judgments and decisions throughout the audit process. We use the term *judgment* to refer to assessments that people make as a prelude to taking action. We use the term *decision* to mean actions that people take to perform a task or solve a problem. The auditor makes judgments and decisions within dynamic, strategic settings. To better understand, and ultimately improve the auditor’s diverse set of judgments and decisions, we conduct research.

Some persons define *research* as systematic, controlled, empirical and critical investigation of natural phenomena guided by theory and hypotheses about the presumed relations among such phenomena (see Kerlinger and Lee, 2000, p. 14). This view may be termed a ‘theory then research’ view. Other persons adopt a composite view—‘theory then research’ preceded by ‘research then theory.’ Under this view, stages of research activity are distinguished: (1) exploratory, (2) descriptive, and (3) explanatory (see Reynolds, 1971, pp. 154-155). *Exploratory research* is an early stage during which the researcher informally gathers observations with respect to some phenomena. The usual output of this research stage would be ideas suggestive of possible associations between variables, and these ideas become the focus for the next stage of research. *Descriptive research* is the stage during which patterns that were suspected based on the former exploration are carefully described with the goal of developing empirical generalizations. *Explanatory research* is the stage at which explicit theory is developed (to explain the previously identified empirical generalizations), tested and then recursively reformulated. Thus, as the researcher moves across these three stages — from exploratory to descriptive research and from descriptive to explanatory research — the degree of scientific understanding increases.

Notably, the degree of one’s scientific understanding can increase regarding phenomena that occur, phenomena that ought to occur, and phenomena that promote convergence between the former two. Research with these three foci often are labeled descriptive research, normative research, and prescriptive research, respectively. Aspiring audit researchers might be able to improve their mental model of research by reference to Exhibit 1. In it, we have depicted the stages reflecting the degree of scientific understanding as the y-axis (exploratory, descriptive, and explanatory) and the foci of scientific understanding as the x-axis (descriptive, normative, and prescriptive).

In our view, it is appropriate to make at least a couple of statements regarding this exhibit. Generally, a stream of research matures as the scientific basis increases for conducting studies in the exhibit’s northeastern sector. The scientific basis for conducting prescriptive studies, however, critically depends on whether the existing degree of understanding has attained the ‘explanatory stage’ for both descriptive and normative research. Prudent researchers are justifiably reluctant to make prescriptions based on an incomplete understanding of the causal relations that govern the auditor’s behavior or the behavior that the auditor should exhibit. Similarly, they are reluctant to specify the behavior that the auditor should exhibit unless they have a rich understanding of the auditor’s (often conflicting) objectives and their antecedents.

**Exhibit 1: A Basic Typology of Research**

<table>
<thead>
<tr>
<th>Degree of Scientific Understanding</th>
<th>Exploratory</th>
<th>Descriptive</th>
<th>Explanatory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foci of Research</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Descriptive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normative</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prescriptive</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Numerous research methods could be used to study auditor’s judgments and decisions, and the foci of research (see exhibit 1) generally does not constrain one’s choice of research method. Certain research methods, however, are well-suited for attaining the highest degrees of scientific understanding. In particular, the **experiment** is one of the most powerful of these methods for helping one test and refine theories about causal relationships. Relative to other research methods, experimentation allows the researcher to penetrate the farthest into the explanatory stage of scientific understanding.

We use the term experiment to mean a method of inquiry in which the researcher *randomly assigns* participants within a controlled setting that allows reproduction of some phenomenon (e.g., an individual’s decision process), *actively manipulates* antecedents hypothesized to affect the nature of the phenomenon and then makes various measurements of the phenomenon, often measurements that could not be made in a natural setting (Reynolds, 1971, p. 156). In combination, *random assignment* and *active manipulation* enable the researcher to make causal inferences about relationships that underlie observed phenomena. They do so by inoculating the research against several potential sources of invalidity, especially invalidity due to uneven distribution and influence of factors other than those manipulated, measured or controlled by the experimenter.

One alternative to experimentation is **individual observation** while another alternative is a **survey**. Both of these methods empower the researcher to advance their scientific understanding of phenomena from the exploratory to the descriptive stage, but neither afford the researcher with the ability to proceed far in the explanatory stage. If one directly observes some phenomenon in a natural setting and attempts to document an unbiased description of the observed phenomenon, one is using individual observation. If one measures individual characteristics of people or social systems (usually based on a sample), one is surveying.

Earlier, we stated that experiments help researchers build or test theories. We use the term **theory** to mean ‘...a set of interrelated constructs (concepts), definitions, and propositions that present a systematic view of phenomena by specifying relations among variables with the purpose of explaining and predicting phenomena’ (Kerlinger and Lee, 2000, p. 11). Further, we note that theories can take several forms. For example, Reynolds (1971) identifies three forms of theories: (1) set-of-laws (i.e., a set of well-supported empirical generalizations), (2) axiomatic (i.e., an interrelated set of definitions, axioms, and propositions (derived from axioms) often expressed in mathematical terms); and (3) causal processes (i.e., a set of definitions, conditioning statements describing when the processes are expected to occur, and causal statements that identify the effect of one or more variables on other variables).

Importantly, theories specify the relationship among variables, thereby providing a basis for the researcher to identify the factors (i.e., **independent variables**) that should be examined to discern an impact on other factors (i.e., **dependent variables**). Independent variables are either manipulated or measured. In an experiment, the researcher manipulates (i.e., varies across levels) at least one of the independent variables. In contrast, in a **quasi-experiment**, the independent variables are measured (sometimes called **attribute variables**). Researchers use measured variables when it is not possible or it is impractical to manipulate a variable. For example, human characteristics (e.g., gender, economic status) generally are established well in advance of the time a person would participate in an experiment.

The two most fundamental types of validity often distinguished by experimentalists are: (1) internal validity and (2) external validity (see Cook and Campbell, 1979). We use the term **internal validity** to refer to the extent to which one can attribute observed dependent variable effects to differences in levels of independent variables. While internal validity is the *sine qua non* of experimentation (Aronson et al., 1998), external validity also is important. **External validity** refers to the extent to which an inferred cause-and-effect relationship can be generalized beyond the experimental specifics (e.g., setting, tasks, participants).

Internal validity is strengthened by **random assignment**, a hallmark of experimentation. We use the term random assignment to mean assignment of participants or objects to the cells of an experimental design in such a way that every participant or object has an equal chance of appearing in each cell of the design (see Kerlinger & Lee, 2000, p. 170). The audit
researcher’s goal is to spread participants with varying characteristics approximately equally among the design cells so that extraneous factors (i.e., factors other than those manipulated) will have an equal effect across the cells. Randomly populating each cell in the research design with a sufficient number of participants allows the researcher to rely on the law of chance. The law of chance permits the researcher to focus on the causal role of the independent variables and to deliberately ignore alternative possible causes of variation within dependent variables (Fisher 1971).

In the absence of random assignment, one must engage in quasi-experimentation. In quasi-experimentation participants appear within cells of the research design in accord with manifestations of variables that already have occurred (e.g., an auditor’s rank has been determined prior to the quasi-experiment). There is, therefore, no assurance that participants with varying characteristics will be spread approximately equally among the design cells. Consequently, extraneous factors could have unequal effects across the cells.

RESEARCH TRAPS

In this section we identify and discuss four inter-related research traps. For brevity and emphasis our list of research traps is by no means exhaustive. Indeed, we selected traps that reflect the knowledge we have gleaned from first-hand experiences conducting our own studies and navigating them through the journal publication process, supervising doctoral students, and serving as journal referees. We have chosen not to cover research traps that are directly related to topics covered in other papers in this same issue (e.g., construct validity, experimental confounds and between-subject vs. within-subject research designs are discussed in Trotman, 2001).

In our experience, the traps we discuss in this section fairly regularly ensnare researchers, especially less experienced researchers, and the damage these traps do usually is pervasive and serious enough to mortally wound a research study. On a more positive note, we have found that among the best ways for a researcher to avoid such traps is to develop a deep understanding and appreciation for the concepts and definitions we presented in the preceding section of this paper.

The Rush to the Explanatory Research Stage Trap

Most auditing research, especially studies published in the better scholarly journals, falls into the explanatory research category. Consequently, theory-testing research is of relative high visibility. It is our experience that, because of its relative lack of visibility, many aspiring scholars do not sufficiently appreciate the full import and extent of the effort required at the pre-explanatory research stages. The indicia to this type of failure are well known to journal referees (e.g., manipulating the wrong variables, setting manipulated variables at uninteresting levels, manipulating variables so that they encompass multiple constructs with material qualitative differences, and failing to control extraneous variables).

The common antecedent to these problems is that the theory being tested (or its implications) is (are) insufficiently developed and thus, the myriad of required pre-experimental judgments and choices are insufficiently informed by theory. As discussed earlier, a primary goal of the exploratory and descriptive phases of research is to provide the scholar with a basis for articulating a reasoned theory during the explanatory research phase. The researcher then determines implications of the theory and expresses them as hypotheses. Such hypotheses effectively are propositions that can be subjected to empirical testing.

Importantly, researchers use theory to guide which factors to include as independent variables in the experiment as well as which factors to exclude. The researcher also uses theory to determine the conceptual levels at which each independent variable should be manipulated or, in the case of measured independent variables, the criteria for inclusion in the study (e.g., participants who have at least a designated level of specialized-industry experience). The researcher again relies on theory to, at least preliminarily, determine the number of participants that are likely to be necessary to be able to have a good chance of detecting the hypothesized relations, should they exist. While this list of key preliminaries to experimentation
that are informed by theory is fairly substantive, it is far from complete (e.g., dependent variables also are determined based on the theory).

A final observation for this research trap is that we believe careful pre-explanatory effort can help ensure theoretical continuity in the audit judgment and decision making literature. Audit tasks and processes are changing rapidly, just as change in business ecologies is exhibiting high velocity. While these changes may warrant expanding or changing the nature of theories that researchers bring to bear or construct themselves, they frequently will warrant new operationalizations of theoretical constructs (Bell et al., 1997). Pre-explanatory effort can help discern whether either or both of these possibilities are warranted.

Suppose, for example, a researcher was interested in how auditors integrate multiple pieces of evidence to judge whether accumulated evidence is sufficiently persuasive for rendering an opinion. Historically, some researchers have used belief perseverance theory to guide examinations of this issue. Unless future researchers were to study and recognize the dimensions along which the evidence gathered by today’s auditors, many of whom adopt a strategic-systems approach to auditing, structurally differs from and is similar to the evidence gathered by the ‘transaction-cycle, risk-based’ auditors of the 1980’s-mid-1990’s, the theoretical implications of much of the extant audit literature would be forfeited. In the absence of conducting such a pre-explanatory-stage analysis, the researcher might unnecessarily conclude that a theoretical discontinuity in the audit research is warranted (e.g., discard belief perseverance theory as a guide to his/her scholarly inquiry).

In our opinion, therefore, it is impossible to overstate the importance of theory as there is no substitute when it comes to developing an informed basis for making these many, key design choices. But, as noted earlier, rich and sound theory is not developed by accident. Rather, rich and sound theory results from rigorous and systematic processes like those contemplated within the exploratory and descriptive research stages.

The Mundane Realism Trap

Mundane realism refers to the extent to which an experimental setting and task mimic the participant’s real-world experiences (Swieringa and Weick, 1982). Despite the appeal of making experimental materials look like or incorporate real-world stimuli, mundane realism generally is unnecessary and insufficient for internal validity or external validity. In some forms, it can undermine a study by distracting participants’ attention away from the primitive independent and dependent variables of interest. Indeed, we generally become concerned when scholars fixate on the level of an experiment’s mundane realism.

In an experiment, the researcher effectively creates and places the participants in a setting that by design is a simplification of the real world in which auditors operate. The task that the experimental participant faces also is a simplification. The main advantage but also, at least as seen through the eyes of the relatively uninhibited, main disadvantage of such an experimental setting and task is that they are simplifications. Let us first consider the disadvantage perspective. Persons adopting this view generally contend that the utility of an experiment is a function of the ability to generalize results. In turn, they typically argue that the generalizability of experimental results is inversely related to the extent to which the experimental setting and task are simplifications of the real world. They customarily develop a laundry list of factors present in the real world but absent from the experiment. For such persons, discordance between the real world in which the audit judge or decision maker resides and the experimental setting and task diminishes the value of the experiment.

Proponents of the advantage perspective primarily focus on the generalizability of theory, not results. Consequently, they view experimental settings and tasks as being analogous to mathematical models, both of which, by definition, are purposeful simplifications. The utility of such experimental settings (much like mathematical models) resides in what can be learned from manipulation of factors, under tightly-controlled conditions, that theory suggests are important. By tightly-controlled conditions, at a minimum, we mean conditions of random assignment, researcher control over factors that
theory suggests are extraneous (i.e., elimination or measurement of extraneous factors), and researcher-determined (i.e., manipulated by the researcher or measured) exposure to theoretically important factors. Such tight control empowers the researcher to separate out the influence of potentially theoretically relevant factors that co-occur in the real world as well as to test the import of factors that do not co-occur in the real world but theory suggests to be of potential import. Consequently, from this perspective, simplification (and even artificiality) of experimental settings and tasks is (are) not a threat to the value of an experiment, but actually is part of what creates value in an experiment.

We fear that this message increasingly is not well understood by some audit researchers as we are observing, in recent years, an elevated tendency for researchers to, in effect, trade off internal for external validity. Admittedly it often is easier to enhance the mundane realism of an experiment than it is to enhance the internal validity of an experiment. In our view, for virtually all situations, attempts to elevate external validity (via enhanced mundane realism or otherwise) at the expense of internal validity are a grave mistake. In fact, it has been argued that internal validity is a necessary condition for external validity (Cook and Campbell, 1979). In the limit, in the absence of internal validity, one learns nothing about one’s theory which, in turn, means that there is nothing to generalize.

We also worry that aspiring audit researchers who fall prey to the mundane realism trap mistakenly presume that experimentation that lacks mundane realism cannot ultimately contribute to the more primitive objectives of understanding and improving auditor’s real-world judgments and decisions. In fact, all discussion of the extent to which experimental findings regarding causal relationships are generalizable involves inductive reasoning. By testing theoretically-derived hypotheses, experimentation helps organize and constrain the amount of inductive reasoning that the researcher must conduct to have reasonable confidence about generalizability. In the absence of theory and valid experimental tests thereof, questions of generalizability never could be addressed satisfactorily as inductive reasoning has no natural bounds. One can always imagine yet another context or yet another set of participants, as the passage of time, for example, arguably changes both.

The Measured Variable Trap

Earlier, when defining key terms, we noted that experimentation is characterized by several features, but most notably by random assignment and active manipulation. Subsequently, we noted that these features were central to high internal validity. We also noted, however, that researchers sometimes are unable to achieve random assignment and/or must employ measured variables. When that is the case, researchers bear an added burden if they are to be able to confidently attribute variation in the dependent variable to variation in their independent variables. When researchers do not sufficiently address this burden, they fall into the measured variable trap. Unfortunately, it seems that some audit researchers do not fully understand the difference between real and quasi-experiments.

At best, when measuring an independent variable, let us say X, together with one’s dependent variable, Y, in the absence of active experimental manipulation, one can conclude only that a correlation exists between the two measured variables. One cannot, however, equate empirical regularities with causation. Even if one were to presume a causal relation were to exist, one could not reasonably infer the direction of causation. Finally, the absence of active manipulation allows for the possibility that a third variable (or sets of variables) jointly cause X and Y. Because active manipulation helps attain the degree of control needed for scientific inference, the researcher should be creative when trying to identify ways to manipulate independent variables that initially appear to be non-manipulable (Kerlinger and Lee 2000).

A particularly dangerous version of the measured variable trap sometimes occurs when a researcher attempts to actively manipulate an independent variable and includes post-test manipulation checks. The measured variable trap is sprung when the researcher observes no reliable association between the actively manipulated variable and the dependent variable, but a reliable association is observed between the manipulated variable and the manipulation check responses as well as between the manipulation check responses and the dependent
variable. The temptation arises to characterize
the manipulation check response, ex post, as a
valid proxy for the manipulated independent
variable and infer causality in describing its
relation to the dependent variable. One cannot,
however, justify describing such an association
in causal terms for reasons mentioned in the
preceding paragraph. As an example, pre-exper-
imental determinants of the participants’
responses to both the manipulation check
question and the dependent variable could exist
that are unrelated to the construct that the
researcher attempted to operationalize by
manipulating the independent variable.

To illustrate, suppose a researcher hypothe-
sizes that task-specific experience improves task
performance. The researcher decides to use an
abstract task in which participants try to
complete an incomplete sequence of symbols
(e.g., $\bigoplus \bigotimes \cap \bigcap \bigcirc \bigcirc$). The researcher manipu-
lates the amount of task-specific experience via
the number of practice tasks that subjects
complete. After the manipulated number of
practice sequences, participants complete a
series of test sequences. After the test sequences,
participants respond to two manipulation check
questions. One question asks participants to
draw all the symbols they recall as having
finished a sequence. Another question asks par-
ticipants to guess how many (or the proportion
of) sequences they correctly answered.

The researcher observes no association
between task-specific experience and perform-
ance, so the hypothesis fails. The first
explanation the researcher explores is a failed
manipulation check. The researcher, however,
oberves a positive association between the
number of practice sequences completed and
responses on the two manipulation check
questions (i.e., participants with more task-
specific experience recall more symbols and
believe they correctly completed more
sequences). Next, the researcher observes that
the responses on the manipulation check
questions are positively associated with per-
formance. The researcher correctly deduces that
one cannot rule out a successful but fairly weak
experimental manipulation given this pattern of
positive associations.

At this juncture, however, the researcher needs
to avoid falling into the measured variable trap.
The manipulation check questions may be noisy
proxies of task-specific experience, but it may be
more correct to think of the manipulated
variable as a noisy proxy of an entirely different
construct that drives performance (e.g., analytical
thinking ability). Consequently, the
researcher must conduct further research.

Of course, even after trying to identify ways
to manipulate independent variables, the
researcher sometimes will conclude that reliance
on measured variables is necessary. One cannot
readily manipulate factors such as personality
traits or forms of intelligence (e.g., analytical
thinking ability), and one cannot readily observe
latent thought processes or subconscious deter-
minants of judgment and decision making.
There are, however, experimental design and
statistical steps researchers can take to make the
best of these difficult situations (see, e.g.,
Schwarz et al., 1998; Kerlinger and Lee 2000).

As one example, when multiple independent
variables are of interest, a researcher can pre-
experimentally measure some independent
variables and randomly distribute the levels of
the manipulated independent variables across
matched levels of measured independent
variables. Suppose that the researcher from the
previous illustration now is interested in task-
specific experience after explicitly controlling for
analytical thinking ability. A reasonable follow
up design would help discriminate from a weak
manipulation in the first experiment and alter-
native accounts for its pattern of findings. In the
follow-up experiment, he or she could increase
the magnitude of the difference across manipu-
lated levels of task-specific experience and use a
valid measure of analytical thinking ability to
balance participants across these two manipu-
lated levels. If this approach were adopted, the
researcher may once again fail to observe
support of the hypothesis, but at least he or she
would be unconcerned about analytical-
thinking ability causing the same pattern of
findings.

**The Audit Practitioner as Participant Trap**

We have saved what probably will be our most
controversial research trap for last. There is a
tendency, especially for aspiring researchers, to
assume that practicing auditors always make
the best participants for audit judgment and
decision experiments. Consistently, there is a
tendency to assume that the more experienced
the participant, the better. And, lately we have
noted an increasing number of studies for which the apparent motivation for the study was nothing more than to learn how experienced auditing practitioners would perform an experimental task. In our view, audit researchers who think along these lines have fallen into one of the more insidious of the audit research traps.

Numerous papers have been published over the past thirty years discussing the validity of accounting and auditing experiments in which students are the participants (e.g., Dickhaut et al., 1972; Abdel-khalik, 1974; Ashton and Kramer, 1980; and Walters-York and Curatola, 1998). It is not our goal to repeat these discussions. Further, like the extant papers, we do not produce flat rules concerning the use of student participants. Rather, our goals are to raise fundamental questions about some of the criticisms that have been attached to the use of student participants, and especially to suggest a different, more productive, way to frame the issues to be considered in connection with the student-participant issue in the context of audit judgment and decision research.

We begin by joining the chorus of researchers who say that participants should be chosen for an experiment on the basis of an ensemble of experiment-specific factors. While such a statement is hard to refute, it is equally difficult to ascribe much practical value to it. Moreover, in many respects, as typically framed, this guidance has created the wrong default condition. Specifically, unless the purpose of the study is to investigate an issue that would require student participants, audit judgment and decision researchers default to practitioner participants, and for that matter, the most experienced participants that can be attracted to the study. Indeed, the position seems to be that using practitioner participants ensures the external validity of the study (such a position is fallacious because of the inductive reasoning argument we discussed earlier under the mundane realism trap). It is worthwhile, however, to closely consider the conditions under which external validity actually will be harmed by usage of student participants. In our view, such conditions are rather limited.

Of the various readily discernable, negative views about external validity and student participants, we consider two that have had, in our opinion, the greatest impact on auditing scholars over the past few years. The first of these arguments is that, by necessity, student samples are convenience samples and thus, are not representative of any real-world population. In our experience, however, irrespective of whether one is employing student or practitioner participants, the audit researcher rarely, if ever, can employ a strict and formal random sampling plan. Human subjects requirements at universities, for example, usually mandate that both student and practitioner participants be given the option of not participating as well as the option of withdrawing from an experiment once participation has begun. The problem can be even more acute for practitioner participants, as they have many, often competing, demands on their time. Further, researchers often can observe how characteristics of students who participate differ from those who do not participate at low cost. For example, a researcher could compare the grade point average of participants versus the population of students from which student participants were drawn. Still, attrition is a non-trivial issue irrespective of whether one is dealing with student or practitioner participants. Such attrition poses a serious threat to participant representativeness irrespective of whether one is considering student or practitioner participants. Consequently, this form of non-representativeness is unlikely to provide a basis for preferring practitioners over students as participants.

The second of these arguments is that students differ systematically from the populations in which the researcher really has an interest. In a somewhat more precise form, this argument is that audit students and practitioners differ in terms of some characteristic (either omission or commission) that is of theoretical interest to the researcher. In our view, the last few words in this re-stated argument are of critical import. Their import arises because they reveal that the fundamental issue being raised concerns internal validity more than external validity. To elaborate, let’s assume that auditing students and auditing practitioners differ on some dimension, say average age. Now, let’s consider the following question — how much of a threat does such a difference pose to the external validity of the study?

Some persons seemingly would want to answer the question with the response that a significant threat is present. We believe, however, that a far superior answer is that we do not yet
know the extent of the threat but we do know that whatever its magnitude, its theoretical importance rests much more squarely on internal than on external validity. Specifically, we contend that to appropriately answer the question, one needs to know the research question of the study and, in particular, if a theory exists for expecting that the factor differentiating audit students and practitioners would interact with at least one of the factors that the researcher’s theory tells him/her is a causal agent. Moreover, to conclude that a causal relation observed within experimentation is of no concern in the natural ecology, the expected interaction between the factor differentiating audit students from practitioners would have to feature another key property — that of mitigating or eliminating the causal relation. Thus, even if audit students and practitioners were to differ in terms of age, there would be no threat to validity, if age does not interact with one of the elements of the theory being investigated. Even if age were to interact with a theoretically relevant factor, the threat would be to one’s ability to attribute the change, in the dependent variable, to the manipulated or measured differences in the independent variable. Such a threat is to internal, not external, validity.

Some may object to this line of argument and prefer employing auditors as participants when testing a theory that involves constructs such as domain-specific knowledge and task-specific experience. By using auditors to test such a theory, however, experimentation generally is not viable. In the field, one’s ability to employ random assignment and active manipulation regarding specialized knowledge or task-specific experience is quite limited. While there are ways of dealing with measured variables, as discussed and alluded to in the earlier section on The Measured Variable Trap, none provide the researcher with the level of control inherent in sound experimentation. We believe it would be profitable if researchers would contemplate ways that they could manipulate such constructs within a laboratory setting. Students—participants’ domain-specific knowledge, for example, could be manipulated by randomly assigning them to pre-experimental training sessions that feature systematic and controlled variations in instructional or experiential learning opportunities.

Upon reflection, the researcher may come to realize that relatively few characteristics cannot be manipulated within laboratory settings featuring students as participants. This realization, coupled with the reality that practitioner participants are a limited and costly resource, lead us to conclude that practitioners should not be used unless necessary. In assessing necessity, the researcher should judge carefully the theoretical bases for anticipating specific forms of interactions to exist between factors that differ across audit student and practitioner populations and the other factors of theoretical import. If such interactions cannot reasonably be predicted, we believe that students should be thought of as being the ‘default’ condition for experimental participants. Employment of this participant protocol could significantly alter the distribution of participants used in the modal audit judgment and decision experiment.

CONCLUDING REMARKS

Contributing to the audit literature requires one to become knowledgeable about extant scientific frontiers and well-versed in the design and execution of research. While the scientific frontiers change over time, we have found that the research traps that ensnare audit scholars remain remarkably constant. This paper, we hope, will help aspiring researchers avoid these common research traps by heightening their awareness of them and by enabling them to understand the value of regularly conversing with seasoned audit scholars who understand their symptoms and implications. Throughout this article, we have tried to avoid flat statements, but we have adopted positions that may challenge the conventional thinking of some. Our positions emphasize careful selection of theory, the perils of exchanging internal validity for external validity, two hallmarks of experimentation (manipulated variables and random assignment), and consideration of students as ‘default condition’ participants. When contemplating these positions, the aspiring researcher ought to take a holistic viewpoint instead of fixating on one position at a time. Tradeoffs are inherent in experimentation, and, in hindsight, it almost invariably seems to be the case that a better version of the experiment could have been run. A reasonable objective for the aspiring researcher would be to
become familiar with each research trap and to develop an appreciation of the implications of being ensnared by it. If the researcher were to do so, and if his/her research also were well motivated in light of the dynamic scientific frontier, the chances of the research informing the scholarly auditing literature would be greatly enhanced.

NOTES

1. The reader should not confuse the acquisition of a scientific understanding with everyday belief revision. Decision makers justifiably revise their beliefs without engaging in experimentation or other forms of scientific inquiry. The methods of science, however, emphasize the disciplined creation and advancement of knowledge (i.e., theory) more than the revision of beliefs. Certainly, in evaluating phenomena, the threshold of decision usefulness often falls below the threshold of making a scientific contribution.

2. There are, of course, exceptions to this general tendency. For example, we would characterize Koonce et al., (1995) as a pre-explanatory stage research study.

3. As others have noted, experimentation can help inform the understanding of those whose business it is to appraise whether it would be desirable for such factors to occur in the real world (see, e.g., McDaniel and Hand, 1996).

4. We discuss measured variables and non-random participant assignment together because measured variables often are the reason that non-random assignment was employed. For example, human characteristics (e.g., gender) sometimes are the reason that participants effectively self select into a cell of the experimental design.

REFERENCES


AUTHOR PROFILES

Associate Professor Mark E. Peecher is a Deloitte & Touche Teaching Fellow at the University of Illinois at Urbana-Champaign. Professor Peecher’s experimental auditing judgment and decision research has appeared in the top scholarly journals in the field.
Professor Solomon is the Robert C. Evans Endowed Chair in Commerce and Professor of Accountancy at the University of Illinois at Urbana-Champaign. He is a former President of the American Accounting Association Auditing Section and twice received the American Accounting Association Auditing Section’s award for Chair of the Outstanding Auditing Dissertation.