Design Considerations for Research on Analytical Procedures

Stephen K. Asare*1 and Arnold Wright2
1University of Florida, USA
2Boston College, USA

This article discusses research design considerations for conducting behavioral research on auditors’ performance of analytical procedures (APs). With the trend in practice towards increasing reliance on APs, it is essential that auditors are proficient in completing such tests. Therefore, research to understand and improve auditors’ performance of APs is important.

Once an unexpected fluctuation is identified, APs involve three phases: generation of plausible hypotheses (likely causes); gathering evidence to examine plausible hypotheses; and identification of the most likely cause followed by appropriate follow-up actions. Although prior research has focused on these phases in isolation, they are, in fact, interrelated. Important research design issues and trade-offs for each of the phases of APs are discussed. For instance, in examining hypothesis generation there are choices as to the amount and nature of case background information, number of ratios or accounts to explain; ex-post evaluation of the quality of the hypothesis set; and instructions on number of causes that account for the fluctuation. Significant considerations are identified for making informed decisions among design choices.

Key words: Analytical procedures, behavioral research methods, hypothesis evaluation, hypothesis generation, information search.

Auditors routinely perform analytical procedures (APs) to identify and investigate unexpected fluctuations in the financial statements. In investigating unusual fluctuations, auditors typically generate or inherit potential causes or hypotheses and gather evidence, varying in diagnostic implications, to evaluate the veracity of the hypotheses.

In essence, these tasks are diagnostic in nature and are important because of the ramifications that they can have on audit efficiency and effectiveness. For instance, failing to correctly diagnose a misstatement that materially overstates a company’s earnings can lead to misleading financial statements with the concomitant misallocation of resources and litigation. Efficiency is also a concern since hypothesis generation, evidence gathering, and hypotheses evaluation are costly from both a cognitive and financial perspective.

*Correspondence to: Fisher School of Accounting, College of Business Adm, University Of Florida, Gainesville, FL 32611-7166, USA, kwaku@ufl.edu
AP investigation entails three interrelated phases (Koonce 1993; Asare and Wright 1997a): 1) generation of plausible hypotheses (i.e., likely causes) to account for significant discrepancies between expectations and realizations; 2) gathering evidence to test plausible hypotheses; and 3) identification of the most likely cause followed by appropriate follow-up actions (e.g., correct the error, revise disclosures, etc.). The purpose of this paper is to discuss and evaluate prevalent research methods that have been used to examine the three phases, along with critical design issues. In so doing, we wish to sensitize the reader to the design tradeoffs inherent in this line of research and their associated limitations. Also, such a discussion should be helpful to potential researchers contemplating a choice of research method. To provide context for the evaluation of the design issues, we also provide a brief overview of the various substantive issues for each phase. The next section focuses on hypothesis generation, followed by information search, hypothesis evaluation and hypothesis selection. The final section contains concluding remarks.

HYPOTHESIS GENERATION / INHERITANCE

Once an unexpected fluctuation is identified, APs, typically, commence with hypothesis generation that, in turn, drives further information search and bounds the information search space. Therefore, this phase is critical since it can lead to a search of irrelevant evidence (efficiency concerns) or the exclusion of highly relevant evidence (effectiveness concerns). Following Libby’s (1985) argument that generating a large hypothesis set increases the probability of the set including the correct hypothesis, most subsequent research on APs has focused on conditions that can enhance or impair hypothesis generation. Based on this line of research, it has been found that task and industry experience (e.g., Libby 1985; Libby and Frederick 1990; Wright and Wright 1997; Solomon et al. 1999) and team deliberations (e.g. Ismail and Trotman 1995) can enhance the number of hypothesis generated while the presence of a client explanation (typically a non-error explanation) interferes with the generation of error causes (Hoch 1984; Bedard and Biggs 1991; Frederick 1991; Heiman-Hoffman et al. 1995) and leads to the generation of fewer hypotheses (Asare et al. 1998).

Researchers of the hypothesis generation phase have either employed a hypothesis-listing task or a process tracing approach. In both cases, participants are provided background information about a hypothetical client and pointed to a specific ratio or set of ratios that materially differ from expectations. They are then asked to generate potential causes to explain the difference(s). In the hypothesis-listing task, participants complete the task by writing down potential causes that become the basis for subsequent analysis. In the process tracing approach, participants are asked to think aloud while completing the task (verbal protocols). An audio recording of the experimental session is then subject to protocol analysis. Important design considerations, irrespective of method employed, include amount and type of background information to provide; number of ratios or accounts to explain; ex-post evaluation of the quality of the hypothesis set; and instructions on number of causes that account for the fluctuation.

Amount and type of background information

The amount and type of background information to be provided to the participants is an important design consideration. There are various options ranging from minimal background information (which leads to a very short case) to very extensive background information (which leads to a very long case). The option chosen should not be arbitrary but tied to the research goals as well as participants’ willingness to participate, which is increasingly becoming a binding constraint. If the research goals include evaluating the accuracy of the participants’ hypothesis set then it is important to have sufficient background information to give the participants an opportunity to develop a relevant mental model of the client. Such mental models are the foundation of expertise and explain why most firm manuals emphasize ‘learning the business and accounting/control systems’ of a client. As noted by Libby (1981), experts through their training and experience develop a complex associative memory that relates evidence to prototypes of problem solutions. In general, therefore, providing more
background information gets closer to the ‘real audit environment’ which makes claims of accuracy more tenable. While it is extremely difficult to prescribe the volume of information needed to make credible claims on accuracy, it seems that at a minimum some information on the client’s control architecture, business objectives, key risks, industry trends and financial statements are warranted. Further, the findings by Anderson and Koonce (1995) suggest auditors may frequently engage in quantification to determine whether a hypothesis can sufficiently account for the magnitude of the fluctuation. Therefore, information enabling quantification is desirable. Needless to say, the more background information that is provided, the longer the case and the more difficult it is to get participants’ cooperation.

Number of fluctuations to explain

The number of fluctuations that have been used in prior AP research vary from providing auditors with a specific ratio (e.g., explain the fluctuation in the gross margin ratio) to asking them to explain the patterns in a set of ratios and account balances. An example of the latter is the Bedard and Biggs (1991) case which required participants to consider six ratios and four account balances. Typically, these latter cases seed an unique error that requires participants to reconcile several discrepancies. While such cases are difficult to construct, their advantage lies in having only one hypothesis that can account for the discrepancies. In contrast, cases that focus on a specific ratio are easier to construct but more equivocal for evaluating accuracy. For instance, imagine the set of plausible causes that can account for a fluctuation in the gross margin ratio!

Ex-post evaluation of quality of hypothesis set

Researchers are interested in evaluating the quality of the hypothesis set and the extent to which this goal can be achieved is linked to the research design phase. If auditors’ performance will be evaluated based on the selection of a unique hypothesis then it is strongly recommended that a case based on an actual audit be used. Further, as noted, the case will be enhanced if it requires the reconciliation of several discrepancies. It is rather difficult to develop such a case and identify an unique cause; thus, researchers alternatively tend to evaluate performance based on the number of ‘plausible’ hypotheses generated.

If the number of plausible hypotheses is used as a benchmark, care must be taken in defining what is plausible. This is especially so when very limited background information has been given and the subject has been asked to explain a rather all-embracing ratio such as the gross margin ratio. Because of the interrelatedness of the items in the financial statements and the potential complexities of a client’s business environment, it is difficult to determine whether an hypothesis proposed by a knowledgeable person is, in fact, ‘implausible.’ Further, it must be made quite clear to auditors that their performance will be evaluated based on the ‘quantity’ of the hypotheses. This is important so as not to disadvantage participants whose strategy may be to generate a manageable, as opposed to a comprehensive, hypothesis set with the assumption that the set may be broadened if necessary. Further, in practice, auditors may generate only a few hypotheses or only a single hypothesis particularly for very well known clients. Also, as suggested by Bhattacharjee et al. (2000), there may be some efficiency in parsimony. In any event, with only one prior study (Bhattacharjee et al. 2000) on this issue, the empirical linkage between the quantity of hypothesis and performance has yet to be established (see Asare and Wright 2001).

Instructions on number of causes that account for the fluctuation

With few exceptions, studies of hypothesis generation ask auditors to assume that there is only one cause for the unexpected fluctuation. This assumption is necessary to rule out a situation where participants hypothesize that several immaterial errors in various accounts have aggregated to the material fluctuation. Yet, this assumption alters the task in a fundamental way. First, seldom, if ever, will a participant in a real situation know how many causes have led to a fluctuation. Second, if participants in their natural ecology perform this task with an open mind on the number of causes that could lead to a fluctuation, then the experiment forces them to perform a task that is contrary to their accumu-
lated experiences. The effect of such a change in the task is largely unknown. An alternative, surprisingly less utilized approach, is simply to ask auditors to investigate the cause of the discrepancy without any restrictive assumptions on the number of causes, as was done in Bedard and Wright (2000). We wish to emphasize that asking subjects to assume there is only cause is not problematic per se. However, to the extent that such an assumption may not be reflective of auditors’ experiences or beliefs regarding a particular experimental case setting, it calls for caution in drawing inferences about performance.

**Coding Hypotheses**

Both hypothesis listing and protocol analysis require an independent coding of participants’ hypotheses. Coding necessarily involves subjectivity making it necessary that multiple coders are used to assure coding reliability. Auditors often use technical terms and/or are unclear about their hypotheses. The researchers should agree on a coding scheme before the actual coding begins. Ideally, the coders should have a good level of auditing experience, familiarity with the task, and be unaware of the experimental hypotheses or conditions to which participants are assigned. A high level of coding agreement is essential to ensure that the findings are reliable. Prior protocol studies, for instance, have reported levels of about 80% (Bedard and Biggs 1991). If possible, disagreements must be resolved, preferably, by a third independent party. However, for practical reasons, it may be difficult to get another independent party. In those instances, both of the researchers can resolve the disagreement as long as disagreements are noted in a way that does not reveal the experimental condition of the participants.

**INFORMATION SEARCH**

Subsequent to hypothesis generation the auditor would typically conduct audit tests (information search) to examine the veracity of potential causes. This phase is important because, even if the auditor has generated the correct cause(s), a limited or inappropriate search will likely lead to an incorrect conclusion. Further, if the auditor either self générates or inherits an incorrect cause, a confirmatory or truncated search can result in an improper conclusion (Kinney and Haynes 1990). Finally, a thorough information search provides an opportunity to compensate for inadequate hypothesis generation by obtaining evidence that may signal the correct cause(s) (Asare and Wright 2001; Bedard et al. 1998). Auditors may conduct a depth, breadth or balanced search strategy (Asare et al. 2000).

Research on information search has been limited. Bedard et al. (1998) find that hypothesis generation impacts information search. In particular, an incorrect management or self-generated explanation impairs the effectiveness of planned audit tests. Asare et al. (2000) report that accountability leads to an increase in the extent and breadth of testing but does not affect the depth of testing. The results also show that a time budget decreases the extent and depth of testing but does not affect the breadth of testing. Bhattacharjee et al. (2000) find that the size of the hypothesis set impacts the efficiency of search.

Research on this phase has employed an experimental approach where auditors are provided a case containing an identified material, unexpected fluctuation. Participants then plan an information search using one of two methods: choice of tests from a menu (e.g., Asare et al. 2000); or a listing of planned tests (Bedard et al. 1998).

**The menu driven approach**

A menu driven approach provides subjects with a defined list of tests or audit procedures. If a menu approach is employed, it is important to randomize the tests to mitigate potential order effects, e.g., a tendency to select those tests identified earlier. Further, there is the problem of which tests to include in the menu listing and how they are labeled. It is clearly not possible to include all potential tests that pertain to all potential hypotheses. Therefore, the researcher needs to conduct extensive pilot testing to identify the most likely, common tests that would be conducted for a particular type of fluctuation. Before a participant selects a test, it must be labeled or named to indicate what the test examines, e.g., ‘sales cutoff test.’ Again, care must be taken to ensure the label is descriptive and clear. If the research issue is whether auditors will select procedures to test a particular hypothesis such as a client explanation, the
Design Considerations for Research on Analytical Procedures

label must clearly indicate it will provide evidence with respect to that hypothesis. There is also the problem that a particular test may provide evidence that is pertinent to more than one hypothesis (e.g., receivable confirmations, to some degree, bear on the assertions of existence and valuation). Again, careful development and pilot testing are necessary to justify the linkage of a particular test with an hypothesis.

Test listing approach

With the menu approach, discussed previously, participants choose a test and then receive the audit evidence obtained. In the test listing approach auditors are asked to identify a set of tests and may not get evidence. If participants are asked to list tests, coding procedures must be planned. In this event, the researcher must be cognizant of all the issues that were discussed earlier under coding of hypotheses. In addition, if program ‘effectiveness’ is examined, there is the problem of developing a normative benchmark. The ‘quality’ of a set of tests in investigating a fluctuation is one of professional judgment and entails a number of underlying dimensions such as breadth (testing of likely causes), depth (appropriate extent for a test), persuasiveness (strength of tests), and defensibility. Further, audit efficiency is an important consideration. Prior studies (e.g., Bedard et al. 1998) have used a Delphi panel of experienced auditors to identify and weight important tests. However, this criterion has limitations (e.g., the validity of relying on the responses of a small set of auditors, appropriateness of weighting scales, etc.).

Simulating a realistic search environment

Information search, as noted earlier, is also affected by a number of variables present in the audit environment, including previously inherited or generated hypotheses, time budgets, accountability, and client/business risks. If these factors are not controlled or randomized, the external validity of the findings is likely to be reduced. For instance, allowing auditors to conduct testing with no time budget constraints is unrepresentative of practice. This is likely to lead to considerably greater testing than would actually be performed. If time budgets are introduced, there is the problem of establishing realistic levels, both for the investigation phase in total and for each individual test. To be reflective of practice, auditors may be allowed to terminate information search when desired and exceed the time budget. Time deadline pressures, however, are difficult to simulate in an experiment.

HYPOTHESIS EVALUATION

As audit evidence is obtained, the auditor must evaluate the evidence and revise his/her likelihood assessments of the potential causes. Prior studies indicate auditors have difficulty performing hypothesis evaluation (e.g., Jamal et al. 1995, 1997; Anderson and Koonce 1995), particularly when there are multiple hypotheses. Further, Bedard and Biggs (1991) find that auditors often fail to recognize the underlying pattern of the evidence and, thus, focus on hypotheses that explain a fluctuation in a single account rather than for related fluctuations in other accounts. This phase is important, since it guides the auditor’s decision as to which of the hypotheses is (are) the most likely cause(s) of the fluctuation and the attendant actions to pursue.

The research approaches that have been employed to examine hypothesis evaluation in auditing include likelihood rating (e.g., Asare and Wright 1997a), choice (e.g., Asare and Wright 2001) and process tracing (verbal protocol) (e.g., Mock et al. 2000). In the first approach, participants are provided several hypotheses and are asked to provide a likelihood rating for each hypothesis as they receive evidence. This allows the researcher to evaluate the impact of the evidence on the hypothesis that is directly implicated by the evidence as well as the competing hypotheses. In the choice approach, subjects are given a menu of likely hypotheses and asked to select the most likely cause after being provided with varying amounts of audit evidence. Therefore, the researcher can only track the most likely hypothesis after each evidence iteration. In process tracing studies auditors verbalize their thoughts about various hypotheses as they are provided or choose audit tests. This approach provides a rich trace of the decision process but entails significant costs (e.g., participant time, researcher coding of responses), resulting in reduced sample sizes.
Benchmark

An important substantive issue which, affects design choices, is the appropriate benchmark with which to compare revisions. The Bayesian school of thought suggests that auditors will update likelihoods via Bayesian calculus that presumes complementary revisions and additivity (i.e., summed probabilities approximate one). One implication from a design standpoint will be to provide auditors with response scales for each hypothesis but force the responses to be additive. However, Asare and Wright (1997a, 1997b, 1995) find that auditors revise assessments only for the hypothesis directly implicated by the evidence obtained rather than the entire hypothesis set. That is, auditors may use simpler cognitive strategies such as focusing on one hypothesis at a time. This finding suggests that a choice approach wherein auditors identify the most likely hypothesis has more external validity than a likelihood approach wherein all hypotheses are evaluated.

Further, an analytical study by Srivastava et al. (2001) finds that the Bayesian pattern of revisions is only appropriate under certain circumstances (e.g., a mutually exclusive, exhaustive hypothesis set), which may not be representative of many situations in audit practice and thereby not assumed by participants. Using a Bayesian framework, they show that the pattern of revisions depends on the assumed conditions of the hypothesis set and may vary from complementary to independent. Further, super-additivity, sub-additivity (less than one), or additivity (one) may be logical. In a verbal protocol study Mock et al. (2000) report auditors frequently describe assumed interrelationships (most notably, a positive correlation) between hypotheses (i.e., an hypothesis set that is not mutually exclusive) and/or potential other yet unidentified causes (i.e., non-exhaustive hypotheses). These findings suggest that to evaluate the appropriateness of revisions the researcher must determine participants’ framing of the hypothesis set and whether it conforms with that anticipated in the experimental setting. Prior experimental studies have attempted to frame the hypothesis set as mutually exclusive and exhaustive (e.g., Heiman 1990), but it is not clear participants internalize this framing given their experiences in practice. Ideally a manipulation check is desirable to determine that the experimental framing has been appropriately encoded.

Response scale

Another important issue to consider in research on hypothesis evaluation is how participants interpret the response scale. In the experimental studies it is assumed that responses are viewed as probabilities. However, the frequent finding of super-additivity may be because auditors are not accustomed to thinking in terms of probabilities and/or the assumed decision context is not well suited for precise probability assessments. Thus, responses may be more akin to belief functions than probabilities (Shafer 1976). Varying response modes and/or manipulation checks are desirable to determine that participants are responding on the scale intended by the researcher. Alternatively, as suggested by Srivastava et al. (2001), other paradigms may be explored such as the use of belief-functions (Srivastava and Shafer 1992).

Evidence attributes

Four attributes of the evidence itself are also likely to impact probability revisions: direction; strength; inference; and order. In terms of direction, with respect to a particular hypothesis evidence may be confirming, disconfirming, or neutral. Evidence strength may also vary from weak to strong, and an audit test may have inferences only for one hypothesis or implicate more than one hypothesis. To interpret the appropriateness of revisions, prior research has assumed the direction, strength, and inference of the evidence are encoded by participants as experimentally intended. However, manipulation checks are important (and/or pilot testing) to ensure that such assumptions are appropriate. Further, the order of tests may impact revisions, e.g., a ‘recency’ effect (Hogarth and Einhorn 1992). To mitigate this potential confounding effect, researchers should consider giving each subject a different random order of the tests or have alternative versions of the experiment where order is varied to statistically evaluate if ‘order’ impacts the pattern of revisions.
HYPOTHESIS SELECTION AND ACTIONS

The final phase of APs involves making a decision as to the cause(s) of the fluctuation to determine what actions should be taken, ranging from nothing further need be done (e.g., when the fluctuation is due to business conditions) to the need to propose an audit adjustment (material error) or go to a high level of management or the audit committee (e.g., a fraud). Decision performance in this phase is clearly very important, impacting audit effectiveness (if a material error is not identified and subsequent litigation ensues) and/or efficiency and client relations (if a material error is incorrectly identified resulting in additional unwarranted testing and client ill will). There has been very little research on the final decision phase.

The correctness of the auditor’s final decision is likely impacted by the joint effects of the previous phases of APs. Asare and Wright (2001) examine the links between hypothesis generation, information search and decision performance. The results reveal that both correct hypothesis generation and a balanced information search significantly impact decision performance. That is, neither on its own is a sufficient condition. Identification of the correct hypothesis, on average, does improve decision accuracy but often, nonetheless, diagnostic errors occur when information search is not thorough enough. Further, an incorrect hypothesis set cannot be fully compensated for by strong subsequent testing, which provides evidence that points towards the correct cause. Moreover, the results suggest that the situation most common in practice wherein auditors generate hypotheses and do their own search results in the lowest level of performance. These results suggest research should ideally entail multiple phases of the analytical procedure process to capture the interactive affects present. However, such an analysis is by its nature complex to design and involves additional, costly participant time. Alternatively, if a single phase of the process is studied, conclusions should be tempered with caution.

Bhattacharjee et al. (2000) examine the impact of the hypothesis set size on evaluation and decision performance. They posit that a large set (e.g., five hypotheses) may make the hypothesis generation and decision phases cognitively very complex and unwieldy, thus, resulting in poorer overall efficiency and effectiveness. On the other hand, a very small set (e.g., one hypothesis) may result in inefficiency, having to go through multiple rounds in considering additional hypotheses, and ineffectiveness in not considering a broad set of hypotheses. They find that a hypothesis set size of three is optimal, with greater efficiency and as good decision accuracy as the five hypotheses set condition. However, Bhattacharjee et al. did not consider the impact of hypothesis set size on information search. Their findings suggest the hypothesis set size requires careful consideration in the design of research studies. If auditors are allowed to self-generate hypotheses, the size of the set should be controlled for statistically in the analyses.

An additional research design consideration in examining decision accuracy is the nature of the seeded cause of the fluctuation. It is important to establish that either the pattern of cues present are uniquely associated with the cause (e.g., Bedard and Biggs 1991) and/or the audit evidence available clearly points toward the cause. As noted previously, many fluctuations (e.g., an increase in a client’s gross margin) could be the result of many potential causes (overstatements of revenues, understatements of costs of goods sold, or the net effect of multiple causes). Further, to test the relative importance of potential explanatory factors on decision performance the cause should not be so straightforward and simple to be unrepresentative of practice and require little skill in conducting the various stages of analytical procedures. Therefore, the choice of the nature of the fluctuation and the seeded cause(s) require careful development and pilot testing. Ideally, the fluctuation(s) should be material and the underlying cause not easy to identify (i.e., represent detection risk of some concern). Further, the pattern of available evidence, if gathered and well considered, should lead a reasonably experienced auditor to be able to identify the underlying cause(s) through evaluation of the evidence and quantification, if needed (Anderson and Koonce 1995).

Finally, task experience and industry knowledge are likely to affect performance on the accuracy and efficiency of auditors’ final decisions in performing analytical procedures, as suggested by the findings of several studies.
on hypothesis, generation (e.g., Wright and Wright 1997; Bedard and Biggs 1991; Libby and Frederick 1990; and Libby 1985). Thus, experience should be controlled for either statistically or through the research design. Participants should have at least the minimal task experience necessary to arrive at a final decision.

COMPUTER OR INTERNET-BASED APPLICATIONS

While disaggregating the AP investigation task into its components for detailed study has yielded many insights, it appears that concurrent examination of these phases is a particularly promising avenue for future research. Historically, concurrent examinations have been rare partly because of the importance of understanding the component parts but also because of the cumbersomeness of using manual technology to design such an experiment. With the recent widespread use of computers in auditing and the availability of the internet, this concern should be less formidable. Indeed, a number of recent studies on analytical procedures have already employed a computer or internet-based administration (for example, see Asare and Wright 1997a,b, 1995 and Asare et al. 2000; Bhatarchjee et al. 2000). There are a number of advantages of such an approach, including:

- the ability to examine multiple phases of APs concurrently;
- the ability to incorporate strong research process controls, e.g., preventing participants from going back and altering previous responses, controlling the order and nature of information, and randomizing order conditions and treatments;
- ease of sending and returning participant responses, potentially improving the response rate; and
- the ability to obtain detailed process information such as evidence search and detailed participant time tracking.

Of course, with alternative research approaches there are always potential tradeoffs and concerns. First, the researcher must ensure the application is widely compatible with participants’ software and hardware capabilities. Second, since the participant may complete the exercise when desired, there may be greater concerns about internal validity than when the exercise is done under controlled conditions, e.g., in the presence of the researcher at a training session (e.g., did the desired individual actually complete the exercise? Were decision tools or other individuals consulted?).

Third, the application must be ‘user friendly’ in allowing the participant to perform the task as closely as possible as it would be done in practice. For instance, auditors often go back and forth between working papers so the application should provide such interfaces. Also, there is the concern that numerous, detailed computer screens are tiresome and may be more difficult to analyze than usual working papers. This creates the worry that the presentation mode may alter the normal decision process and/or reduce the response rate. Therefore, the researcher should be careful to avoid excessive detailed screens and may consider asking the participant to print out some detailed schedules or information, resulting in a combination ‘paper based’ and computer based design (e.g., see Asare and Wright 1995).

Fourth, there are additional costs and time required for programming. Finally, with computer or internet-based applications there are concerns about protecting anonymity and potential viruses. Computer and internet files may be traced back to the sender. If anonymity is assured, the researcher must be careful to ensure that the identities of participants are never revealed or considered in the reporting of the results or alternatively that identities are removed from files. The researcher should be careful to utilize recent virus protection packages to minimize the risk of either sending or receiving viruses.

CONCLUSION

Analytical procedures (APs) are widely used in auditing practice due to their potential to significantly enhance efficiency and effectiveness. Further, APs are likely to be relied on to an even greater extent in the years to come as auditing firms have shifted from a primarily substantive testing approach to one utilizing holistic, business risk, knowledge-based audit strategies (e.g., Bell et al. 1997, KPMG; and Winograd et al. 2000, PWC), which have significant potential to more fully utilize APs. This article identifies important research method and design issues to consider in conducting studies on the hypothesis generation, information search, hypothesis
evaluation, and hypothesis selection phases of APs.

The inferential value of the research in informing practice and theory depends on the care taken by the researcher in designing the study and in readying the data for statistical analysis. Neither a solid theory nor an aggressive application of the latest statistical application can cure a flawed design. The researcher, interested in studying diagnostic judgments such as APs, is therefore well advised to be aware of some of the ‘tips and traps’ of design noted in this article. While we have pointed out many significant and pervasive design issues, it is not our intention (nor is it feasible) to provide an exhaustive list. Our objective is to help guide researchers to consider potential design considerations and trade-offs in conducting research on analytical procedures.

NOTES

1. Analytical procedures also require the auditor to develop an expectation that becomes the basis for determining whether there is a significant fluctuation. Given scope limitations, we do not address this phase in this paper but many of the design issues, especially the portion dealing with background information, discussed under hypothesis generation will be relevant in this phase as well. Cognitive theories of expectation formation are provided by Klein (1998) and Bédard and Chi (1993). Koonce (1993) provides a good literature review of each of these phases.

2. On a continuum, a depth strategy would be to employ multiple tests for only one hypothesis, while breadth entails conducting one test for many hypotheses. Balanced testing falls within these extremes. Asare et al. (2000) discuss the cost and benefits of each of these strategies.

3. Complementary revision occurs when the likelihood of one hypothesis is increased (decreased) with a corresponding, offsetting decrease (increase) in another hypothesis (hypotheses) such that the summed revisions are zero.

REFERENCES


**AUTHOR PROFILES:**

**Stephen Kwaku Asare** is an associate professor at the University of Florida. He teaches auditing and financial accounting and his research focuses on improving auditors’ judgements and decisions. He has a BSc. (First Class) from the University of Ghana and has published papers in various journals. He is currently the co-editor of *Journal of Accounting Literature*.

**Arnold Wright** is the Andersen Professor at Boston College. His research interests are in the judgement and decision-making area, primarily applied to the field of auditing and assurance. He has published over 50 articles and is currently the editor for *Auditing: A Journal of Practice & Theory* (7/99-7/02).