Analytical Review: Two Memory Aids Found to Improve Auditors' Effectiveness

Robert Marley[a]

[a]Georgia Southern University
School of Accountancy
P.O. Box 8141
Statesboro, GA 30460-8141
(912) 478-8008
rmarley@georgiasouthern.edu

Spring 2013

This is a manuscript working paper. Please do not cite or distribute without the author's permission.

I thank Mark Mellon, Uday Murthy, Stephanie Bryant, Rosann Collins, Terry Engle, and the University of South Florida workshop participants for their helpful and valuable feedback.

Abstract

Empirical evidence suggests that auditors have difficulty performing analytical review even when the relevant information is identified. In this study, I examine whether two memory aids, designed to help with the information retrieval process, improves auditors' analytical review effectiveness. In a between-subjects experiment, I find each aid to help auditors propose, identify, and select the cause of the unexpected fluctuation, improving analytical review effectiveness.

Keywords: Analytical review, analytical procedures, analytical review effectiveness, decision aids, memory aids

I. Introduction

Research suggests auditors perform analytical review ineffectively (Bedard and Biggs 1991; Anderson and Koonce 1995; Bedard, Biggs, and Maroney 1998; Asare and Wright 2001; Asare and Wright 2003; Green and Trotman 2003; Green 2004). Consistent with the definition used in audit practice, I define effective analytical review as when an auditor ultimately selects the "true" cause of the unexpected fluctuation. Analytical review is the process of identifying and resolving unexpected fluctuations (Koonce 1993). An unexpected fluctuation arises when there is a significant difference between a client's reported balance and the auditor's expectation for that balance. When performing financial statement audits, external auditors around the world employ analytical review because it is consistent with conducting an audit from a holistic, risk-based approach (Trompeter and Wright 2010) and because performing analytical review is presumptively mandatory under audit standards (AICPA 2010; IAASB 2010).[1]

Auditors use analytical review as an attention-directing device to identify risk and to obtain audit evidence. While auditors perform analytical review during the planning, fieldwork, and overall review stages of an audit, this study focuses on improving auditors' analytical review effectiveness when performed to evaluate financial statement assertions during fieldwork. That is, this study examines whether two memory aids help auditors propose and select the cause of the unexpected fluctuation.

There are negative consequences associated with ineffective analytical review. An auditor's failure to select the cause of an unexpected fluctuation negatively affects both audit effectiveness and audit efficiency. For example, failing to correctly attribute an unexpected increase in gross margin to an accounting error leads to misstated financial statements,

[1] The PCAOB adopted AU 329's promulgation as part of the body of interim auditing standards. Conducting analytical review during the audit planning stage and final review stage is presumptively mandatory.

decreasing audit effectiveness. Audit failure results if the auditor fails to identify misstatements that materially misrepresent a company's financial position. Conversely, conducting ineffective analytical review has audit efficiency consequences. For example, if an auditor fails to select the cause of an unexpected fluctuation, audit fieldwork can be inappropriately channeled based on the auditor's incorrect selection. Although the cause of the unexpected fluctuation might ultimately be uncovered before an audit opinion is rendered, wasted audit resources (such as time and effort) accompany the discovery because additional testing is required to obtain additional audit evidence.

Performing analytical review ineffectively poses a significant, and growing, risk to public accounting firms: analytical review constitutes an increasingly large part of audit engagement budgets, representing approximately 25 percent of engagements' budgeted hours compared to 21 percent of budgeted hours at the turn of the millennium and 15 percent of budgeted hours during the early 1990's (Trompeter and Wright 2010; Ameen and Strawser 1994). Traditionally, analytical review was performed by senior auditors. However, a recent study by Trompeter and Wright (2010) reveals novice auditors to conduct analytical review 48 percent of the time as compared to ten percent of the time in the early 1990's (Hirst and Koonce 1996).[2]

Through a set of empirical findings I term "the analytical review paradox," research suggests auditors usually identify the relevant pieces of information needed to propose the cause of an unexpected fluctuation when performing analytical review (Bedard and Biggs 1991; Bedard, Biggs, and Maroney 1998). Nonetheless, auditors frequently fail to effectively utilize the identified, relevant pieces of information to propose the unexpected fluctuation's cause. Further,

[2] Trompeter and Wright (2010) note that although novice auditors now frequently *conduct* analytical review, experienced auditors (i.e. audit seniors and above) still generally *design* the analytical review tests. Consequently, novice auditors are often the ones who propose, evaluate, and select the cause of the unexpected fluctuation. As a result, these three stages are investigated in this study.

when auditors do manage to propose the cause, they often fail to select it (Bedard, Biggs, and Maroney 1998; Green and Trotman 2003; Asare and Wright 2003; Green 2004). I suggest one reason the paradox exists is that auditors fail to accurately recall the numerous relevant cues needed to propose and select the cause of the unexpected fluctuation. Using research in human cognition as the basis for designing two memory aids created to help auditors accurately recall relevant cues from memory, I investigate whether my aids improve the analytical review effectiveness of novice auditors.[3]

The first memory aid, the activity relationship diagram (ARD), presents the auditor with a graphical depiction of basic accounting relationships. As auditors should already be familiar with the basic accounting relationships depicted by the ARD, and hence reside in long-term memory, the aid's purpose is to assist the auditor with accurately recalling these relationships from long-term memory during analytical review.

At key stages of analytical review, the second memory aid, the pattern-consideration aid (PCA), presents each auditor with the cues she self-identified as relevant. As the relevant cues must be self-identified during analytical review, these cues will reside within short term memory. By improving an auditor's cue recall accuracy, the PCA should help auditors perform analytical review more effectively.

Figure 1 presents a complete list of analytical review's stages and identifies the key stages investigated in this study.

<<INSERT FIGURE 1 HERE>>

[3] Since this study is concerned with investigating the analytical review effectiveness of novice auditors tasked with performing analytical review, hereafter the term "auditor" in this study refers to novice auditors and not those with relatively high levels of audit experience such as senior and manager level auditors.

In a between-subjects experiment, I manipulate memory aids at three levels: unaided, ARD, and PCA. In an analytical review task, 122 masters and accounting senior students proxy as novice auditors who propose, evaluate, and select the cause of an unexpected fluctuation.

As predicted, I find each memory aid to improve auditors' analytical review effectiveness when compared to auditors who perform analytical review unaided. That is, I find both memory aids to improve the odds that cause of the unexpected fluctuation is proposed, identified, and ultimately selected. Further, I find empirical evidence indicating that both aids improve analytical review effectiveness by assisting auditors with information recall.

Responding to Bonner's (1999) call for research to mitigate known person-judgment and decision making deficiencies, this study introduces and empirically validates two new decision aids that can be used to attenuate the analytical review paradox. Further, this study contributes to practice by providing practitioners with two relatively easy-to-deploy decision aids found to improve auditors' analytical review effectiveness.

The remainder of this study is organized as follows. In the Background and Hypotheses section, I present a literature review and develop my predictions. In the Method section, I discuss the experimental design, the task, and my dependent variable. In the Results section, I discuss this study's findings. Finally, the Discussion and Conclusions section contains a discussion of this study's limitations, contributions, and future research opportunities.

II. Background and Hypotheses

An unexpected fluctuation arises during analytical review when there is a significant difference between a client's reported balance and the auditor's expectation for that balance. Although there are numerous analytical review techniques, such as ratio and/or trend analysis, the rationale for using techniques is to provide assurance that financial statement assertions are

not materially misstated. The audit literature recognizes analytical review as a useful tool for detecting misstatements, whether caused by error or fraud (Loebbecke and Steinbart 1987; Kinney 1987; Wells 2007).

Once an unexpected fluctuation has been identified, the audit literature categorizes analytical review into three distinct and sometimes iterative stages: hypothesis proposal, hypothesis identification, and hypothesis selection (Koonce 1993; Asare and Wright 1997; Asare and Wright 2001). During the hypothesis proposal stage an auditor generates potential causes of the unexpected fluctuation using domain and context specific knowledge (Koonce 1993; Asare and Wright 2001).[4] To effectively propose hypotheses an auditor needs to organize the relevant information into a pattern and apply her accounting knowledge to make valid inferences. Although auditors are generally able to identify the cues needed to propose the cause of the unexpected fluctuation, auditors have trouble considering the cues as a pattern, often failing to propose a hypothesis that addresses all the cues (Bedard and Biggs 1991; Bedard, Biggs, and Maroney 1998). Thus, auditors frequently fail to propose the cause of the unexpected fluctuation.

After hypothesis proposal, the auditor moves to the hypothesis identification stage where she evaluates each of her proposed hypotheses against the cues to determine if any hypothesis satisfactorily explains the unexpected fluctuation's cause (Koonce 1993; Hirst and Koonce 1996). Auditors have difficulty evaluating multiple hypotheses during analytical review, causing them to perform this stage poorly (Bedard and Biggs 1991; Asare and Wright 2003; Green and

[4]Alternatively, an auditor may obtain an explanation directly from the client instead self-generating hypotheses. However, obtaining an explanation from the client can lead to inappropriate fixation, clustered substantive testing, and fewer self-generated hypotheses (Bedard and Biggs 1991; Anderson et al 1992; Church and Schneider 1993; Asare and Wright 1997; Bierstaker et al 1999; Asare and Wright 2003; Green 2004). Recognizing this, I examine auditor hypothesis self-generation in this study. Consequently, future references to "hypothesis proposal" imply self-generation.

Trotman 2003; Green 2004). Thus, auditors often fail to identify the unexpected fluctuation's cause during the hypothesis identification stage.

Finally, during the hypothesis selection stage the auditor chooses the hypothesis she has identified as the unexpected fluctuation's cause. Although hypothesis selection is the last stage, it is also the most critical because analytical review is not effective unless hypothesis selected by the auditor is the actual cause of the unexpected fluctuation. Research consistently finds that even experienced auditors frequently fail to select the unexpected fluctuation's cause during this stage (Bedard, Biggs, and Maroney 1998; Green and Trotman 2003; Asare and Wright 2003; Green 2004). I created two memory aids to help auditors overcome these difficulties identified. Both decision aids are developed based upon research in human cognition as explained in the next section.

Human Cognition and Information Retrieval

Information retrieval refers to the process where an individual recalls information from memory. Information can be retrieved from either short term memory or long term memory. Short term memory can be thought of as the amount of space available to an individual for retaining, processing, and manipulating information (Baddeley 1992). Although researchers agree that individuals' short term memory is not unlimited, the size of its limits is widely debated. Sweller suggests that short term memory can contain approximately seven pieces of information at any one time while Miller (1956) suggests it is even more constrained. Regardless of short term memory's exact size, there is consensus that it is substantially limited (Simon 1974; Penney 1989; Baddeley 1992). Long-term memory represents a more permanent state in which information can exist in a dormant, unused state until retrieved by the individual into consciousness.

Failure to accurately retrieve information is the source of "forgetting" information "known" by the individual (Anderson et al. 1994). Within analytical review, failure to retrieve information accurately can result in an auditor's failure to recall the relevant information needed to propose, identify, and select the unexpected fluctuation's cause.

Decision Aids and Memory Aids

Decision aids are tools designed to improve decision quality by addressing specific aspects of the judgment and decision making process. A considerable body of research supports the notion that decision aids can overcome cognitive constraints and improve judgment and decision-making performance (Ashton 1992; Rose 2002; Bonner 2008). Memory aids are a type of decision aid whose purpose is to assist individuals in retrieving relevant information from memory in a judgment and decision making context (Bonner 2008). Because auditors often fail to retrieve all the relevant information while performing a task, memory aids can help with the cognitive process of retrieving knowledge from memory (Frederick 1991; Bonner 2008). Accordingly, I create and empirically examine two memory aids designed to help auditors recall information relevant to the judgment and decision making process, thereby improving the quality of the information consider during analytical review.

Although auditors frequently use memory aids, such as audit manuals and audit programs, to help with cue recall during various phases of an audit engagement, public accounting firms rarely use memory aids during analytical review (Hirst and Koonce 1996; Bonner 1990; Trompeter and Wright 2010). However, given public accounting firms extensive use decision aids to improve audit effectiveness (Dowling and Leech 2007), and since research suggests memory aids memory aids greatly enhance information processing in auditing contexts, I suggest accounting firms may be receptive to using the memory aids examined in this study if

they are found to significantly enhance analytical review effectiveness.[5] Further, the use of decision aids that foster the use of auditors' professional judgment, such as memory aids, instead of supplanting auditors' professional judgment find greater acceptance in practice (Kneschel 2000; Knechel 2007).

Activity Relationship Diagram Aid

Within the context of analytical review, auditors need to accurately recall basic accounting relationships since the cause of the unexpected fluctuation is deduced by analyzing unanticipated fluctuations among one or more specific accounts. As shown in Figure 2, the activity relationship diagram (ARD) aid is a flowchart-type diagram that gives an auditor a graphical depiction of basic accounting relationships. The ARD pictorially displays the duality of account relationships by demonstrating how an increase (decrease) in one account results in a concomitant decrease (increase) in related accounts. By graphically depicting basic accounting relationships, the ARD eliminates the need for users to wholly rely upon their long-term memory to retrieve domain-specific information (i.e., accounting knowledge). Even though auditors should know the basic relationships depicted within the ARD as part of their accounting education, an auditor's ability to accurately recall these relationships can vary due to a variety of factors such as cognitive interference, memory decay, or lack of retrieval cues. Further complicating information recall is that little of the knowledge an individual has in long-term knowledge is consistently available at any given point in time (Anderson et al 1994).

[5] In semi-structured interviews with audit partners and audit managers from the Big Four and one other international public accounting firm, Dowling and Leech (2007) find that only two use decision aids related to analytical review: one firm uses a decision aid that helped with ratio calculation while the other firm uses "tools to extract and analyze data" (p. 99). The decision aids examined in this study differ from the decision aid used to calculate ratios because this study's aids do not focus on ratio calculation, rather my aids heighten the salience of information needed to effectively perform analytical review. Thus, only one of the accounting firms examined by Dowling and Leech may already use a decision aid comparable to the two examined in this study.

Consequently, I suggest the ARD serves as a memory aid by helping auditors to recognize basic accounting relationships during analytical review. The application of a diagram to a context where the auditor may fail to accurately recall information relevant to the task should improve the accuracy of the information considered. Hence, by improving the accuracy of the information retrieval process, the ARD should improve the quality of the information considered, leading to more effective analytical review.

Within the hypothesis proposal stage, the ARD should enable users to more accurately recall basic accounting relationships. By improving the quality of the information available during hypothesis proposal, the ARD should help auditors to propose the cause of the unexpected fluctuation more often than non-ARD users. Within the hypothesis identification stage, the accurate retrieval of account relationships should permit ARD users to more easily determine which proposed hypothesis is most consistent with the unexpected fluctuation, leading to the identification of the unexpected fluctuation's cause more often than when auditors conduct hypothesis identification unaided. Finally, because auditors performance during the hypothesis selection stage is a direct result of their effectiveness during the preceding two stages, I predict ARD users will select the unexpected fluctuation's cause more frequently than unaided auditors. However, it remains an empirical question as to whether the ARD will improve auditors' analytical review effectiveness because the aid merely provides auditors with basic accounting relationships that should be already known. Thus, analytical review effectiveness may not improve because the aid does not provide auditors with any information not previously known. Accordingly, I predict:

H1: Auditors using the activity relationship diagram will perform analytical review more effectively than auditors who perform analytical review without the use of a memory aid.

<<INSERT FIGURE 2 HERE>>

Pattern-Consideration Aid

To perform analytical review effectively, auditors need to accurately recall the cues they believe to be relevant to the unexpected fluctuation's cause because the cues form the basis of information the auditor uses to propose, identify, and select the correct cause of the unexpected fluctuation. In a result pattern I term "the analytical review paradox," audit research suggests that though auditors generally identify the cues relevant to an unexpected fluctuation (Bedard and Biggs 1991; Bedard, Biggs, and Maroney 1998), they often fail to propose the unexpected fluctuation's cause. Further, even when the unexpected fluctuation's cause is proposed, auditors often fail to identify and select the cause during analytical review tasks (Bedard, Biggs, and Maroney 1998; Green and Trotman 2003; Asare and Wright 2003; Green 2004).

I suggest one explanation for the analytical review paradox is that auditors while auditors may identify all the relevant cues, they are often unable to accurately recall all the cues on demand. Given that cues relevant to explaining the unexpected fluctuation's cause are specific to the circumstances of each analytical review task, the cues must be identified and maintained by short term memory. That is, the auditor's lack of familiarity with the relevant cues does not lend the cues to be encode into long-term memory. Due to limitations in short-term memory retrieval, I suggest auditors frequently perform hypotheses proposal, identification, and selection without the full cue set, causing them to perform analytical review ineffectively.

Thus, the PCA should serve as a memory aid by accurately retrieving the cues self-identified by an auditor as relevant, enabling the auditor to perform the hypotheses proposal, identification, and selection stages of analytical review with a full cue set, leading to more

effective analytical review.[6] More specifically, by accurately retrieving the auditor-identified cues during hypotheses proposal, the PCA enables the auditor to consider the cues holistically while proposing potential causes of the unexpected fluctuation. During the hypotheses identification stage, the PCA helps with the auditor's decision making process by requiring auditors to view each of their proposed hypotheses and "check the box" next the relevant cues supporting each proposed hypothesis. After the auditor has evaluated each hypothesis in terms of the number of relevant cues that support it, the PCA displays the number of cues supporting each hypothesis. PCA users can then click on each hypothesis to see the specific cues supporting it. Non-PCA users must employ their working memory to accurately recall the cues, in determining how many cues support each proposed hypothesis, and to keep track of the cues supporting each hypothesis. Thus, PCA users should perform the identification stage more effectively than those who must rely upon their working memory. Finally, during the hypothesis selection stage, the PCA's presentation of the user's proposed hypotheses and the number of cues supporting each should make it easier for PCA users to select the hypothesis most consistent with the auditor-identified relevant cues. Accordingly, I predict that auditors using the PCA will select the true cause of the unexpected fluctuation more frequently than auditors who conduct analytical review unaided:

H2: Auditors using the pattern-consideration aid will perform analytical review more effectively than auditors who perform analytical review without the use of a memory aid.

[6] A key aspect of the PCA is that it does not provide a user with any information the individual did not self-enter into the aid. That is, the PCA does not provide a user with any information the user did not type into it.

III. Method

To test the hypotheses, I use a 1 x 3 between-subjects design. The decision aid factor is manipulated at three levels: unaided, ARD, and PCA. The research design is illustrated in Figure 3.

<<INSERT FIGURE 3 HERE>>

Procedure

The experiment consists of six parts: an online pre-test, analytical review training, case familiarization, hypothesis proposal, hypothesis identification and hypothesis selection, and an online post-experiment questionnaire. All experimental sessions took place within a university lab setting. First, participants completed an online pre-test designed to gauge warm their minds up for the training session to follow and to guage their understanding of basic accounting relationships. Next, participants were given a 20 minute training session to provide them with a basic understanding of analytical review. To avoid introducing bias into the training sessions, the experimental treatment condition was randomly determined by drawing a slip of paper from a hat at the end of each training session. Figure 4 illustrates the study's experimental procedure.

<<INSERT FIGURE 4 HERE>>

Task Stage One: Case Familiarization and Cue Identification

After training, participants were told to assume the role of novice auditors asked to identify the cause of an unexpected fluctuation.[7] Participants were asked to spend 15 minutes becoming familiar with Bean Co., a fictitious company buying and selling one product: commoditized Jamaican coffee beans. Participants were given two pages of client background

[7] To permit a meaningful interpretation of the results the case was designed to have only one correct answer. I abstract from practice for experimental tractability, particularly with regards to interpreting the effect each decision aid has on task effectiveness. While designing a task to have more than one correct answer may improve external validity, doing so introduces more variation into the setting, complicating the interpretation of results.

information and one page of financial information. The case materials informed participants that their engagement senior had discovered that the client's cost of sales ratio was lower than expected. The instructions asked all participants to identify and record the cues they felt were related (i.e., relevant) to the cause of the unexpected fluctuation.

The financial information was adapted from Kinney (1987) and developed by first establishing the correct account and ratio balances, then seeding the appropriate discrepancies. Participants were provided with fourteen cues, although six were relevant to deducing the true cause of the unexpected fluctuation: (1) the inventory turnover ratio was lower than expected, (2) the accruals ratio was higher than expected, (3) the gross margin ratio was higher than expected, (4) purchases are lower than expected, (5) accounts payable was less than expected, and (6) inventory costs did not decrease. When the six relevant cues are considered as one pattern, the resulting conclusion is that the client failed to record an inventory purchase that was subsequently recorded as a sale.

The materials focused participant attention on the unaudited information provided by the client, not on the accuracy of the information provided by the audit firm by presenting account balances and ratios as "Audited," "Expected," and "Unaudited." Audited information was described as the result of last year's audit and, therefore, could be considered completely accurate. Expected information was described as the account expectations developed by the engagement partner based upon her past client experience and current industry trends. To prevent participants from questioning the validity of the partner's expectations, participants were told that the partner's expectations were completely accurate. Unaudited information was described as current year information provided by the client without any verification regarding its accuracy.

Task Stage Two: Hypothesis Proposal

After the Case Familiarization Stage, participants were asked to spend at least 15 minutes proposing potential causes (i.e., hypotheses) of the unexpected fluctuation. Participants in all conditions recorded their proposed hypotheses using a text box in the task's application software. The researcher orally informed all participants that it might be helpful to consider the cues they identified as relevant in formulating their proposed hypotheses.

Before stage two commenced, participants in the ARD intervention were handed a placemat-sized copy of the ARD and orally informed that it might be helpful to refer to the ARD when formulating proposed hypotheses. The application software automatically presented participants in the PCA intervention with the cues they previously self-identified as relevant whenever participants proposed a new hypothesis. Figure 5 illustrates the PCA's cue presentation format.

<<INSERT FIGURE 5 HERE>>

Task Stage Three: Hypothesis Identification and Hypothesis Selection

As each participants individually indicated they were finished proposing hypotheses, each participant was handed paper instructions asking them to evaluate each of their proposed hypotheses to identify the one that best explained the cause of the unexpected fluctuation. During this stage, the PCA required its' users to evaluate each proposed hypothesis against the cues they previously self-identified as relevant before choosing the best explanation. Figure 6 provides a screenshot that illustrates the PCA's check-box functionality. Consistent with the other stages in this experiment, participants were not required proceed to or finish stage 3 until

ready.[8] Then, each participant completed an online post-experiment questionnaire before being dismissed.

<<INSERT FIGURE 6 HERE>>

Dependent Measure

I use one dependent variable to test the hypotheses: the proportion of times each intervention's participants selected the cause of the unexpected fluctuation.

IV. Results

Sixteen experimental sessions, lasting an average of 37 minutes, were conducted with one hundred and thirty nine accounting masters and second-semester accounting seniors from a large state university serving as proxies for novice auditors. I suggest these participants are an appropriate proxy for novice auditors since accounting masters and second-semester seniors constitute the primary candidate pools from which public accounting firms recruit novice auditors.

Participant Exclusion Criteria

Two questions in the post experiment questionnaire served as manipulation checks. The purpose of the questions was to identify whether participants were attentive to the task and whether they performed the experiment conscientiously.[9] Nine participants were eliminated from the analysis because they failed the manipulation check questions.

[8] Although imposing an overall time limit on participants may seem desirable, I suggest such a limit actually constitutes a significant departure from practice and potentially leads to false results. Aside from budgetary considerations , there is no time limit imposed upon auditors to propose, evaluate, and select the cause of an unexpected fluctuation in practitioner settings. Further, within a laboratory setting, imposing a time limit could pressure (at least some) participants to choose an explanation they would not otherwise select. Accordingly, examining the effects of time pressure on the effectiveness of the ARD and PCA is an extension I leave to future research.

[9] The first manipulation check question asked participants to "Please select the unexpected fluctuation you investigated today." The second manipulation check question asked participants to "Please indicate what materials were available to you while conducting analytical review today."

Participants also took an accounting relationship knowledge test before the main task to assess their knowledge of the basic accounting relationships depicted within the ARD.[10] To appropriately attribute any difference in task effectiveness to the memory aid treatments, I only analyzed data from participants who reported placing some degree of reliance on the memory aid available to them. As a result, seven participants were excluded from the analysis because they reported no reliance upon the memory aid given to them. Finally, I excluded one participant from the analysis because her electronic files became corrupted and I was unable to open them. After removing excluded observations, data from 122 participants was used to test this study's hypotheses.

Participant Descriptive Statistics

Aside from the amount of time spent on the task, the demographics of participants across the three interventions were not significantly different, providing evidence that participants were randomly assigned to the three treatment conditions. Using a 7-point Likert scale, participants generally found the case material easy to understand ($\bar{x}$ = 2.3) and found it somewhat easy to identify the pieces of information related to the unexpected fluctuation ($\bar{x}$ = 3.1). Participants indicated they did not feel the task was hurried ($\bar{x}$ = 2.2) and reported exerting a relatively high amount of mental effort ($\bar{x}$ = 5.0). Finally, participant responses indicated that it was of high importance for them to find the correct answer to the task ($\bar{x}$ = 5.4). Thus, participant responses

[10] Originally the author intended to exclude participants who failed to answer all questions on the knowledge test correctly because answering an accounting relationship question incorrectly introduces the possibility the ARD provides users with knowledge not already known to the participant. However, since this study suggests both decision aids function as memory aids which help participants to recall knowledge they know (but may be unable to recall at a specific point in time) a maintained assumption is that the accounting relationships presented in the ARD are so basic that every accounting senior or masters student knows them. A participant may have answered incorrectly due to a variety of factors: a participant may have not known the correct answer (i.e., incorrectly believed that general and administrative expenses are related to cost of sales), a participant may have known the correct answer but was unable to recall it, or a participant may not have attended to the question. Regardless, the results of this study are not only robust, but even stronger when excluding participants that did not answer every question on the knowledge test correctly.

suggest they did not find the case confusing or rushed and suggest participants were diligent in performing the task. Participants spent an average of 37 minutes performing the task. There was a significant difference in time spent on the task between the no-aid intervention and the PCA intervention. The difference is not unexpected since PCA users must evaluate each of their proposed hypotheses using the PCA's checkbox functionality. Finding that participants within the PCA intervention took approximately three minutes longer than unaided participants (40 minutes on average vs. 37 minutes on average, respectively) to complete the task provides evidence that PCA users attended to the aid. Further, task time was not found to be significant covariate in testing this study's hypotheses.

Data Coding

The researcher and one doctoral student served as data coders. To prevent coding bias, neither coder was provided with information that could be used to identify the data's intervention treatment. Thus, even though coders were not blind to the study's hypotheses, they were unable to identify the intervention from which any given piece of datum originated. After both coders classified the data, they met to disclose the classification that each coder assigned to every piece of datum. Coder responses were then compared and any discrepancies were reconciled. Prior to resolving discrepancies, the Cohen's Kappa coefficient for the dependent variable was 0.88, suggesting a high degree of coding agreement.

Assumption Testing: Logistic Regression (LOGIT)

I use LOGIT to test my hypotheses because the dependent variable is discrete. The assumption of observation statistical independence is satisfied by using a between-subjects experiment design.

Hypotheses Testing

Hypothesis 1 (H1) predicts that auditors using the activity relationship diagram will perform analytical review more effectively vis-à-vis auditors who do not use a memory aid.

Analytical review is performed successfully when the cause of the unexpected fluctuation is selected by the auditor.[11] Fourteen out of 45 ARD users (31.1 percent) selected the cause of the unexpected fluctuation as compared to four out of 40 unaided participants (10.0 percent). Using a LOGIT model that controls for cognitive load, the results reported in Table 1 indicate the odds a participant selects the unexpected fluctuation's cause increase when the ARD is used (χ^2=6.295, df = 1, p=0.011, one-tailed).[12] Thus, H1 is supported.

<<INSERT TABLE 1 HERE>>

Hypothesis 2 (H2) predicts that auditors using the pattern-consideration aid will perform analytical review more effectively vis-à-vis auditors who do not use a memory aid. Ten out of 37 PCA users (27.0 percent) selected the cause of the unexpected fluctuation as compared to four out of 40 unaided participants (10.0 percent). Using a LOGIT model that controls for cognitive load, the results reported in Table 2 indicate the odds a participant selects the true cause increase when the PCA is used (χ^2=3.456, df = 1, p=0.032, one-tailed). Thus, H2 is supported.

<<INSERT TABLE 2 HERE>>

Additional Analysis: How Does Each Memory Aid Improve Analytical Review Effectiveness?

In developing this study's hypotheses, I suggested that by helping participants to accurately retrieve the information they self-identified as relevant, the memory aids would improve participants' effectiveness during three key stages of analytical review: hypotheses proposal, hypothesis identification, and hypothesis selection. Having demonstrated both aids

[11] On other words, when the actual cause of the unexpected fluctuation is chosen by the auditor. Naturally, an auditor engaged in analytical review will, by definition, select a cause she believes explains the unexpected fluctuation. However, whether the cause selected by the auditor is the actual cause of the unexpected fluctuation is the distinguishing characteristic of whether analytical review is performed effectively or not.

[12] Results are robust to statistical method as support is found for both hypotheses using Fisher's exact test (one-tailed p-values 0.015 and 0.050 for the ARD and PCA, respectively). Fisher's exact test has been demonstrated to be more conservative than logistic regression (Liddell 1976; Berkson 1978; D'Agostino et al. 1988).

improve hypothesis selection, I now I perform additional analysis to determine whether these memory aids improve the odds hypotheses proposal and hypothesis identification are performed effectively.

To test whether the memory aids improve participants' hypotheses proposal effectiveness vis-à-vis participants who perform the task unaided, I use a LOGIT model to regress the use of each aid against whether the unexpected fluctuation's cause is proposed. Sixteen out of 45 ARD users (35.5 percent) and 13 out of 37 PCA users (35.1 percent) propose the cause of the unexpected fluctuation, as compared to nine of 40 unaided participants (22.5 percent). Controlling for cognitive load, I find the odds a participant proposes the true cause increase when the ARD is used (χ^2=2.771, df = 1, p=0.048, one-tailed) and when the PCA is used (χ^2=2.404, df = 1, p=0.061, one-tailed). Thus, there is empirical evidence to suggest the memory aids improve participants' hypothesis proposal effectiveness.

To test whether the memory aids improve participants' hypothesis identification effectiveness as compared to participants who conduct the task unaided, I use a LOGIT model to regress the use of each aid against whether the cause is identified.[13] Fourteen out of 16 ARD users (87.5 percent) and ten out of 13 PCA users (76.9 percent) identified the cause of the unexpected fluctuation when proposed, as compared to four out of nine (44.4 percent) unaided participants. Controlling for cognitive load, I find using the ARD increases the odds a participant identifies the unexpected fluctuation's cause (χ^2=4.170, df = 1, p=0.021, one-tailed) and using the PCA increases the odds a participant identifies the unexpected fluctuation's cause

[13] I operationalize a participant's identification of the cause by whether the participant ultimately selects it when proposed. I suggest this is an appropriate way to operationalize hypothesis identification because the participant ultimately selects the proposed hypothesis she has evaluated to be the best explanation of the cause of the unexpected fluctuation. If the memory aids help participants to evaluate their proposed hypotheses, I would expect when the cause of the unexpected fluctuation is proposed, it will be selected more often by aid users than by unaided participants.

(χ^2=3.456, df = 1, p=0.032, one-tailed).[14] Thus, there is empirical evidence to suggest that both the ARD and PCA improve overall analytical review effectiveness by helping participants perform the hypotheses proposal stage and hypothesis identification stage.

VI. Discussion and Conclusions

This study examines whether use of two memory aids, the Activity Relationship Diagram (ARD) and the Pattern-Consideration Aid (PCA), improve novice auditors' analytical review effectiveness across three stages: hypotheses proposal, hypothesis identification, and hypothesis selection. In a laboratory setting, novice auditors performed analytical review to identify the cause of an unexpected fluctuation. The presence or absence of a memory aid (ARD, PCA, or unaided) was varied between subjects.

Compared to novice auditors who perform the task without a memory aid, I find both the ARD and the PCA to increase the odds that novice auditors perform analytical review effectively. That is, each memory aid increases the odds a novice auditor will ultimately select the cause of the unexpected fluctuation.

Additional analysis reveals that both the ARD and the PCA improve analytical review effectiveness because each aid increases the odds a novice auditor proposes (i.e., self-generates) the cause of the unexpected fluctuation during the hypotheses proposal stage and because each aid increases the odds an auditor identifies the cause of the unexpected fluctuation (when it is proposed) during the hypothesis identification stage.

Further, conducting post-hoc analysis, I find empirical evidence to suggest that cognitive load is negatively associated with analytical review effectiveness, extending the cognitive load literature outside the traditional learning context, to a judgment and decision making task.

[14] Results are robust to statistical method as support is found for both hypotheses using Fisher's exact test (two-tailed p-values 0.003 and 0.047 for the ARD and PCA, respectively).

My results have a number of implications for theory and practice. By creating and empirically validating two memory aids found to mitigate the analytical review paradox, this study answers Bonner's (1999) call for research addressing known person-specific accounting judgment and decision making deficiencies. Further, this study extends Bonner's (2008) theoretical work on memory aids to a new judgment and decision making context: analytical review, finding memory aids to improve decision quality. Finally, finding cognitive load to negatively affect decision quality in a non-learning, judgment and decision making context extends the literature on cognitive load to a new setting.

From a practitioner perspective, because I find novice auditor using either memory aid to significantly outperform those who conduct analytical review unaided, it seems logical to suggest that auditors should use some form of decision aid during analytical review.[15] Thus, a major contribution of this study is in providing audit practitioners with two relatively easy-to-deploy memory aids than can be used to improve auditors' analytical review effectiveness. Further, public accounting firms may also find it beneficial to minimize the amount of cognitive load placed upon auditors as they perform analytical review.

Although one of the strengths of this study is that we use a validated analytical review task with one seeded cause of the unexpected fluctuation (Kinney 1987), I acknowledge that an unexpected fluctuation's cause can stem from more than one underlying cause. Thus, this study's results may be less generalizable as the number of causes associated with an unexpected fluctuation increases beyond one. Further, because I employ masters and senior accounting

[15] Time is always an important practitioner consideration since auditors have to accomplish their work under tight time budgets. Thus, although it took auditors using the PCA approximately three minutes longer to conduct analytical review vis-a-vis auditors who conducted the task unaided (or approximately eight percent longer) I suggest that in light of the effectiveness of the PCA, this relatively small time differential is not large enough to dissuade public accounting firms from deploying the PCA on client engagements.

students as surrogates for novice auditors, the results of this study may not generalize to settings where experienced auditors conduct analytical review. Thus, examining whether the decision aids investigated within this study can improve the analytical review effectiveness of experienced auditors is an extension I leave to future research. Finally, this study only examines an unexpected fluctuation caused by client error. Although I have no basis to predict my findings will not generalize to settings where the unexpected fluctuation is caused by factors besides client error, such as fraud, whether the effectiveness of these memory aids will hold in a context outside of client error remains an empirical question.

Future Research Opportunities

Extending the ARD and PCA to an analytical review experiment employing experienced auditors is a natural and logical extension of the current study. This study provides evidence that both the ARD and PCA improve the analytical review effectiveness of novice auditors, who conduct analytical review approximately 48 percent of the time (Trompeter and Wright 2010). However, it is less clear whether these findings can be generalized to the remaining 52 percent of settings where experienced auditors conduct analytical review. Extending the ARD and PCA to a setting comprised of experienced auditors is important not only to determine whether the results found within this study hold among experienced auditors, but also because experienced auditors may possess characteristics that differentially impact each aid's effectiveness. For example, since experienced auditors usually identify the cues needed to propose the true cause of the unexpected fluctuation (Bedard and Biggs 1991; Bedard, Biggs, and Maroney 1998), the PCA may be especially useful to experienced auditors during hypothesis proposal because it automatically displays the auditor-identified relevant cues, making the cues more available to the auditor while she attempts to propose the true cause. Future research could also extend the use

of novice auditors to two-person groups, as prior research has only examined the analytical review effectiveness of experienced auditors within two-person groups (Bedard, Biggs, and Maroney 1998). Since inexperienced auditors might not perform analytical review as effectively as experienced auditors and considering the fact that novice auditors are increasingly performing analytical review, the use of two-person groups may be an avenue available to public accounting firms to improve the analytical review effectiveness of novice auditors.

There are a few reasons to expect that investigating the analytical review performance of novice auditors may yield different results that those previously found in the experienced auditor group literature. First, a novice auditor has less overall accounting knowledge than an experienced auditor. Thus, putting two novice auditors together may increase the pool of available knowledge more than a group comprised of two experienced auditors, whose knowledge pool may not increase as much because each experienced auditor has a larger knowledge base. Second, due to their inexperience, novice auditors may be more willing to consider each other's input during analytical review as compared to a group of experienced auditors who may place more confidence in their own judgment and be less willing to consider alternative points of view. Finally, I suggest that novice auditors may be more open to conducting analytical review in groups as compared to experienced auditors, who may view working with another auditor as unnecessary and a nuisance.

To conclude, the main findings of this study are that both memory aids improve the odds that novice auditors perform analytical review effectively. Both the ARD and the PCA improve the odds that a novice auditor proposes the cause of an unexpected fluctuation, identifies the cause when it is proposed, and ultimately select the cause of an unexpected fluctuation.

References

AICPA. 2010. American Institute of Certified Public Accountants. AU 329: Analytical Procedures. *AICPA Professional Standards.* AICPA: New York, NY: 339-344.

Ameen, E., and J. Strawser. 1994. Investigating the use of analytical procedures: An update and extension. *Auditing: A Journal of Practice & Theory.* 13(2): 69-76.

Anderson, J. C., and S. Kaplan, and P. Reckers. 1992. The effects of output interference on analytical procedures judgments. *Auditing: A Journal of Practice & Theory*. 11: 1-13.

Anderson, U., and L. Koonce. 1995. Explanation as a method for evaluating client-suggested causes in analytical procedures. *Auditing: A Journal of Practice & Theory*. 14: 124-132.

Asare, S. K., and A. M. Wright. 1997. Evaluation of competing hypotheses in auditing. *Auditing: A Journal of Practice & Theory.* 16(1): 1-13.

Asare, S. K., and A. M. Wright. 2001. Design considerations for research on analytical procedures. *International Journal of Accounting*. 5: 205-214.

Asare, S. K., and A. M. Wright. 2003. A note on the interdependence between hypothesis generation and information search in conducting analytical procedures. *Contemporary Accounting Research*. (20)2: 235-251.

Ashton, R. H. 1992. Effects of justification and a mechanical aid on judgment performance. *Organizational Behavior and Human Decision Processes*. 52: 292-306.

Baddeley, A. D. 1992. Working memory: The interface between memory and cognition. *Journal of Cognitive Neurosciences.* 4: 281-288.

Bedard, J.C., and S. F. Biggs. 1991. Pattern recognition, hypothesis generation, and auditor performance in an analytical task. *The Accounting Review.* 66(3): 622-642.

Bedard, J. C., and S. F. Biggs, and J. J. Maroney. 1998. Sources of process gain and loss from group interaction in performance of analytical procedures. *Behavioral Research in Accounting*. 10(Supplement): 207-233.

Bierstaker, J. L., and J. C. Bedard, and S. F. Biggs. 1999. The role of problem representation shifts in auditor decision processes in analytical procedures. *Auditing: A Journal of Practice & Theory.* 18: 18-36.

Bonner, S. 1999. Judgment and decision-making research in accounting. *Accounting Horizons.* 13(4): 385-398.

Bonner, S. 2008. Judgment and Decision Making in Accounting. New Jersey: Prentice Hall.

Church, B. K., and A. Schneider. 1993. Auditors' generation of diagnostic hypotheses in response to a superiors' suggestion: Interference effects. *Contemporary Accounting Research.* 10: 333-350.

Coglitore, F., and R. G. Berryman. 1988. Analytical procedures: A defensive necessity. *Auditing: A Journal of Practice & Theory*. 7(2): 150-163.

Dowling, C., and S. Leech. 2007. Audit support systems and decision aids: Current practice and opportunities for future research. *International Journal of Accounting Information Systems*. 8: 92-116.

Green, W. 2004. Impact of the timing of receipt of inherited explanation on auditors' analytical procedures judgments. *Accounting and Finance*. 44: 369-392.

Green, W. J., and K. T. Trotman. 2003. An examination of different performance outcomes in an analytical procedures task. *Auditing: A Journal of Practice & Theory.* 22(2): 219-235.

Hegarty, M., and K. Steinhoff. 1997. Individual differences in use of diagrams as external memory in mechanical reasoning. *Learning and Individual Differences*. 9(1): 19-42.

Hirst, D. E., and L. Koonce. 1996. Audit analytical procedures: A field investigation. *Contemporary Accounting Research*. (13)2: 457-486.

IAASB. 2010. International Auditing and Assurance Standards Board. International Standard on Auditing 520: Analytical Procedures. *Handbook of International Quality Control, Auditing, Review, Other Assurance, and Related Services Pronouncements*. 2010. IFAC: New York, NY: 433-440.

Kinney, W. E. 1987. Attention-directing analytical review using accounting ratios: A case study. *Auditing: A Journal of Practice & Theory.* 6(2): 59-73.

Koonce, L. 1993. A cognitive characterization of audit analytical review. *Auditing: A Journal of Practice & Theory*. 12(Supplement): 57-76.

Loebbecke, J. K., and P. J. Steinbart. 1987. An investigation of the use of preliminary analytical review to provide substantive audit evidence. *Auditing: A Journal of Practice & Theory*. 6(2): 74-89.

Paas, F. 1992. Training strategies for attaining transfer of problem-solving skills in statistics: A cognitive-load approach. *Journal of Educational Psychology*. 84: 429-434.

Paas, F., and J. J. van Merrienboer. 1994. Variability of worked examples and transfer of geometrical problem solving skills: A cognitive-load approach. *Journal of Educational Psychology.* 86: 122-133.

Paas, F., and J. J. van Merrienboer, and J. J. Adam. 1994. Measurement of cognitive load in instructional research. *Perceptual and Motor Skills*. 79: 419-430.

Rose, J. M. 2002. Behavioral decision aid research: Decision aid use and effects. In *Researching Accounting as an Information Systems Discipline*, edited by V. Arnold and S. G. Sutton. 2002. American Accounting Association: Sarasota, FL: 111-133.

Sweller, J. 1988. Cognitive load during problem solving: Effects on learning. *Cognitive Science*. 12: 257-285.

Trompeter, G., and A. Wright. 2010. The world has changed – Have analytical procedure practices? *Contemporary Accounting Research.* 27(2): 669-700.

Wells, J. T. 2007. *Corporate Fraud Handbook: Prevention and Detection*. 2007. John Wiley & Sons: Hoboken, New Jersey.

1. ^ [a] [b] Liddell, Douglas (1976). "Practical tests of 2x2 contingency tables". *The Statistician* **25** (4): 295–304. doi:10.2307/2988087. JSTOR 2988087.
2. ^ [a] [b] Berkson, Joseph (1978). "In dispraise of the exact test". *Journal of Statistic Planning and Inference* **2**: 27–42. doi:10.1016/0378-3758(78)90019-8.
3. ^ [a] [b] D'Agostino, R. B., Chase, W., and Belanger, A. (1988). "The Appropriateness of Some Common Procedures for Testing Equality of Two Independent Binomial Proportions". *The American Statistician* **42** (3): 198–202. doi:10.2307/2685002. JSTOR 2685002.

Figure 1: Illustration of All Stages of Analytical Review

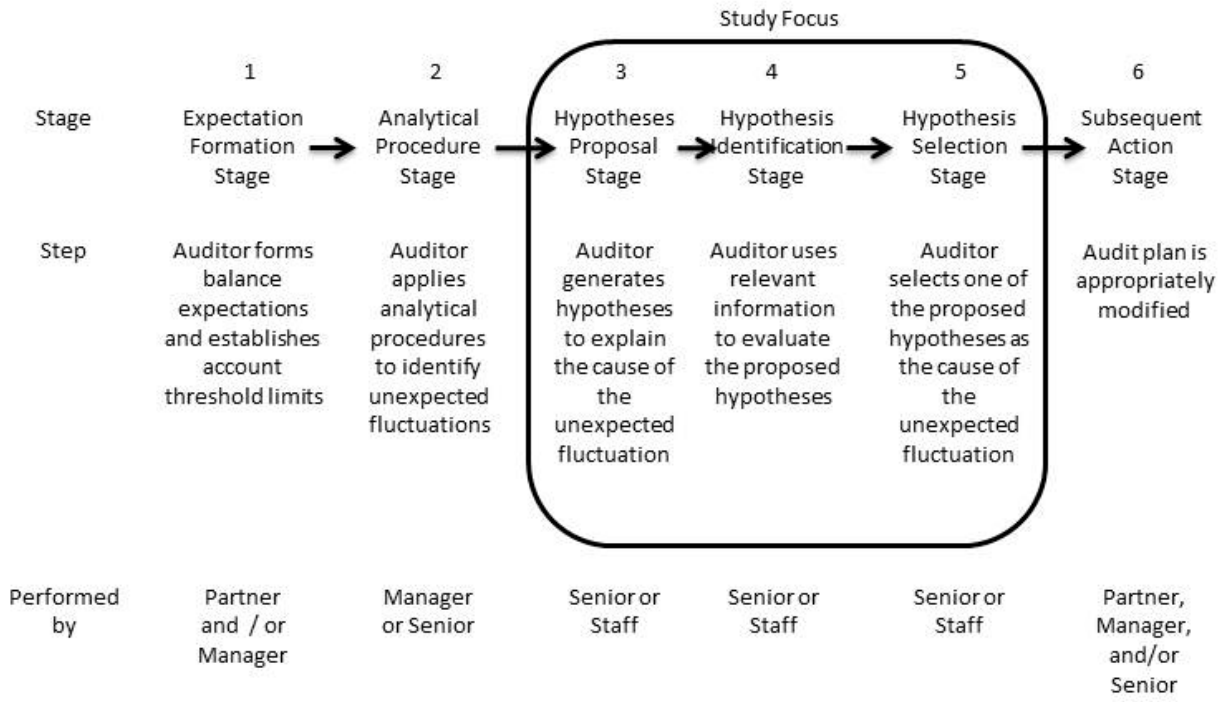


Figure 2: Activity Relationship Diagram (ARD)

(Zoomed in for illustrative purposes, Page 1 of 2)

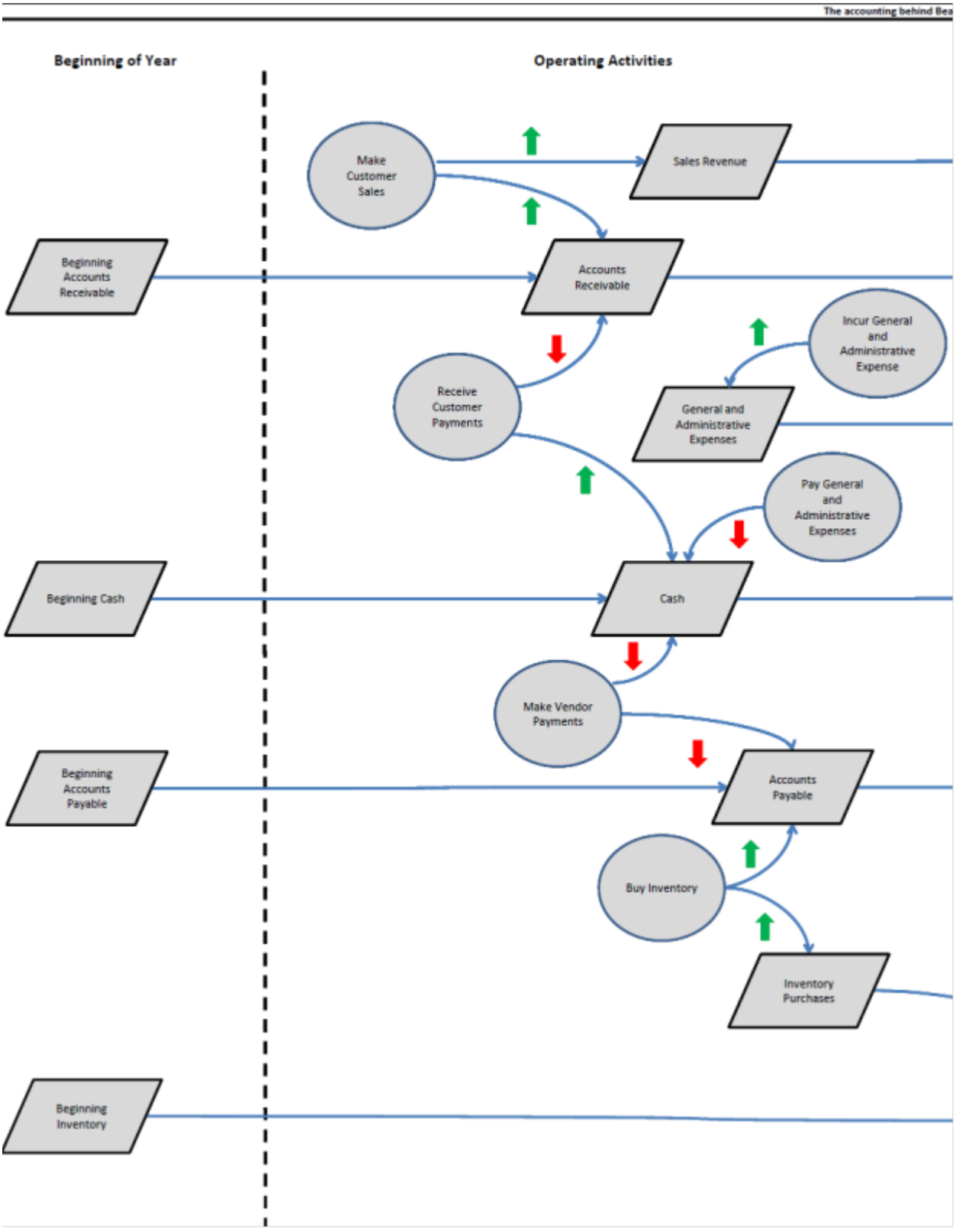


Figure 2 (continued): Activity Relationship Diagram (ARD)

(Zoomed in for illustrative purposes, Page 1 of 2)

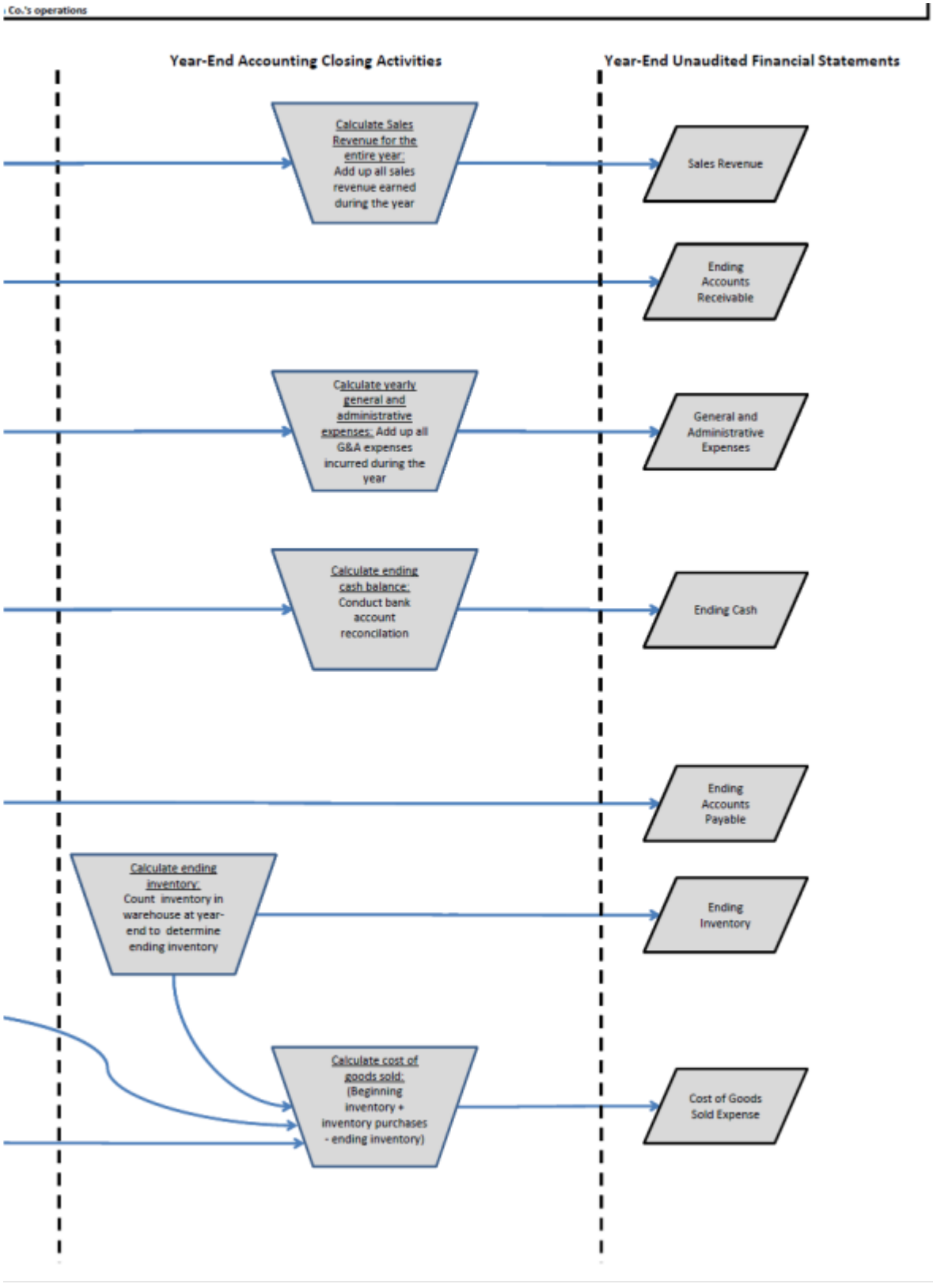
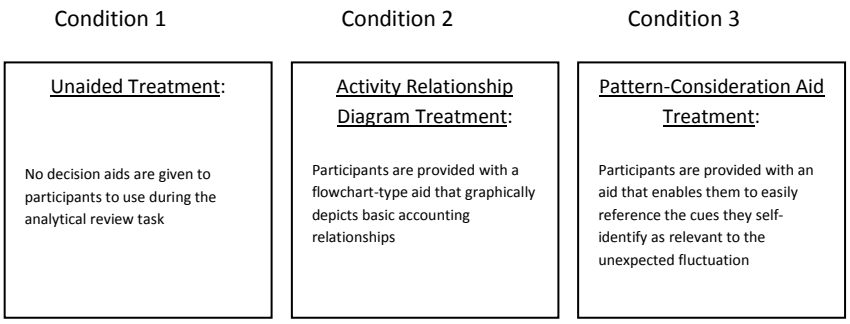


Figure 3: Research Design



Task: Explain the cause of an unexpected fluctuation by conducting analytical review across the hypotheses proposal stage, hypothesis identification stage, and hypothesis selection stage.

Participants: Accounting masters and senior students as a proxy for novice auditors

Figure 4: Experimental Procedure

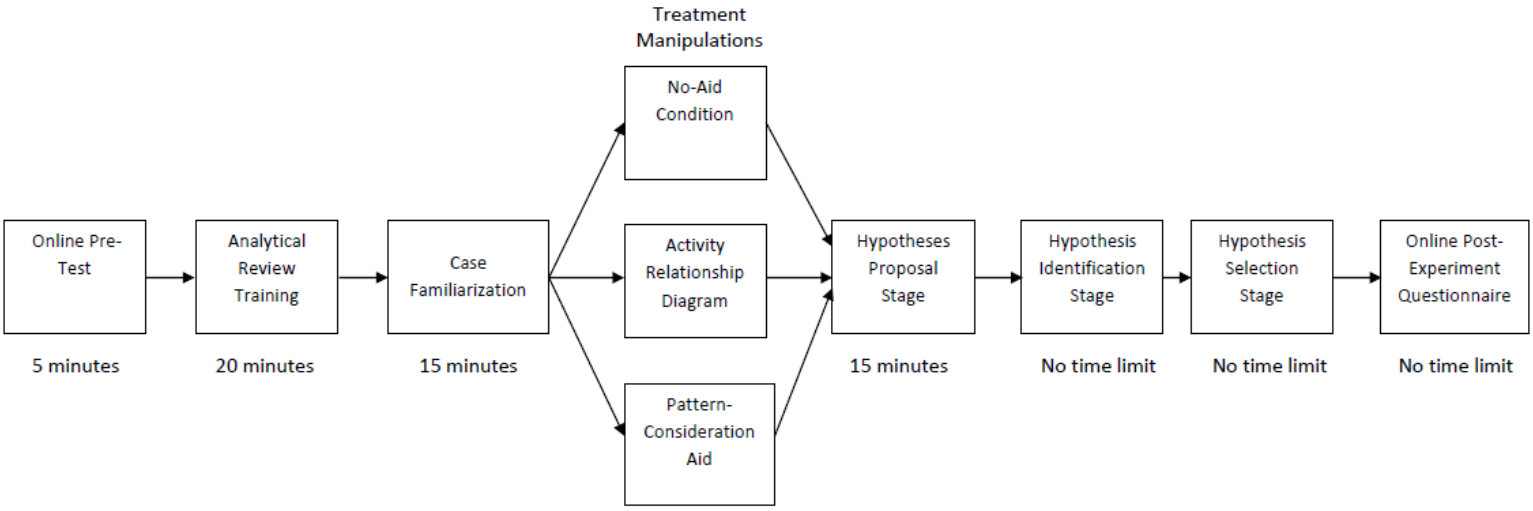


Figure 5: Screenshot Example of Participant-Identified Cues Displayed by the Pattern-Consideration Aid (PCA) During the Hypothesis Proposal Stage

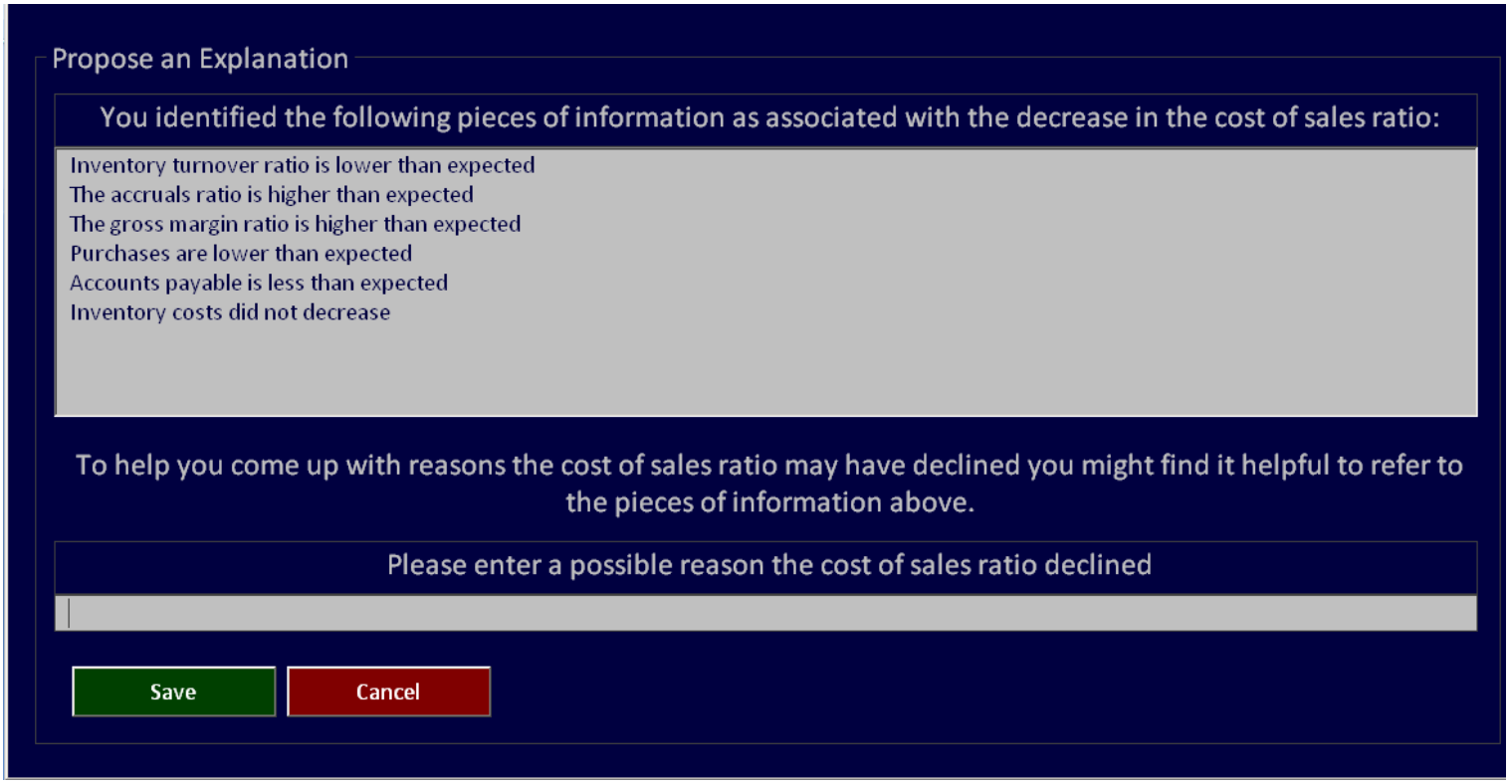


Figure 6: Screenshot Example of a Hypotheses Evaluation Checklist Displayed by the Pattern-Consideration Aid (PCA) During the Hypothesis Identification Stage

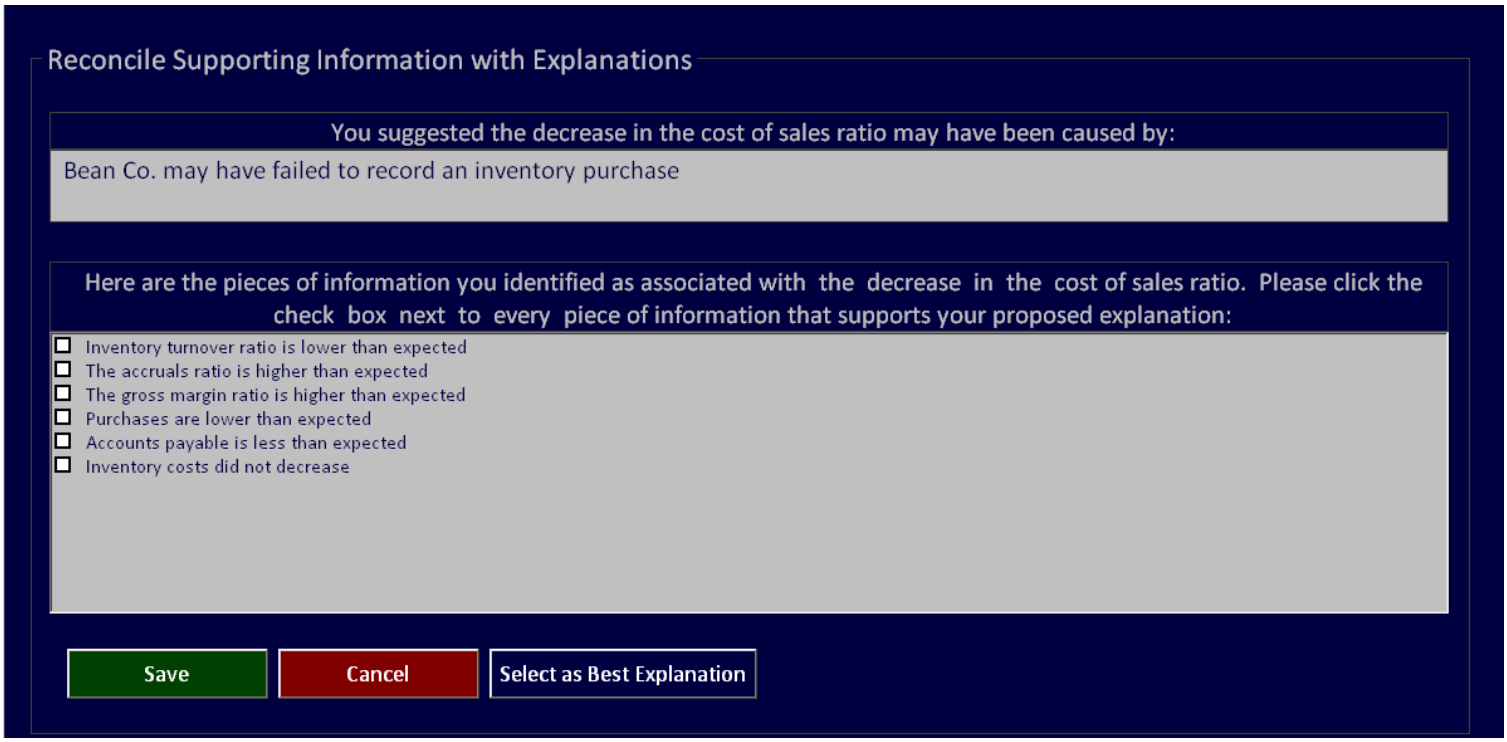


Table 1
H1: Analytical Review Effectiveness
Activity Relationship Diagram (ARD) vs. No Memory Aid (No-Aid)

Panel A: Analytical Review Effectiveness Descriptive Statistics

Treatment	N	Number of Times the Unexpected Fluctuation's Cause Was Selected	Proportion of Times the Unexpected Fluctuation's Cause Was Selected
No Aid	40	4	10.0%
ARD	45	14	31.1%

Panel B: Overall LOGIT Model and Significance

LOGIT	Overall Model χ^2	Degrees of Freedom	Significance
Model	14.771	2	0.001

Panel C: LOGIT Model Detailed Statistics

	B	Exp(B)	Standard Error	Wald χ^2	df	Significance
Memory Aid (0 = No Aid, 1 = ARD)	1.516	4.554	0.622	5.251	1	0.011*
Covariate: Cognitive Load (1 = Very Low, 7 = Very High)	-0.539	0.583	0.197	7.484	1	0.006**
Constant	-0.429	0.651	0.771	0.0309	1	0.578

Dependent Variable: The proportion of times participants selected the correct cause of the unexpected fluctuation.
*One-tailed.
**Two-tailed.

Table 2
H2: Analytical Review Effectiveness
Pattern-Consideration Aid (PCA) vs. No Memory Aid (No-Aid)

Panel A: Analytical Review Effectiveness Descriptive Statistics

Treatment	N	Number of Times the Unexpected Fluctuation's Cause Was Selected	Proportion of Times the Unexpected Fluctuation's Cause Was Selected
No-Aid	40	4	10.0%
PCA	37	10	27.0%

Panel B: Overall LOGIT Model and Significance

LOGIT	Overall Model χ^2	Degrees of Freedom	Significance
Model	7.104	2	0.029

Panel C: LOGIT Model Detailed Statistics

	B	Exp(B)	Standard Error	Wald χ^2	df	Significance
Memory Aid (0 = No-Aid, 1 = PCA)	2.629	13.854	1.414	3.456	1	0.032*
Covariate: Cognitive Load (1 = Very Low, 7 = Very High)	-0.744	0.475	0.410	3.287	1	0.070**
Constant	1.624	5.074	1.225	1.759	1	0.185

Dependent Variable: The proportion of times participants selected the true cause of the unexpected fluctuation.
*One-tailed.
**Two-tailed.